Population Genetics and Demographic Resilience in Three Aquatic Invertebrates

Thesis submitted for the degree of Doctor of Philosophy

by

Hannah C Macdonald

School of Biosciences

Cardiff University

July 2016

Supervisors: Prof. Mike Bruford, Dr Hefin Jones and Prof. Steve Ormerod

"In many of the more relaxed civilizations on the Outer Eastern Rim of the Galaxy, the Hitchhiker's Guide has already supplanted the great Encyclopaedia Galactica as the standard repository of all knowledge and wisdom, for though it has many omissions and contains much that is apocryphal, or at least wildly inaccurate, it scores over the older, more pedestrian work in two important respects.

First, it is slightly cheaper; and secondly it has the words

DON'T PANIC

inscribed in large friendly letters on its cover"

Douglas Adams, Hitchhikers guide to the galaxy

Acknowledgements

There is one person, without which I can say for certain I'd never have completed this thesis, my husband and best friend, Greg Macdonald. From deciding whether to start the PhD in the beginning, which involved moving to the other side of the country and having a long distance relationship for more than three and a half years, to the countless trials during, he gave me his full support and encouragement. He also lent me his computer programming skills and his OCD spelling and formatting checking skills, without which this document would be filled with my dyslexic spelling. When my funding came to an end he has financially supported me, and during my write up whilst being pregnant and all of the things that come with that, he has cooked, cleaned, walked the dog and generally cared for me so that I could concentrate on trying to hand this in. He's done all of this with endless patience, kindness and love. I don't deserve you. I promise I'll be a better wife from now on (well, I'll try) - I love you so much!

"Love doesn't make the world go round. Love is what makes the ride worthwhile."
Franklin P. Jones

To all of my lab family, thank you so much for all your help, support, advice and friendship: Isa-Rita Russo, Mario Barbato, Dave Stanton, Mafalda Costa, Josie Jackson, Jez Smith, Rosemary Moorhouse-Gann, Luis Cunha, Caitlin Pearson, Marian Pye, Beky Weiser, Pablo Orozco-terWengel, Jen Stockdale, Tano Cánovas, Frank Hailer, Niall McCann, Hugh Feeley, Pierfrancesco Sechi, Maria da Silva (Joana); Nia Thomas, Matt Dray and Steve Thomas. You are all so amazing, you are some of the kindest people I know, so giving of your time and (considerable) expertise, and I love you all! Particularly to Isa, Mario, Dave, Maf and Pablo – you were always there to help me and you always knew what to do, from day 1 till the end - thank you!

"I'll get by with a little help from my friends." The Beatles

Thank you to my supervisors, Mike Bruford, Hefin Jones and Steve Ormerod, firstly for giving me this opportunity in the first place, and for your continued help and guidance along the way. It was also a pleasure to work within the DURESS team lead by Isabelle Durance.

To my parents and whole family (Burtons and Macdonalds), thank you for your love and constant support; telling me that I could do it and understanding that I couldn't visit as much as I wanted to.

"In every conceivable manner, the family is link to our past, bridge to our future."
Alex Haley

Small shout out to J.K Rowling and Stephen Fry for the Harry Potter audiobooks which kept me sane through all the late nights and weekends alone in the lab completing the 12 000+ PCRs.

My dog, Merry (full name Meriadoc Brandybuck Macdonald), who got me out of the house for a bit of perspective and was always there for a cuddle, he also made working from home a lot less lonely.

"Happiness is a warm puppy." Charles M. Schulz

And lastly to my unborn son, you gave me the strictest deadline that any PhD student could have, and I'm so grateful, without you I'd have been dragging my feet. For you, I tried to stay (relatively) calm, because I didn't want you to have physiological issues because your mum was stressed during pregnancy; you made me look after myself (without you I'd have survived on cereal, illuminous coloured tinned soup, and caffeine); and I was kept sane during writing up because I could think of the amazing new adventure you, me, your dad and Merry will be starting when you arrive! I'm so looking forward to meeting you! Not long now! I love you. Mum x

"Babies are such a nice way to start people." Don Herold

Myself and this project was funding by The President Research Scholarship from Cardiff University and NERC funded DURESS project.

Table of Contents

List of Figures ix

Chapter 1 ix

Chapter 2 ix

Chapter 3 xi

Chapter 4 xii

Chapter 5 xii

List of Tables xiv

Chapter 1 xiv

Chapter 2 xiv

Chapter 3 xvi

Chapter 4 xvii

Chapter 5 xvii

Thesis Summary xix

Chapter 1 - Introduction 1

Chapter 1 - Part 1 2

1.1 General introduction 2

1.1.1 Why do we care about biodiversity? 2

1.1.2 Importance of freshwater ecosystems 3

1.1.3 Freshwater invertebrates 4

1.1.4 Importance of genetic studies 5

1.2 Main aims of study 6

1.3 Study species 7

1.3.1 *Amphinemura sulcicollis* 9

1.3.2 *Isoperla grammatica* 10

1.3.3. *Baetis rhodani* 12

1.4 Why microsatellites were chosen as the genetic marker 13

1.5 Invertebrate genetics in freshwater environment 15

Chapter 1 - Part 2 25

Species-Genetic Diversity Correlation Conceptual Framework 25

1.6.0 Abstract 25

1.6.1 Introduction 26

1.6.1.1 SGDC Theory 26

Case one: How can heterogeneous environments cause parallel responses? 27

Case two: How could genetic diversity affect species diversity? 28

Case three: How could species diversity affect genetic diversity? 31

1.6.2 Conclusions 32

1.7 Empirical studies 33

1.7.1 Studies showing a positive SGDC 34

1.7.2 Studies showing a negative SGDC 38

1.7.3 Studies showing no SGDC 39

1.8 Limitations/ knowledge gaps 42

1.8.1 Species 42

1.8.2 Genetic analysis 43

1.9 Conclusion 45

Chapter 2 - Methods 52

Highlights 52

2.0 Abstract 53

2.1 Site Selection 54

2.1.1 All Sites Visited 54

2.1.2 Choosing the final sites used for genetic analysis 55
2.1.2.1 Amphinemura sulcicollis 56
2.1.2.2 Isoperla grammatica 56
2.1.2.3 Baetis rhodani 56
2.2 Genetic Methods 62
2.2.1. Sample collection 62
2.2.2 Extraction methods 62
2.2.3 DNA Barcoding 64
2.2.3.1 Sequencing and analysis methods 65
2.2.3.2 Results and discussion 67
2.2.3.2.1 Amphinemura sulcicollis 67
2.2.3.2.2 Isoperla grammatica 67
2.2.3.2.3 Baetis rhodani 68
2.2.3.3 Cryptic species discrimination tool for *B. rhodani* 69
2.3 Next Generation Sequencing 74
2.3.1 *PrimerPipeline* development 75
2.4 Developing novel microsatellite markers 75
2.4.1 Method 76
2.4.2 Data analysis 77
2.4.3 Results and discussion 77
2.5 Final Dataset 78
2.5.1 Choosing Final Microsatellites 78
2.5.2 Data quality 79
2.5.3 Binning alleles 80
2.6 Null Alleles 84
2.6.1. Introduction 84
2.6.2 Data analysis 85

2.6.3 Results....85

2.6.4 Further investigation into effect on genetic structure....91

2.6.4.1 Data analysis....91

2.6.4.1 Results....92

2.6.5 Discussion....96

Chapter 3 - Genetic Structure and Diversity of Three Aquatic Invertebrates: Response to Environmental Stressors....99

Contributions....99

Highlights....99

3.0 Abstract....100

3.1 Introduction....101

3.2 Methods....105

3.2.1 Sampling and genotyping....105

3.2.2 Stressors....105

3.3 Statistical analysis....106

3.3.1 HWE and descriptive statistics....107

3.3.2 Genetic structure....108

3.3.3 Genetic diversity correlations....108

3.3.3.1 Genetic diversity....108

3.3.3.2 Stressors....109

3. 4 Results....110

3.4.1 HWE and descriptive statistics....110

3.4.2 Genetic structure....114

3.4.3 Genetic diversity correlations....121

3.4.3.1 Genetic diversity....121

3.4.3.2 Stressors....122

3.5. Discussion....131

3.5.1 HWE and descriptive statistics 131

3.5.2 Population genetic structuring 132

3.5.3 Genetic diversity and correlations with stressors 134

Chapter 4 - The Relationship between Species and Genetic Diversity in Freshwater Invertebrates 138

Contributions 138

Highlights 138

4.0 Abstract 139

4.1 Introduction 140

4.2 Methods 143

4.2.1 Invertebrate sample collection 143

4.2.2 Species diversity indices 144

4.3 Statistical analysis 145

4.3.1 Species genetic diversity correlation 145

4.3.2 Species diversity *vs* environmental stressor 146

4.4 Results 146

4.4.1 Species genetic diversity correlation 146

4.5 Discussion 154

Chapter 5 - A preliminary investigation of demographic resilience in three invertebrate species 158

Highlights 158

5.0 Abstract 159

5.1 Introduction 160

5.2 Methods 161

5.3 Statistical analysis 162

5.3.1 BOTTLENECK 162

5.3.2 MSVAR 163

5.3.3 MPVal 165

5.4 Results 167
5.4.1 BOTTLENECK 167
5.4.2 MSVAR 169
5.4.3 MPVal 175
5.5 Discussion 181
5.5.1 Statistical inference 181
5.5.2 Conclusions 184
5.6 Further work 185
Chapter 6 - General discussion 187
6.1 Overview 187
6.2 Synthesis of outcomes and results 188
6.3 Limitations 190
6.3.1 Observation study 190
6.3.2 Number of different sites 191
6.3.2 Hardy-Weinberg disequilibrium 192
6.4 Future directions 192
6.5 Conclusions 193
References 194
A 194
B 196
C 201
D 206
E 209
F 211
G 215
H 217
I 221

J 221
K 223
L...225
M 228
N...232
O...233
P 235
Q...239
R 239
S 241
T 245
U...248
V 248
W..250
X 253
Y 254
Z 254

Appendix A - Site Information and Sequencing Data 256

Appendix B - Genomic Resource Note 268

Appendix C - PrimerPipeline Instruction Manual 292

Appendix D - Microsatellite Development.. 320

Appendix E - Aquatic Conservation Manuscript 350

Appendix F - Final Data Set 377

Appendix G - Descriptive Statistics per Loci 457

Appendix H - Chapter 3 Supporting Information.. 466

Appendix I - Chapter 3 Reduced Loci 483

Appendix J - R Script 504

Appendix K - Species Diversity Data 508

Appendix L - Chapter 5 Supporting Information 526

List of Figures

Chapter 1

Figure 1.1 *Amphinemura sulcicollis* a) larva; b) adult (Photo credit: James Lindsey at Ecology of Commanster). 9

Figure 1.2. *Isoperla grammatica* a) larva (Photo credit: Lars L. Iversen); b) adult (Source http://www.orthoptera.ch/orthoptera-galerie/item/steinfliegen). 10

Figure 1.3. *Baetis rhodani* a) larva (Photo credit: Lars L. Iversen); b) adult (Source http://www.naturamediterraneo.com/forum/topic.asp?TOPIC_ID=10692). 12

Figure 1.4. 'Potential connections between genetic diversity and species diversity'. Modified from Vellend and Geber (2005). 27

Chapter 2

Figure 2.1. The location of all 60 freshwater invertebrate sampling sites throughout upland Wales. Blue circles represent sampling in 2012, yellow circles represent sampling in 2012 and 2013. Site codes numbered 1 to 60 which correspond with Table A1 in Appendix A. 57

Figure 2.2. The 14 sites from which *Amphinemura sulcicollis* samples were analysed. Blue circles represent sampling in 2013, Red circles represent sampling in 2012, yellow circles represent sampling in 2012 and 2013. All are labelled with their site code (See Table 2.1a). 58

Figure 2.3. The 12 sites from which *Isoperla grammatica* samples were analysed. Blue circles represent sampling in 2013, Red circles represent sampling in 2012, yellow circle represents sampling in 2012 and 2013. All are labelled with their site code (See Table 2.1b). 59

Figure 2.4. The 9 final dataset sites from which *Baetis rhodani* samples were analysed. Blue circles represent sampling in 2013, yellow circle represents sampling in 2012 and 2013 and are labelled with their site code (see Table 2.1c). 60

Figure 2.5. A neighbour joining tree of a) *Amphinemura sulcicollis* and b) *Isoperla grammatica* samples from 13 and 14 sites, respectively. 'F&R' = sequenced forward and

reverse. 'H1-11' represents the haplotype the sample belongs to. Circle's = haplogroup 1 (used for network diagram). GenBank accession numbers between KU955863-KU955923 (Appendix A, Table A2). 71

Figure 2.6. A neighbour joining tree containing samples of *Baetis rhodani* from 19 sites. Within the name: 'F&R' = forward and reverse was sequenced, and 'H1-13' represents which haplotype the sample belongs to. Green circle = haplogroup 1 (used for network diagram); Pink square = haplogroup 2, Black triangle = haplogroup 3; GenBank accession numbers: KU955924-KU955988 (Appendix A, Table A2). 72

Figure 2.7. Median Joining Network of a) *Amphinemura sulcicollis*, b) *Isoperla grammatica*, and c) *Baetis rhodani*, showing the diversity of haplotypes within the main haplogroup of each species. Black dots represent single mutations between haplotypes. 73

Figure 2.8. Photograph of a 3% ethidium-bromide stained agarose electrophoresis gel under UV light, showing PCR products after incubation with two restriction enzymes (*Bam*H1 and *Xba*1). Samples left to right: 100bp ladder; positive control; negative control; 102B14; 102B15; 102B16; 94B3; 94B4; 94B5; 94B6; 94B7; 94B8; 100bp ladder; 94B9; 94B10; 94B11; 94B12; 94B13; 94B14; 94B15; 94B16; 94B17; 94B18; 100bp ladder. 74

Figure 2.9. STRUCTURE plots showing results for *Amphinemura sulcicollis* using the whole dataset (10 loci including 4 loci that show evidence of null alleles: Amp_9, 10, 11 and 13) a) assumed two clusters; K=2, b) assumes three clusters; K=3 and c) assumes four clusters K=4. Site numbers shown below (grey shows partial sites), sites ordered from south to north of Wales.. 93

Figure 2.10. STRUCTURE plot showing results for *Amphinemura sulcicollis* using six loci only, not suspected of null alleles, a) assumed two clusters; K=2 and b) assumes three clusters; K=3. Site numbers shown below (grey shows partial sites), sites ordered from south to north of Wales. 94

Figure 2.11. STRUCTURE plot showing results for *Amphinemura sulcicollis* using six loci without evidence of null alleles plus Amp_9 a) assumed two clusters; K=2 and b) assumes three clusters; K=3. Site numbers shown below (grey shows partial sites), sites ordered from south to north of Wales. 95

Figure 2.12. STRUCTURE plot showing results for *Amphinemura sulcicollis* using six loci without evidence of null alleles plus Amp_10 for K=2 (identified by structure harvester to be the most likely value of K). 95

Figure 2.13. STRUCTURE plot showing results for *Amphinemura sulcicollis* using six loci without evidence of null alleles plus Amp_11 for K=2 (identified by structure harvester to be the most likely value of K).....96
Figure 2.14. STRUCTURE plot showing results for *Amphinemura sulcicollis* using six loci without evidence of null alleles plus Amp_13 for K=2 (identified by structure harvester to be the most likely value of K).....96

Chapter 3

Figure 3.1. *Amphinemura sulcicollis* STRUCTURE plots showing a) two clusters (K=2), b) three clusters (K=3) and c) four clusters (K=4). White bars within plot divide different sites, site numbers shown below (grey shows partial sites), sites ordered from south to north of Wales.....118
Figure 3.2. *Isoperla grammatica* STRUCTURE plots showing a) two clusters (K=2), b) three clusters (K=3) and c) four clusters (K=4). White bars within plot divide different sites, site numbers shown below (grey shows partial sites), sites ordered from south to north of Wales.....119
Figure 3.3. *Baetis rhodani* STRUCTURE plots showing a) two clusters (K=2), b) three clusters (K=3) and c) four clusters (K=4). White bars within plot divide different sites, site numbers shown below (grey shows partial sites), sites ordered from south to north of Wales. 120
Figure 3.4. Interpolation maps creating using QGIS showing mean allelic richness across sites sampled in Wales, UK, for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. If data were available at a site for more than one year, a mean value was used. .125
Figure 3.5. Correlations between genetic diversity indices (expected heterozygosity and allelic richness) and environmental stressors for: a) - d) *Amphinemura sulcicollis* showing correlations with pH and cadmium*;* e) - f) *Isoperla grammatica* showing correlations with cadmium and g) - h) *Baetis rhodani* showing correlations with total oxidised nitrogen. Where correlation exists a line of best fit has been added, created from a linear mixed model in RStudio. Stars above graphs denote significances: * = p-value <0.05, ** = p-value <0.01, *** = p-value <0.001.....130

Chapter 4

Figure 4. 1. Interpolation maps created using QGIS across sites sampled in Wales, UK, showing a) species richness, b) Shannon diversity and c) functional diversity. Values for all indices are available in Appendix K Table K4. If data were available at a site for more than one year, a mean value was used. 148

Figure 4.2. Correlations between genetic diversity indices (expected heterozygosity and allelic richness) and species diversity indices (species richness, Shannon diversity and functional diversity) for: a-f) *Amphinemura sulcicollis* (purple datum points); g-l) *Isoperla grammatica* (blue datum points) and m-r) *Baetis rhodani* (green datum points). Line of best fit created from a linear mixed model in RStudio. Stars above graphs denote significances: * = p-value <0.05, ** = p-value <0.01. 150

Figure 4.3. Correlations between species diversity indices (species richness, Shannon diversity and functional diversity) and environmental stressors a-c) pH, d-f) total oxidised nitrogen. Line of best fit created from a linear mixed model in RStudio. Stars above graphs denote significances: * = p-value <0.05, ** = p-value <0.01, *** = p-value <0.001. 152

Figure 4.4. Side by side interpolation maps of a) the mean allelic richness of *Amphinemura sulcicollis*, b) species richness and c) pH. These three factors have all been shown to correlate with one another. 153

Chapter 5

Figure 5.1. Depicting the three scenarios of effective population size that were tested using MSVAR runs 1-10: stable, bottleneck and expansion. 165

Figure 5.2. Showing the posterior distributions calculated using MSVAR comparing N_1 (effective population size at time Ta) = darker red line to N_0 (current effective population size) = lighter red line and posterior distributions of Ta (Time between N_0 and N_1) = green line, along with priors (dotted lines) for representative sites and clusters for Amphinemura *sulcicollis*. Each site/cluster contains ten independent runs: a and b) Site 104; c and d) Site 96; e and f) Site 95; g and h) Site 93; I and j) Cluster 3; k and l) Cluster 4. 172

Figure 5.3. Showing the posterior distributions calculated using MSVAR comparing N_1 (effective population size at time Ta) = darker red line to N_0 (current effective population

size) = lighter red line and posterior distributions of Ta (Time between N_0 and N_1) = green line, along with priors (dotted lines) for representative sites and clusters for *Isoperla grammatica*. Each site/cluster contains ten independent runs: a and b) Site 108; c and d) Site 93; e and f) Cluster 2. 174

Figure 5.4. Showing the posterior distributions calculated using MSVAR comparing N_1 (effective population size at time Ta) = darker red line to N_0 (current effective population size) = lighter red line and posterior distributions of Ta (Time between N_0 and N_1) = green line, along with priors (dotted lines) for representative sites and clusters for *Baetis rhodani*. a and b) Site 94; c and d) Site 102; having 9 and 3 independent runs, respectively. 175

List of Tables

Chapter 1

Table 1.1. A summary of certain similarities and differences between the three target freshwater invertebrate species that may be relevant to this study. 13

Table 1.2. Literature review of previous studies on freshwater invertebrates. This table represents the result of a comprehensive search for all studies of freshwater invertebrates of the class Insecta, which utilised microsatellite markers; also includes some examples of other invertebrates from the phylum Arthropoda (e.g. Crustaceans). References include description of microsatellite loci if published separately. 19

Table 1.3. Summary of 42 empirical studies and models whose focus was to compare genetic and species diversity (most recent - oldest). In the *Species used* column, blue indicates when a plant species has been used. In the conclusion column: pink = positive, blue= no correlation; green = mixed results and yellow =negative. 46

Chapter 2

Table 2.1a-c. Details of the final sites used for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. Sites are ordered approximately from south to north Wales, and by catchment. 61

Table 2.2. PCR reagents for the mtDNA PCR using DNA barcoding primers (LCO1490: GGTCAACAAATCATAAAGATATTGG and HCO2198: TAAACT TCAGGGTGACCAAAAAATCA, Folmer *et al.* (1994)). PCR was carried out in a final volume of 20 µl. 65

Table 2.3. PCR conditions for the mtDNA PCR using Folmer *et al.* (1994) primers (LCO1490: GGTCAACAAATCATAAAGATATTGG and HCO2198: TAAACT TCAGGGTGACCAAAAAATCA). PCR was carried out in a final volume of 20 µl. 65

Table 2.4. Details of reagents of restriction enzyme mix used to determine haplogroup of *Baetis rhodani* sample. Carried out in a final volume of 20 µl and incubated at 37°C for one hour. 70

Table 2.5. PCR reagents for singleplex PCR for microsatellites in Appendix E, Table E1. PCR was carried out in a final volume of 5 µl. 76
Table 2.6. PCR conditions for microsatellites in Appendix E, Table E1. PCR was carried out in a final volume of 5 µl. 76
Table 2.7a-c. Summary of final dataset for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhondoni*, per loci and then in total. *No. Repeated* = the number of individual samples that were repeated at least twice; *Repeated as %* = the percentage of total samples that were repeated at least once; *Total successful* = total number of successful samples genotyped, including repeats; *Total failed* = total number failed, including repeats; *Missing data* = the number of individual samples that could not be genotyped; *Missing data as %* = the percentage of data that are missing from the whole dataset per loci; *Not Included* = the number of samples not included in Appendix F final dataset, due to failure to amplify, usually due to species miss-identification; *Negatives* = number of negative samples (blank) sent for fragment analysis, (not included in Appendix F Final dataset); *Total samples* = total number of samples sent for genotyping; and *No. of alleles* = the number of different alleles in the dataset. 82
Table 2.8a-c. Summary of descriptive statistics for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*, per loci, per site (only sites with <10 samples per site were included). N = number of individuals; N_A = number of alleles per locus; H_O = observed heterozygosity; H_E = expected heterozygosity; HWE = p value of Hardy-Weinberg Equilibrium test; no. sig = total number of sites that are significantly out of HWE with and without Bonferroni correction (critical P value p ≤0.005)); % = percentage of sites significantly out of HWE with and without Bonferroni correction, and PIC = polymorphism information content. 87
Table 2.9a-c. Describing results from Micro-Checker for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*, per loci, per site (only sites with greater than 10 samples were included). Where 'yes' = there is evidence of null alleles and 'no' = there is no evidence of null alleles ([1] denotes evidence for scoring error due to stuttering). Total= the total number of sites that show evidence of null alleles per loci. 90

Chapter 3

Table 3.1a-c. Microsatellite diversity in a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*, using the full dataset. Mean (across loci) and standard error (SE) values calculated per site. N = number of individuals; Na = number of alleles; Ho = observed heterozygosity; uHe = expected heterozygosity; all calculated in GenAIEx. HWE = p-value of Hardy-Weinberg Equilibrium test calculated per site using Genepop, F_{IS} = the inbreeding co-efficient and p = the F_{IS} p value. **Bold** represents significance (critical p-value for HWE after Bonferroni correction = ≤0.005, ≤0.005, ≤0.0038, for *A. sulcicollis*, *I. grammatica* and *B. rhodani*, respectively). 111

Table 3.2a-c. Pairwise genetic differentiation (F_{ST}, distance method) of a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*, between sites, using the full dataset. Values with a p value of <0.05 in **bold**..116

Table 3.3a-b. Pairwise genetic differentiation (F_{ST}, distance method) of a) *Amphinemura sulcicollis*, b) *Isoperla grammatica*, between clusters, using the full set of loci. Values with a p value of <0.05 (*) in **bold**. 117

Table 3.4. Comparing pairwise allelic richness between sites, using full datasets for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. Only significant differences in allelic richness are presented: * p-value = <0.05, ** p-value = <0.01, *** p-value = <0.001. Blue = positive coefficient value, Red = negative coefficient value. The horizontal axis is the reference site, if the regression coefficient value was positive (blue) this means that site on the vertical axis has higher mean allelic richness than the reference site above. If the value is negative (red) the site to the left of the value has a lower mean allelic richness compared to the reference site above. 126

Table 3.5a-b. Comparing the allelic richness between each cluster, using full datasets for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. Only significant differences in allelic richness are presented: * p-value <0.05, ** p-value <0.01, *** p-value <0.001. Blue = positive coefficient value, Red = negative coefficient value. The horizontal axis is the reference cluster, if the regression co-efficient value was positive (represented here with blue asterisks) this means that site on the vertical axis has higher mean allelic richness than the reference cluster above. If the value is negative (represented here with

red asterisks) the cluster to the left of the value has a lower mean allelic richness compared to the reference cluster above. 127

Table 3.6. Summary of all linear mixed models preformed between the two indices of genetic diversity; (always on the y-axis) H_e = expected heterozygosity; AR = mean allelic richness) and four environmental stressors (on the x-axis); aluminium (Al, µg/l), cadmium (Cd, µg/l), pH (pH-units) and total oxidised nitrogen (TON, mg/l). Showing adjusted R-squared (R^2), p value (p) and number of sites (n) for each regression. Blue = a positive correlation (i.e. as the environmental stressor value increases the genetic diversity indices value increases); Green = a negative correlation (i.e. as the environmental stressor value increases the genetic diversity indices value decreases); Black = no correlation. Significant values in bold: * = p-value <0.05, ** = p-value <0.01, *** = p-value <0.001. 128

Chapter 4

Table 4.1. Summary of all mixed linear models preformed between the two indices of genetic diversity; (always on the y-axis) H_e = expected heterozygosity; AR = mean allelic richness and three indices of species diversity (on the x-axis); species richness, Shannon diversity and functional diversity. Showing adjusted R-squared (R^2), p-value (p) and number of sites (n) for each regression. Significant values in bold: * = p-value <0.05, ** = p-value <0.01. 149

Table 4.2. Summary of all mixed linear models preformed between the three indices of species diversity (always on the y-axis); species richness, Shannon diversity and functional diversity, and four environmental stressors (on the x-axis); aluminium (Al, µg/l), cadmium (Cd, µg/l), pH (pH-units) and total oxidised nitrogen (TON, mg/l). Showing adjusted R-squared (R^2), p-value (p) and number of sites (n) for each regression. Significant values in bold: * = p-value <0.05, ** = p-value <0.01, *** = p-value <0.001. 152

Chapter 5

Table 5.1. Range of priors and variances tested in ten independent runs and what scenario they represent, using MSVAR. N_0 = current effective population time; *var* = variance; N_1 =

Ancestral effective population size before event; μ = mutation rate; *Ta* = time interval between N_1 and N_0. 165

Table 5.2a-c. Showing results from the Wilcoxon test (two-phase (T.P.M.) model) and the mode shift test using BOTTLENECK for a) *Amphinemura sulcicollis*; b) *Isoperla grammatica* and c) *Baetis rhodani*. *P* = probability and *H* = heterozygosity. **Bold** = significance (p = <0.05). 168

Table 5.3a-c. Results from software MPVal and Critical M showing; *n* = sample size (2 x number of diploid individuals); *Site* = Site (cluster represented); *Average M* = as calculated using MPVal; M_c = Critical M value calculated assuming one of two extremes of θ (pre-bottleneck *theta*) and *d* = the different between the *Average M* and the M_c (*Average M* - M_c), where positive (blue) values indicate no signature of bottleneck, and negative (red) values indicate presence of a bottleneck. Calculated for a) *Amphinemura sulcicollis* b) *Isoperla grammatica* and c) *Baetis rhodani*. 177

Table 5.4a-c. Summary table showing: *Site (Cluster)* = site code with which cluster it represents in brackets; *n* = number of individuals; H_o = mean (across loci) observed heterozygosity; uH_e = mean (across loci) expected heterozygosity; *HWE* = p-value of Hardy-Weinberg Equilibrium test; F_{IS} = the inbreeding co-efficient; *p* = the F_{IS} p value; *AR* = mean (across loci) allelic richness adjusted for sample size. **Bold** represents significance. *BOTTLENECK*, *MSVAR* and *MPVal* represents a summary of these three methods, 'Yes' = this method found evidence of a bottleneck, 'No' = no evidence of bottleneck. For MPVal, this was based on the maximum θ tested for each species but * denotes when the result was consistent across both θ's tested. 179

Thesis Summary

Freshwater environments are threatened worldwide by external stressors and biodiversity decline, with major implications for ecosystem resilience. The genetic consequences so far have been neglected, especially for freshwater invertebrates, though their abundance, diversity, ease of sampling and functional importance renders them ideal candidates for genetic appraisal.

For three freshwater invertebrates (*Amphinemura sulcicollis*, *Isoperla grammatica* and *Baetis rhodani*) novel microsatellite markers were developed so that genetic structure, and genetic diversity could be assessed throughout upland Wales. The aim was to investigate dispersal and the genetic response to environmental stressors. Genetic diversity in these species was compared to species diversity across whole macroinvertebrate assemblages to investigate what factors might cause a correlation between these fundamental levels of biodiversity. The demographic history of each species was also investigated with the aim of assessing whether reduced genetic diversity was due to bottlenecks and more broadly, what this indicates in terms of the populations' resilience.

Species differed in their genetic structure and genetic diversity. All three species showed effective dispersal and geneflow, with each species displaying panmixa across catchments in southern and mid-Wales. However, *A. sulcicollis* and *I. grammatica* revealed genetic isolation and reduced genetic diversity at specific northern sites. Genetic and species diversity were correlated positively only in *A. sulcicollis,* where isolation combined with a common driver were the likely cause. There was evidence of recent bottlenecks in all three species.

All these results could be explained by an underlying genetic response to post-industrial acidification: reduced genetic diversity correlated significantly with acidity for *A. sulcicollis*, while reduced species diversity and genetic bottleneck signatures was consistent with chronic and episodic acidification across the Welsh region.

Overall, these results show how a positive correlation between species and genetic diversity can never be assumed, and illustrate how assessments of genetic health expand insights available from traditional biodiversity assessment.

Chapter 1 - Introduction

Part 1

General Introduction, aims, introduction to study species, invertebrate genetics and molecular markers used

Part 2

Specific literature review on species-genetic diversity correlation

Chapter 1 - Part 1

1.1 General introduction

1.1.1 Why do we care about biodiversity?

It is generally accepted that we are in the midst of a 'sixth extinction' (Leakey and Lewin 1995). Accurately estimating the number of species that have already become extinct during this period is extremely difficult, particularly as many species remain undiscovered. According to Dr Dias, (Braulio Ferreira de Souza Dias, Executive Secretary of the Convention on Biological Diversity) estimates of extinction rates vary between 2 and 100 million species (Knight 2012). Humankind relies on the planet's biodiversity for three main reasons. Firstly, for food, medicine, and a whole manifest of other materials and fibres. Secondly, for ecosystem services, for example, natural pest control and crop pollination (Frankham *et al.* 2002), and thirdly, biodiversity provides social benefits in the form of recreation and ecotourism. These services have been valued at between 490 international dollars per hectare per year (int$/hr/yr) provided by environments such as the open ocean, to almost 350,000 int$/hr/yr for environments such as coral reefs (de Groot *et al.* 2012). Ecosystem service demands will increase with an increasing population. Between 1997 and 2011, the global human population increased by 16% (from 5.83 to 7 billion), and is now estimated to be around 7.4 billion. As a consequence, ecosystems and the services they provide will become increasingly more stressed and less functional (Costanza *et al.* 2014; UN 2015).

All levels of diversity are important, and both species and genetic diversity contribute to ecosystem diversity (Hughes and Stachowicz 2004). Greater species diversity has been linked with higher productivity (Moreira *et al.* 2012), as well as increases in a community's resistance and resilience to a range of unpredictable events (Carreño-Rocabado *et al.* 2012; Bernhardt and Leslie 2013). Resilience is defined in ecology as the capacity of an ecosystem to respond to a disturbance by resisting damage and recovering quickly (Hodgson *et al*. 2015). Genetic diversity is an essential determinant of a population's evolutionary potential. It is this diversity that has allowed organisms to evolve and adapt to the environment. It is of particular relevance in the context of climate change, pollution, invasive species, disease,

pests and parasites (Frankham *et al.* 2002). For example, in dominant plants, genetic diversity has been shown to correlate with ecosystem resistance to disturbance (Hughes and Stachowicz 2004) and with ecosystem recovery after climatic extremes (Reusch *et al.* 2005).

1.1.2 Importance of freshwater ecosystems

Freshwater ecosystems are the most threatened habitats worldwide, with their biodiversity declining at a greater rate than any other ecosystem (Sala *et al.* 2000; Dudgeon *et al.* 2006; Sievert *et al.* 2016). Though covering only 0.8% of the Earth's surface, freshwater ecosystems contain 6% of all species (Dudgeon *et al.* 2006) and are worth an estimated 4 trillion US$ in terms of ecosystem services to society (Costanza *et al.* 2014). Rivers and streams provide direct economic benefit to humankind from water supply, fisheries, energy generation and recreation, e.g. recreational fishing, which is a hugely popular and profitable industry (Moss 1998), as well as social and health benefits provided by their cultural importance and aesthetics. Experiencing nature, such as walking along a riverside, has been shown to relax mental fatigue and improve cognitive function, as well as relieving stress and depression (Wolf and Flora 2010). Freshwater ecosystems also provide a range of functions such as biogeochemical cycling, carbon storage and local climate regulation (Anas *et al.* 2015; Williamson *et al.* 2016).

This immensely important ecosystem is, however, being threatened due to over-exploitation, pollution, flow modification, species invasion and habitat degradation (Postel and Richter 2003; Xenopoulos *et al.* 2005; Dudgeon *et al.* 2006; Jaramillo and Destouni 2015). The effects of climate change will exacerbate these effects further (Heino *et al.* 2009; Dohet *et al.* 2015), creating major ecological pressures on freshwaters and their conservation. The effects of fluctuations or extreme events due to climate change are predicted to simplify food web network structure, impair energetic transfer efficiency; and reduce resilience (Woodward *et al.* 2016). The position of freshwater habitats, embedded and downhill of the terrestrial world, makes them vulnerable to the anthropogenic activities occurring in the drainage basin, for example, land clearing, industry, and fertilization (Strayer 2006). Conservation efforts to protect aquatic environments are critical to slowing

the decline in biodiversity and preventing future losses (Master *et al.* 1998; Sievert *et al.* 2016). To plan for long-term biodiversity conservation, assessment of current biodiversity patterns is essential.

1.1.3 Freshwater invertebrates

Freshwater invertebrates are an integral and highly diverse part of the freshwater environment, with around 90,000 species described, representing 17 phyla and 570 families, with new species being frequently discovered even in well know groups and geographical areas (Strayer 2006). They are economically important as food sources, both directly (human consumption, e.g. mussels) or indirectly through their essential role in food webs, which link them to other trophic levels important in industry and conservation (e.g. fish and water fowl (Whiles 2013; Resh and Rosenberg 2014)). These invertebrates have many functional roles within their environment, such as regulating primary production, decomposition, water clarity, thermal stratification and nutrient cycling (Strayer 2006). Macroinvertebrates also represent one of the most important elements in aquatic food webs, primarily because of their consumption of large volumes of terrestrial material which enhances the energetic balance of the whole ecosystem (López-rodríguez *et al.* 2010).

Freshwater invertebrate populations make excellent models for basic ecological studies because of their abundance, diversity, ease of sampling and functional importance (Strayer 2006). Invertebrates are the most commonly used organisms worldwide for biological monitoring of environmental quality, conservation assessment and ecological health (Pfrender *et al.* 2010; Resh and Rosenberg 2014; Buss *et al.* 2015; Cardoni *et al.* 2015). They are used as model organisms for toxicological studies, and as water quality indicators (Lock and Goethals 2008), and are recognised as such by the Water Framework Directive (2000/60/CE) (European Commission 2000). Freshwater invertebrates face the same threats as freshwater ecosystems, though scientific knowledge, although substantial for some groups of invertebrate, is still far less than for vertebrates (Strayer 2006).

1.1.4 Importance of genetic studies

The International Union for Conservation of Nature (IUCN) and the Convention on Biological Diversity (CBD) both recognise that biodiversity must be conserved at three different levels: ecosystem, species and genetic (McNeely *et al.* 1990). Reducing the decline of biodiversity is set out by the 2010 Aichi Targets to the United Nations as a priority international goal (Pereira *et al.* 2013). The importance of genetic diversity is highlighted by the CBD's addition of Aichi Target 13, which seeks to minimise genetic erosion and safeguard genetic diversity. Unfortunately, this remains generally neglected in local, national and international conservation planning (Laikre *et al.* 2009; Taberlet *et al.* 2012), particularly in the freshwater environment (Crook *et al.* 2015). While assessments of the resulting risks in freshwater ecosystems are increasingly widespread, overwhelmingly they are based on traditional abundance and species compositional data and largely ignore indicators of genetic biodiversity, which may directly affect the resilience of the ecosystem (e.g. Villalobos *et al.* 2013; Chessman 2015; Storey 2015). To achieve Aichi Target 13, genetic diversity assessment should be an integral part of any biodiversity study (Taberlet *et al.* 2012) because a strong foundation of genetic knowledge is essential to preserve and protect genetic diversity (Hoban *et al.* 2013).

Molecular markers can be used to aid and inform biodiversity conservation. They can provide more cost-effective and reliable indicators of demographic change than some traditional approaches, providing information relevant to both ecological and evolutionary time-frames (Schwartz *et al*. 2007). For example, DNA barcoding can aid in monitoring the presence of a species, investigating hybridisation, and assessing for cryptic species (species which are morphologically identical but genetically distinct (Bickford *et al.* 2007)), that would otherwise be missed (Schwartz *et al.* 2007; Cardoni *et al.* 2015). Other methods, such as the use of microsatellite loci, can determine current population structure (e.g. connectivity or isolation) and longer-term changes in abundance and distribution (effective population size and genealogy) (Theissinger *et al.* 2011). Without genetic techniques, tracking changes in abundance would be labour-intensive and require sampling over long time-periods. This could be expensive and, for invertebrates, be quite inaccurate. For example, studying invertebrate dispersal without genetic techniques can prove logistically

difficult and could lead to inaccuracies, particularly if the species occur over larger geographical distances, and if dispersal movements are rare.

1.2 Main aims of study

This study sets out to assess and explore biodiversity in the vulnerable environment of river catchments, specifically within the understudied macroinvertebrate community. In order to study this, three specific macroinvertebrate species were chosen, and their genetic health was assessed using newly developed molecular markers. A species 'genetic health' incorporates their evolutionary potential and to some extent their resilience, and in this thesis, is measured by their genetic diversity.

Specifically, the main aims of this thesis were to investigate:

Chapter 3

1) The genetic structure of the macroinvertebrate species to explore dispersal and connectivity of streams in upland Wales.
2) The genetic diversity of macroinvertebrate species and whether this is correlated with certain environmental stressors (e.g. pH and metal concentrations).

Chapter 4

3) Whether species and genetic diversity are correlated within the freshwater macroinvertebrate community.

Chapter 5

4) The demographic resilience of each species. Demographic resilience is defined here as the resilience of a population. For example populations that have undergone a bottleneck but have since grown or expanded could be said to have demographic resilience, because they would have responded to a disturbance by resisting damage and recovering quickly (as in Hodgson *et al*. 2015). Specifically, it will be investigated whether any differences in genetic diversity between sites is due to past bottlenecks.

To provide appropriate background to these investigations, the remainder of Part 1 of this chapter sets out, firstly, to introduce the species studied; secondly, the reasons for choosing

microsatellite markers to achieve the above aims are discussed, and thirdly, to consider the limited availability of microsatellite loci for freshwater invertebrates and discuss previous studies in invertebrate genetics, focusing on the target species and species related to them.

Part two of this chapter introduces the conceptual framework and contains a specific literature review of previous studies comparing species diversity and genetic diversity, (pertaining to aim 3).

In Chapter 2 the study location and data collection methods are described, along with an investigation of cryptic diversity within target species using mitochondrial DNA. The final design and description of novel microsatellite markers used in Chapters 3 to 5 will also be presented. Chapter 3 explores the genetic structure and diversity of each species and how this reflects on the individual's dispersal, and hence their gene-flow. Correlations between genetic diversity and certain environmental stressors present in the freshwater environment are also investigated. In Chapter 4, correlations between genetic diversity and species diversity are investigated to determine the presence, if any, of a species-genetic diversity correlation (SGDC). Chapter 5 investigates the demographic history of each species to assess if any differences in genetic diversity could be caused by past bottlenecks and attempt to infer the resilience of each species. Finally, Chapter 6 synthesises the results found in all the previous chapters and discusses their significance and limitations.

1.3 Study species

Three macro invertebrate species, *Amphinemura sulcicollis*, *Isoperla grammatica* and *Baetis rhodani* were used to investigate the above aims in this study. These species were chosen because they represent different functional groups, yet, by reviewing past species data, were shown to co-occur at the same sites and are widespread enough to act as model organisms to study the genetic effects of a range of ecological pressures (Alp *et al.* 2012). Wide-ranging species with large population sizes are also the most likely to reveal historical population relationships, and allow the largest scope for comparison (Whiteley *et al.* 2006). *Amphinemura sulcicollis*, *I. grammatica* and *B. rhodani's* broad geographical range can be used to compare rivers and streams of contrasting types, while also meeting policy needs,

which increasingly requires biomonitoring and conservation policies over large spatial scales (Statzner and Bêche 2010). All three species have aquatic larval and terrestrial adult stages, allowing a focus on dispersal and connectivity. The three species do, however, differ in size, functional group and trophic level (Table 1.1). *Baetis rhodani* also had the added benefit of a small number of microsatellite loci already being established (Williams *et al.* 2002).

1.3.1 *Amphinemura sulcicollis*

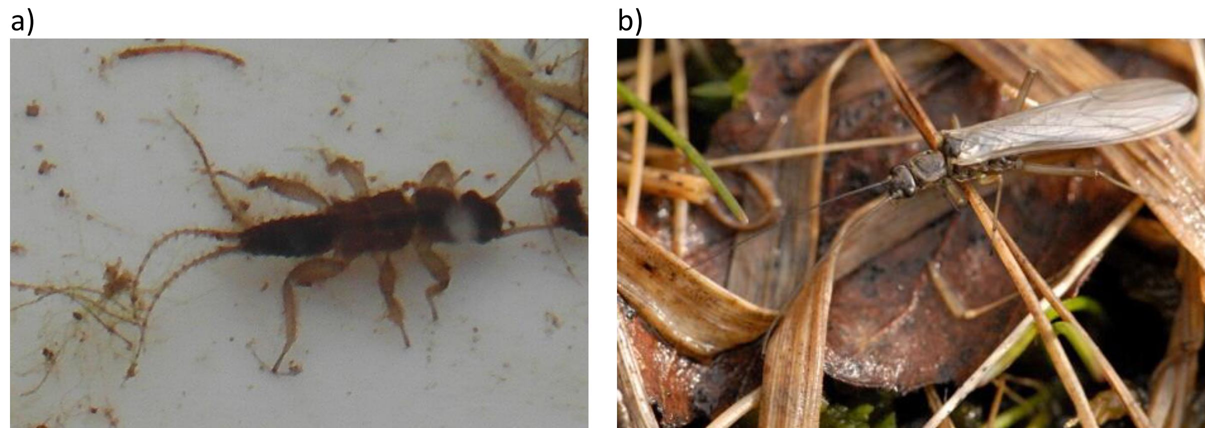


Figure 1.1 *Amphinemura sulcicollis* a) larva; b) adult (Photo credit: James Lindsey at Ecology of Commanster).

Amphinemura sulcicollis (Stephens, 1836; Figure 1.1) is an abundant stonefly (Plecoptera) species. As a widely distributed species in Europe, *A. sulcicollis* inhabits various freshwater habitats (many orders of rivers and lakes (Fochetti and Tierno de Figueroa 2008; de Figueroa *et al.* 2009)) and its ecology is well studied. Amphinemura sulcicollis is the smallest of the three species studied and Graf *et al.* (2002) categorised this species' functional feeding group as 20% shredder (feeding on fallen leaves, plant tissue and coarse particulate organic matter (CPOM)), 30% detritus feeder (collector-gatherer, eating sedimented fine particulate organic matter (FPOM)) and 50% grazer (feeding on epilithic algal tissues, biofilm, and POM). López-rodríguez *et al.* (2010), for example, categorised the species as collector-gatherer as it was recorded feeding principally on detritus and, to a lesser extent, CPOM. However, previous classical studies had noted that A. sulcicollis nymphs were herbivorous and fed mainly on leaves, mosses, detritus and algae (Hynes 1941, 1961a). Intraspecific differences in diet and ontogenetic variations have, however, been frequently found within Plecoptera (López-rodríguez *et al.* 2010) and, as Macneil *et al.* (1997) warns, `functional feeding groups' may actually be `fictional feeding groups'.

The life cycle of *A. sulcicollis* is generally univoltine (one generation per year). This has been observed in many studies (Hynes 1961b; Maitland 1966; Lavandier and Dumas 1971; Smith *et al.* 2000; Fjellheim and Raddum 2008; de Figueroa *et al.* 2009). Nymphs are present in the benthos from November to April-May, growing at a constant rate except towards the end of

their larval stage when the growth rate increases. The flight period of this species ranges from mid-April to June (Svensson 1977; de Figueroa *et al*. 2009).

Though the ecology of *A. sulcicollis* is well studied, the genetic knowledge of this species is very limited. Prior to this present study there were no known microsatellite loci, and GenBank, a comprehensive database that contains publicly available nucleotide sequences (Benson *et al.* 2013), only had 15 sequences associated with the species. Seven of these sequences are from mitochondrial DNA of the cytochrome oxidase subunit I gene (mtDNA COI), but were only used as a comparison for a study on a related species, *Amphinemura palmeni* (Boumans and Baumann 2012). Apart from this, there are a further six genes available from an unpublished study (histone H3 gene, COII and 12S, 18S, 28S, 16S ribosomal RNA), and one sequence each from Thomas *et al.* (2000) and Amore *et al.* (2009) (one 18S ribosomal RNA gene and one hexamerin mRNA sequence, respectively). From numerous literature searches, there is one study which, using somewhat outdated techniques, reports an electrophoretic survey of 16 enzyme loci to compare nine British stonefly species (Lees and Ward 1987). Answers to basic questions on the species' cryptic diversity, genetic structure and genetic diversity are therefore unknown.

1.3.2 *Isoperla grammatica*

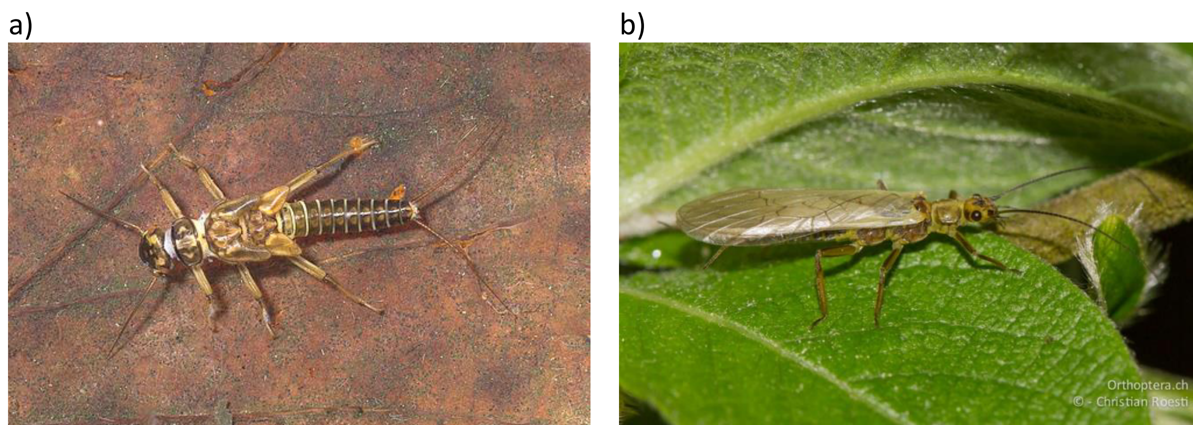


Figure 1.2. *Isoperla grammatica* a) larva (Photo credit: Lars L. Iversen); b) adult (Source http://www.orthoptera.ch/orthoptera-galerie/item/steinfliegen).

Isoperla grammatica (Poda, 1761; Figure 1.2) is also a stonefly species that, based on observations in other European rivers, typically has a univoltine life cycle (Hynes 1941; Hynes 1961b; Malmqvist and Sjöström 1989; Smith *et al*. 2000; de Figueroa *et al.* 2009b). Unlike *A. sulcicollis*, *I. grammatica* is a predator (70% of their diet comes from prey according to Graf *et al.* (2002), but the species can supplement its diet as 10% each comes from shredding, detritus feeding and grazing activities); it is known to prey upon both other target species, particularly *B. rhodani* (Malmgvist *et al.* 1991; Elliott 2003a, b). Although there is considerable variability in the size of the last instar, *I. grammatica* is always larger than *Baetis rhodani* and *A. sulcicollis* (Hynes 1941; Malmqvist and Sjöström 1989).

Nymphs are present in the benthos between November and mid-June; with the exact period depending on temperature (incubation period lasting 36 days at 16 °C (Elliott 1991)). The species has a large size range (explained by sexual dimorphism) and short flight period occurring between May and June (de Figueroa *et al.* 2009).

Genetic knowledge of this species is more limited than for A. *sulcicollis*; there is only one study that looks into respiratory proteins (the same study which also included *A. sulcicollis* and culminated in just two mRNA sequences deposited on Genbank for *I. grammatica* (Amore *et al.* 2009)). The same questions as for *A. sulcicollis* remain unanswered for this species.

1.3.3. *Baetis rhodani*

a) b)

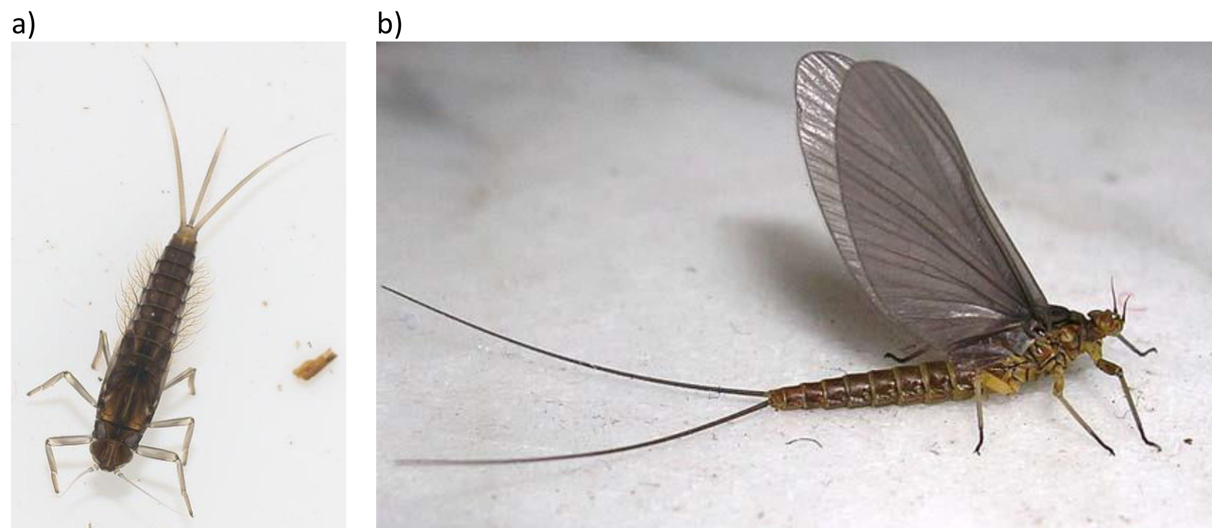


Figure 1.3. *Baetis rhodani* a) larva (Photo credit: Lars L. Iversen); b) adult (Source http://www.naturamediterraneo.com/forum/topic.asp?TOPIC_ID=10692).

Baetis rhodani (Pictet, 1843; Figure 1.3) is a widespread and abundant mayfly (Ephemeroptera) species. Within freshwater ecosystems individuals of the species act as generalist detritivores as they can be both scrapers and collector-gathers, feeding on periphyton and detritus (Baekken 1981). They are mainly found clinging to submerged rocks or plants in the rifle (Elliott and Humpesch 1983). Their high abundance has meant that *B. rhodani* has been well-studied. The species sensitivity to acidity makes them important indicators of water quality (Elliott *et al.* 1988); *B. rhodani* is a pivotal test species for investigating recovery of streams to acidification (Bradley and Ormerod 2002b; Andrén and Wiklund 2013) and metal pollution (Fialkowski *et al.* 2003).

Baetis rhodani can have either one or two generations per year. Adults will emerge between April and August/September, with a second generation over-wintering as nymphs (Elliott *et al.* 1988; Kowalik and Ormerod 2006). This pattern has also been observed within the current study sites (Tywi catchment, (Masters 2002)). Though subtle differences in growth and life cycles within this species have been observed, this can occur over short distances and be attributed to differences in riparian cover and stream water temperature (Imholt *et al.* 2010).

Of the three target species used in this study, *B. rhodani* is the only one that has previously been the test species of a large-scale molecular study (Williams *et al.* 2002; Williams 2003; Williams *et al.* 2006). Williams *et al.* (2002) designed seven microsatellite primers, which have continued to be used in other studies (Table 1.2). Two previous studies, assessing streams in Switzerland (Alp *et al.* 2012) and Italy (Rebora *et al.* 2005) found that *B. rhodani* had weak genetic structure due to its high level of dispersal. It has also been well established through genetic techniques that *B. rhodani* contains cryptic diversity (Williams *et al.* 2006; Bisconti *et al.* 2016).

Table 1.1. A summary of certain similarities and differences between the three target freshwater invertebrate species that may be relevant to this study.

	Amphinemura sulcicollis	*Isoperla grammatica*	*Baetis rhodani*
Phylum	Arthropoda	Arthropoda	Arthropoda
Class	Insecta	Insecta	Insecta
Order	Stonefly (Plecoptera)	Stonefly (Plecoptera)	Mayfly (Ephemeroptera)
Family	Nemouridae	Perlodiae	Baetidae
Genus	*Amphinemura*	*Isoperla*	*Baetis*
Abundance and Distribution	Dominant, widespread	Dominant, widespread	Dominant, widespread
Functional group	Collector-gatherer/ shedder/grazer	Predator	Generalist detritivore or scraper/grazer
Body Size	Smallest	Largest	Medium

1.4 Why microsatellites were chosen as the genetic marker

To study the genetic structure and diversity of these three target species, a suitable genetic marker had to be chosen. When choosing a genetic marker for a study, three factors must be considered: 1) the biological question being asked (Sunnucks 2000); 2) the timescale to be investigated; and, to a lesser extent, 3) cost. For the current study's aims, species-specific microsatellite markers (also known as simple sequence repeats / SSRs) were chosen as the most appropriate marker based on these three criteria.

Microsatellite markers are the most widely used molecular marker in population genetics; this is because of their co-dominance, hyper-variability among individuals, multiallelic nature and reproducibility (Ellegren 2004; Kalia *et al.* 2011; Liu *et al.* 2015). Microsatellites are frequently successfully used to aid in biodiversity conservation in freshwater ecosystems, for example: wetland birds (Corrêa *et al.* 2015), fish (Raeymaekers *et al.* 2005; Abdul-Muneer 2014; Junge *et al.* 2014) and a range of invertebrates (Jones *et al.* 2015; Lopes-Lima *et al.* 2015; Pérez-Portela *et al.* 2015). The variability of microsatellite markers results from variation in the number of repeat-motifs at a locus caused by replication slippage and/or unequal crossing-over during meiosis (Kalia *et al.* 2011).

Microsatellite markers provide a finer resolution for investigating population genetics compared to mitochondrial DNA (mtDNA). This is mainly due to a faster mutation rate (Shaw *et al.* 1999b), meaning microsatellites are better for investigating more recent evolutionary events such as isolation or bottlenecks. Mitochondrial DNA is better suited to phylogenetic studies but not for population differentiation at local scales (Flook and Rowell 1997). Mitochondrial DNA can be used to investigate more ancient branches of the phylogenetic tree and to separate species. For this reason, mtDNA was used to investigate cryptic diversity within the three target species (Section 2.2.3). Also, as mtDNA only provides the maternal side of the 'story', if, for example, levels of dispersal, and therefore gene flow, are different between males and females of the species due to behavioural differences, mtDNA would only reveal female dispersal.

Single nucleotide repeats (SNPs) are frequently used in the study of human (Consortium 2007; Yang *et al.* 2010) and domestic animal (Zhang *et al.* 2012) genetics. The wide availability of sequences in these species provide opportunities to generate millions of SNPs. In less well studied species, and in studies where there is a large sample size (this present study used 701 different individuals without any replication) however, SNPs can be less cost effective as they are identified by sequencing, which is more expensive compared to fragment analysis. Nuclear SNPs also have a slower mutation rate compared to microsatellites (Vignal *et al.* 2002), so unless there are a very large number available, microsatellite markers are more appropriate for studying recent events. Tenesa *et al.* (2007), for example, used 1 million SNPs to determine the recent human effective population size. Additionally, using the more cost effective microsatellite markers allowed

higher levels of repetition for the final genotype data, and therefore provided a more reliable dataset.

Other genetic markers have been previously used for similar population genetic studies. Some of these include: random amplified polymorphic DNA (RAPD, e.g. Baillie *et al.* (2000)), restriction fragment length polymorphism (RFLP; e.g. Carew *et al.* (2003)), amplified fragment length polymorphisms (AFLP, e.g. Gomez-Uchida *et al.* (2003)); intersimple sequence repeat (ISSR, e.g. Abbot (2001) and allozymes (e.g. Daniels *et al.* 2002). Allozymes are no longer used in population genetics and although the other techniques listed can still be useful in some studies, none of these techniques are as polymorphic as microsatellites (Powell *et al.* 1996; Garcia *et al.* 2004). Both RAPD and AFLP are multi-locus approaches, which although technically convenient, have weaknesses and limitations which have been shown in a variety of organisms: 1) a considerable proportion of polymorphism detected by these markers could be non-heritable or not even derived from the target organism; 2) many current genetic analysis software do not support multi-locus markers; and 3) neither RAPD nor AFLP are codominant, like microsatellite markers, therefore can only be scored as present or absent, whereas codominant markers have two alleles at each locus and can therefore be analysed more precisely (Sunnucks 2000). Applications of RAPDs and other techniques are more likely to be used for species discrimination and only for population screening when microsatellites are unavailable (Smith 2005).

Microsatellite markers were therefore deemed the most appropriate molecular marker for investigating fine scale genetic structure and genetic diversity within *A. sulcicollis, I. grammatica and B. rhodani.*

1.5 Invertebrate genetics in freshwater environment

Strayer (2006) suggested that, compared to vertebrate species, invertebrates are understudied. To investigate the current extent of this for genetic studies on freshwater invertebrates, a literature search of studies using microsatellite markers was conducted. The search, using Scopus and Web of Science, was carried out using the key words

'invertebrate', 'microsatellite' and 'freshwater'. Suitable studies listed in the bibliography and 'cited by' references were also included in the database. Thus, it is hoped, that almost all studies within the Class Insecta (the same Class as all three target species, Table 1.1) were included (last search: June 2016). There are some microsatellites that have been described for species within Insecta, but not yet used in a published study; for example, Dragonflies (Kurita *et al.* 2014) and Damselflies (da Silva-Méndez *et al.* 2013; Cao *et al.* 2015). Descriptions of loci are only included in Table 1.2 if they are in the same taxomonic order as one of the target species (i.e. stonefly (Plecoptera) or mayfly (Ephemeroptera)). Though concentrating on the Class Insecta, some examples are included from the Phylum Arthropoda (Table 1.2), for example, from the subphylum Crustacea. Freshwater invertebrates outside of the Arthropoda phylum have however been excluded (e.g. freshwater mussels (Froufe *et al.* 2013) and bryozoa (Freeland *et al.* 2000a, b; Freeland *et al.* 2001)).

Freshwater macroinvertebrate species make up a diverse and abundant community; it is thought that globally, more than 20,000 dipteran species breed in freshwater worldwide (Hutchinson 1993; Covich *et al.* 1999). In light of this, and the fact that freshwater ecosystems are so threatened and well-studied in other aspects, Table 1.2 represents the relative knowledge gap in terms of genetic studies on macroinvertebrates. A thorough literature search returned just 21 studies that used microsatellite markers to investigate population structure of 16 different species of freshwater invertebrates from the Class Insecta (Table 1.2). Paz-Vinas *et al.* (2015) conducted a literature review of all taxa living in river ecosystems to investigate the spatial distribution of genetic diversity. Of 79 case studies, only seven species were not fish or plants, and only three species were from Arthropoda. This highlights that, at least in terms of fine-scale genetic knowledge, the whole phylum is understudied. When looking within orders Ephemeroptera (mayflies) and Plecoptera (stoneflies), there are only five and three species studied respectively (Table 1.2). Mayflies' genetic structure (Gibbs *et al.* 1998; Rebora *et al.* 2005; Leys *et al.* 2016) has generally been studied to investigate either their dispersal (Williams 2003; Alp *et al.* 2012) or to identify areas of high genetic diversity for conservation (Taubmann *et al.* 2011). The number of loci used varies considerably between studies; one study only identifies two loci (Pathirana *et al.* 2012), while others range from five to ten loci (Table 1.2). It is unknown as

to how many unlinked loci are needed to distinguish between recently diverged populations, though a greater number of loci provides greater statistical power (Haasl and Payseur 2011; Putman and Carbone 2014). Many studies choose a minimum of ten polymorphic loci (Cervini *et al.* 2006; Östergren *et al.* 2015). There are three previous studies on target species, *B. rhodani*, which used a maximum of seven loci (4 – 7 loci, Table 1.2). Williams (2003) compared three different catchments; all subsequent studies compare *B. rhodani* across sites within the same catchment (though Alp *et al.* (2012) did include one outgroup population from a different catchment).

The two target stonefly species, *A. sulcicollis* and *I. grammatica*, have no previous studies using microsatellites and the number of studies on related species is small (Table 1.2). *Dinocras cephalotes* is the closest related species to *I. grammatica*, both being predatory stoneflies. This species had high connectivity between the sites analysed; the study used only four loci and all loci developed were significantly out of Hardy-Weinberg equilibrium (HWE). This study also found that *D. cephalotes* did not contain cryptic diversity as has been found for *B. rhodani* (Elbrecht *et al.* 2014). *Arcynopteryx compacta* was studied using six loci, though 11 were described and, again, departures from HWE were observed. Within this study strong population differentiation was found between mountain ranges, and genetic hotspots were identified. This information was used to shed light on freshwater invertebrate's response to Pleistocene climate change (Theissinger *et al.* 2013). For the only other stonefly species, *Brachyptera braueri*, microsatellites have been designed but only tested on a small subset of samples (Geismar and Nowak 2013).

One, if not the main reason resources for invertebrates are still limited when compared to vertebrates is largely due to the requirement to develop species-specific microsatellites. This constraint is diminishing as the cost of Next Generation Sequencing (NGS) declines while sequence output increasing; this has led to an acceleration in microsatellite locus discovery (Yu *et al.* 2011; Fernandez-Silva *et al.* 2013).

The genetics of freshwater invertebrates have usually been studied through methods that do not require species-specific markers. Nukazawa *et al.* (2015) used non-neutral amplified fragment length polymorphism (AFLP) to study four stream invertebrates (three caddisflies, *Hydropsyche orientalis, Stenopsyche marmorata* and *Hydropsyche albicephala*, and one

mayfly, *Ephemera japonica*). This study predicted spatial patterns along an elevational gradient, and attempted to explain the results in terms of adaption whereby lowland habitats only accommodate genotypes adapted to severe thermal conditions. Most studies on stream macroinvertebrates use mitochondrial DNA, usually cytochrome c oxidase subunit I (COI), though as explained in Section 1.4, mtDNA is most appropriate for phylogenetic rather than fine scale population studies. Addison *et al.* (2015) used mtDNA (COI) to examine the population genetic structure and phylogeography of the predatory caddisfly, *Rhyacophila minor*; Wickson *et al.* (2014) investigated the dispersal potential of Australian caddisfly Lectrides varians, and Saito and Tojo (2016) investigated the genetic structure of habitat generalist mayfly, *Isonychia japonica*. Using mtDNA, multiple species can be screened with the same marker making studies involving many species more accessible, whereas most studies with microsatellites focus on just one species, especially when loci do not co-amplify (see only two exceptions in Table 1.2). Chester *et al.* (2015) compared the genetic structure of six different freshwater invertebrate species in Victoria, Australia, identifying many differences between species. They found that the large caddisfly (Lectrides varians) was panmictic indicating strong dispersal capacity, while the small caddisfly (Agapetus sp.) had limited gene flow, indicating that size may impact dispersal ability. Both mayflies (Koorrnonga AV3 and Nousia AV1) showed evidence of gene flow among streams, also Nousia AV1 (the more widespread species) appeared to disperse across land by adult flight. The water penny beetle (Sclerocyphon sp.) showed an unusual pattern of genetic structure that indicated limited dispersal while the freshwater crayfish, Geocharax sp., had high levels of genetic structure indicating limited dispersal among streams (Chester *et al.* 2015).

Though results from mtDNA and microsatellite markers have been known to correlate, mtDNA has also been shown not to reveal genetic structure where microsatellites do (e.g. Feulner et al. 2004; Rodríguez et al. 2010; Teske et al. 2015). Therefore, when using mtDNA there is a chance that the fine scale genetic structure between populations could be missed.

Table 1.2. Literature review of previous studies on freshwater invertebrates. This table represents the result of a comprehensive search for all studies of freshwater invertebrates of the class Insecta, which utilised microsatellite markers; also includes some examples of other invertebrates from the phylum Arthropoda (e.g. Crustaceans). References include description of microsatellite loci if published separately.

No.	Species	Relevance to target species	No. microsats (No. of sites)	Region	Main aims (Aim) and conclusions (Conc.) of study	Reference
1	*Baetis alpinus*	Same Genus as target sp. Alpine mayfly	10 loci (24 sites)	Central Alps	**Aim:** Access cryptic genetic variation within *B. alpinus* **Conc.:** Consists of two reproductively isolated units sharing the same geographical locations, which differed in relative abundance, overall levels of genetic diversity as well as patterns of population structure	Leys *et al.* (2016)
2	Drusus discolor	Same Class (Insecta) as target spp. Montane caddisfly	20 loci (44 sites)	Two central German highlands	**Aim:** How dispersal modes affect population structure, and how landscape structure affects dispersal. **Conc.:** High levels of overland dispersal up to 20 km showing that surrounding landscapes or catchment boundaries did not drive population structure.	Geismar *et al.* (2015)
3	Heterotrissocladius marcidus	Same Class (Insecta, Order Diptera) as target spp. Chironomid, Midge	17 loci (21 loci developed) (6 sites)	High Tatra Mountains, Western Carpathians, Slovakia, Central Europe	**Aim:** Develop microsatellite markers and test their application **Conc.:** Bayesian cluster analysis revealed limited gene-flow between alpine lakes	Goffová *et al.* (2015)
4	*Parabathynellidae*	Same Phylum as target spp. Crustacean	8 loci (8 sites)	New South Wales, Australia	**Aim:** Assess small-scale genetic diversity and structure **Conc.:** Significant intraspecific genetic structuring was found, supporting the view of limited dispersal.	Asmyhr *et al.* (2014)

No.	Species	Relevance to target species	No. microsats (No. of sites)	Region	Main aims (Aim) and conclusions (Conc.) of study	Reference
5	*Dinocras cephalotes*	Same Genus as target spp. Predatory stonefly	4 loci (15 developed) Plus CO1 mtDNA (29 populations)	Ruhr river basin of North Rhine-Westphalia, Germany	**Aim:** Investigate intraspecific genetic diversity and population connectivity and cryptic species. **Conc.:** Found high connectivity among all populations through dispersal of the adult female. Though two distinct haplotypes, *D.cephalotes* is considered a single valid species	Elbrecht *et al.* (2014)
6	Wangiannachiltonia guzikae	Same Phylum as target spp. Crustacean Spring amphipod	11 loci (14 springs)	Springs of the Great Artesian Basin (GAB) in central Australia	**Aim:** Investigate fine-scale dispersal mechanisms to aid conservation management that maintain gene flow genetic diversity. **Conc.:** Low levels of gene flow and significant population differences found. As this sp. has no terrestrial life stage connectivity between springs is needed.	Robertson *et al.* (2014)
7	Paratanytarsus grimmii	Same Class (Insecta, Order Diptera) as target spp. Chironomid	6 loci Plus 2 mtDNA markers (5 populations)	Australia, England, Germany, Japan, and Canada	**Aim:** Compare mtDNA and microsatellite loci **Conc.:** Extremely low mitochondrial diversity but high levels of genotypic endemism suggesting P. grimmii populations have remained relatively isolated after an initial spread.	Carew *et al.* (2013)
8	Brachyptera braueri	Same Genus as target spp. Stonefly	13 loci (2 populations)	Germany	**Aim:** Identify microsatellite loci **Conc.:** Identified loci for use in fine-scale population studies	Geismar and Nowak (2013)
9	*Coenagrion mercurial* (3rd study on this sp.)	Same Class (Insecta, Order Odonata) as target spp. Southern Damselfly	12 loci (19 sites)	Oberaargau region, Switzerland	**Aim:** Investigate dispersal habitats compared to mark and recapture. **Conc.:** Populations well connected. Short-distance dispersal occurred along streams, and discontinuity of streams hindered dispersal. Long-distance dispersal was	Keller and Holderegger (2013)

No.	Species	Relevance to target species	No. microsats (No. of sites)	Region	Main aims (Aim) and conclusions (Conc.) of study	Reference
					suggested to happen along more or less straight lines (≤4500 m).	
10	Abedus herberti	Same Class (Insecta, Order Hemiptera) as target spp. Giant water bug	10 loci (20 populations)	Perennial stream habitats in arid regions of North America	**Aim:** Investigate genetic connectivity and its association with landscape variables (representing hypotheses of landscape-level connectivity) **Conc.:** Strong population structure was found. Concluded that population connectivity may depend on the shape of local overland topography rather than direct connectivity.	Phillipsen and Lytle (2013)
11	*Arcynopteryx compacta*	Same Genus as target spp. Arctic–alpine Stonefly	6 used (11 loci designed) Plus mtDNA CO1 (46 sites)	Variety of European mountain ranges	**Aim:** Investigated the glacial survival and post-glacial recolonization routes (European Pleistocene and Holocene history) of the sp. **Conc.:** Strong population differentiation between mountain ranges and genetic hotspots. Suggests that aquatic organisms may have reacted differently to Pleistocene climate change compared with terrestrial species.	Theissinger *et al.* (2009) & Theissinger *et al.* (2013)
12	1) *Baetis rhodani* (3rd study on this sp.) 2) *Gammarus fossarum*	(1) Target species (2) Same Phylum as target spp. Crustacean, Amphipod	1) 6 loci (11 Sites) 2) 9 loci (14 sites)	Within 1 Swiss pre-alpine catchment (looked within 35.7km)	**Aim:** Compared differences in small-scale connectivity (gene flow) between two spp's contrasting dispersal-related traits. Utilises landscape genetic approaches. **Conc.:** Weak structure in *B. rhodani* compared strong in *G. fossarum*. Attributed to dispersal capacity but also habitat specialisation and potentially the extent of local adaptation could be responsible for the differences in genetic differentiation.	Alp *et al.* (2012)
13	*Coenagrion*	Same Class	12 loci	Oberaargau region,	**Aim:** Assess functional connectivity of	Watts *et al.*

No.	Species	Relevance to target species	No. microsats (No. of sites)	Region	Main aims (Aim) and conclusions (Conc.) of study	Reference
	mercurial (2nd study on this sp.)	(Insecta, Order Odonata) as target spp. Southern Damselfly	(15 sites with ≥10)	Switzerland	populations of the endangered damselfly. **Conc.:** Effective dispersal between sites linked by agricultural land (1.5–2 km), however substantial elevation change and larger forest patches separated populations.	(2004) & Keller *et al.* (2012)
14	Chironomus riparius (Loci also cross amplify in Chironomus piger but have species-specific alleles for discrimination between the two)	Same Class (Insecta, Order Diptera) as target spp. Chironomid	5 loci (over many generations)	Laboratory experiment	**Aim:** Experiment exposing spp. to pesticide tributyltin doses over multiple generations. **Conc.:** Inbreeding and the level of genetic diversity strongly impacts life-history of midges. Increasing genetic diversity rapidly restores population fitness. Highlights importance of considering genetic diversity in ecotoxicology	Nowak *et al.* (2006) & Nowak *et al.* (2012)
15	*Rhithrogena japonica*	Same Genus as target sp. Mayfly	2 loci (7 sites)	Peninsula and central Honshu of Japan.	**Aim:** Identify microsatellite markers **Conc.:** Isolated two polymorphic loci and found impaired genetic flow between populations as a result of geographic distance and physical barriers.	Pathirana *et al.* (2012)
16	*Coenagrion mercurial* (1st study on this sp.)	Same Class (Insecta, Order Odonata) as target spp. Southern Damselfly	14 loci (9 sites)	Southern England, UK	**Aim:** Investigate whether there is demographic isolation between coexisting, but alternate, semivoltine cohorts. **Conc.:** Different cohorts differ in population size but gene flow is sufficient.	Watts *et al.* (2004) & Watts and Thompson (2012)
17	*Macrobrachium australiensis*	Same Phylum as target spp. Crustacean	4 loci (10 sites)	Murray–Darling Basin, western Queensland, Australia	**Aim:** Investigating connectivity and gene flow. **Conc.:** Found weak genetic structure but some significant F_{ST} values suggested that the population isn't entirely panmictic	Huey *et al.* (2011)
18	*Ameletus inopinatus*	Same Genus as target sp.	8 loci (10 loci designed)	Highlands of Central Europe and a more	**Aim:** Identify area's of high genetic diversity therefore areas of priority for conservation.	Theissinger *et al.* (2008) &

No.	Species	Relevance to target species	No. microsats (No. of sites)	Region	Main aims (Aim) and conclusions (Conc.) of study	Reference
		Mayfly	(31 sites, 18 with <10 samples)	continuous distribution across Fennoscandia and Northern, Euro-Siberia.	**Conc.:** High genetic diversity areas isolated, that if preserved would maintain present-day genetic diversity and continue to provide long-term suitable habitat under future climate warming scenarios.	Taubmann *et al.* (2011)
19	*Tasimia palpata*	Same Class (Insecta) as target spp. Stone-cased caddisfly	2 loci Plus mtDNA (COI) (2 populations)	Northern and Southern Australia	**Aim:** Investigate the role of the adult emergence patterns in genetic patchiness and compare patterns of genetic differentiation in subtropical and emerging temperate populations. **Conc.:** Though distinct patterns of genetic structure and deviations from HWE were observed in both populations, no evidence was found suggesting that these were caused by differences in emergence patterns.	Schultheis *et al.* (2008)
20	1) *Plectrocnemia conspersa* (2nd study on this sp.) 2) *Plectrocnemia flavomaculatus*	Same Class (Insecta) as target spp. Caddisfly's	1)7 (45 sites) 2)9 (10 sites)	South-east and north-west England	**Aim:** Investigate the hierarchical genetic structure both spp. and genetic structure of *P. conspersa* comparing the lowland south-east with that in an upland region in the north west. **Conc.:** Weak genetic differentiation in *P. conspersa* but strong genetic differentiation *P. flavomaculatus* suggesting different dispersal abilities.	Wilcock *et al.* (2001b); Dawson and Wilcock (2002) & Wilcock *et al.* (2007)
21	*Baetis rhodani* (2nd study on this sp.)	Target species	4 loci (6 sites from 3 streams)	Northern Italy	**Aim:** Preliminary analysis of the genetic differentiation of the mayfly B. rhodani **Conc.:** No specific population pattern observed though high FST values suggested limited dispersal across sampling sites.	Williams *et al.* (2002) & Rebora *et al.* (2005)
22	*Plectrocnemia conspersa*	Same Class (Insecta) as	6 loci (200 m stretch)	River Medway in south-east England,	**Aim:** Investigating small-scale patterns in oviposition and genetic relatedness.	Wilcock *et al.* (2001b) &

No.	Species	Relevance to target species	No. microsats (No. of sites)	Region	Main aims (Aim) and conclusions (Conc.) of study	Reference
	(1st study on this sp.)	target spp. Caddisfly		UK	**Conc.:** Siblings did disperse away from each other quickly and that kin structure did not persist over time and refutes the patchy recruitment hypothesis	Wilcock *et al.* (2005)
23	*Baetis rhodani* (1st study on this sp.)	Target species	7 loci (9 sites)	Tywi, Wye and Usk catchments in south Wales, UK	**Aim:** Testing hypothesis that *B. rhodani* had limited dispersal. **Conc.:** No genetic structure was detected, concluded that neither dispersal nor geographical isolation will limit the recovery of streams.	Williams (2003)
24	*Siphlonisca aerodromia*	Same Genus as target sp. Tomah mayfly	5 loci (6 sites)	Maine, USA	**Aim:** Assess genetic differentiation. **Conc.:** Found differentiation between sites. Suggested that differentiation was a result of geographical isolation and long-term reproductive isolation, due to climatic differences between these areas result in a substantial gap (~20 d) in emergence times of the reproductively active, short-lived adults.	Gibbs *et al.* (1998)

Chapter 1 - Part 2
Species-Genetic Diversity Correlation Conceptual Framework

1.6.0 Abstract

When facing global biodiversity decline from such threats as climate change and habitat loss, the need to accurately assess and conserve biodiversity has never been so great. One of the major gaps in knowledge is the possibility of a species-genetic diversity correlation (SGDC). Theory suggests that the underlying processes affecting species and genetic diversity could cause them to co-vary in response to environmental heterogeneity. Previous studies suggest that a positive SGDC is found more regularly than not, however caution must be taken because a positive correlation is not universal, and a conservation strategy that only considers one level of biodiversity may be missing vital information, which may affect its success. Further research is needed to discover in what environments and with what species a positive SGDC is found. Chapter 4 represents an empirical study into whether there is a relationship between species and genetic diversity within macro-invertebrate communities in freshwater streams across upland Wales. Meanwhile, here, the conceptual framework is laid out and the previous literature reviewed.

1.6.1 Introduction

During the past decade, studies (e.g. Vellend and Geber 2005; Cleary *et al.* 2006) have hypothesised a correlation between what many believe are the two fundamental levels of biodiversity: species and genetic diversity. Some researchers have speculated that one should be able to predict the other (e.g. He *et al.* 2008) and that species diversity could be used as a surrogate for genetic diversity, and *vice versa*. This possibility is appealing, as it would greatly simplify biodiversity assessments, and make them significantly more affordable. But do empirical studies comparing species and genetic diversity support this assumption, or is an assumed relationship over-simplifying complex ecosystems?

Janis Antonovics was one of the first to link species and genetic diversity, and suggested that a positive correlation might exist between them (Antonovics 1976, 1978, 1992). Prior to this, community ecology and population genetics tended to be treated and studied separately. More recently the number of multidisciplinary studies claiming to be part of either ‘community genetics’ or ‘ecological genetics’ has been gradually increasing (Antonovics 1976, 1992; LeRoy *et al.* 2012; Vellend *et al.* 2014; Xu *et al.* 2016).

1.6.1.1 SGDC Theory

Vellend and Geber (2005), in their review paper, synthesised the theory behind species-genetic diversity correlation (SGDC) and formalised a conceptual framework onto which others have built. They suggested three different ‘cases’ to argue how species and genetic diversity may be linked (Figure 1.4).

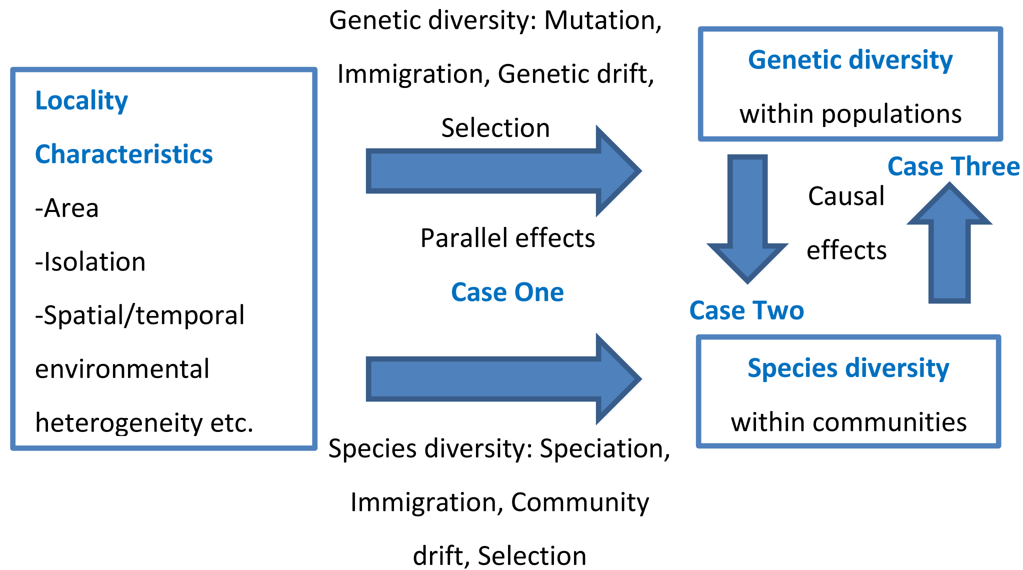


Figure 1.4. 'Potential connections between genetic diversity and species diversity'. Modified from Vellend and Geber (2005).

Case one: How can heterogeneous environments cause parallel responses?

Case one suggests that characteristics of different localities (a locality being defined as an area where the community may be found) offer spatially-varying selection pressures which may affect both species and genetic diversity. This is because the processes affecting both levels may be very similar and therefore cause a parallel response, meaning that both are either increased or reduced. Processes that may cause a parallel response include:

1) Mutation and speciation. These processes create new genes and species, respectively, but each work on very different time-scales. Vellend and Geber (2005) suggest that these particular processes have the least influence on causing a parallel response.
2) Immigration. The isolation and physical barriers of the locality will affect the rate of immigration and, consequently, the rate of individuals of different species and genotypes joining the community.
3) Genetic and community drift. These are random fluctuations in the relative abundance of alleles and species, respectively. The amount of influence drift has on a community can also be a factor of the environment as different localities can

support different sizes of populations within a community; population size is intrinsically linked with drift.

4) Selection. Certain genotypes and species have different relative fitness in different localities with different selection pressures, therefore spatial and temporal environmental heterogeneity can cause diversifying selection gradients.

Within Case one, environmental heterogeneity appears to play a large part in driving parallel responses in species and genetic diversity. There are two main types of heterogeneity that can exist within a locality:

1) Exogenous. This type of heterogeneity can be defined as the spatial or temporal variation in abiotic properties such as slope, temperature and soil quality, or in the freshwater environment, pH and water quality (Bell *et al.* 1993; Vellend and Geber 2005). As different species show individualistic responses to abiotic properties, it is not easy to predict the consequences to diversity of, for example, a pH or land-use gradient.
2) Endogenous. This is the effect of individuals interacting within the community, for example, the level of competition, and the proportions of predators and prey. A population of the same species is likely to be more successful in a locality with low compared to high predation. Again, individuals will respond in different ways to competition and immigration rates.

When investigating SGDCs, Vellend and Geber (2005) recognised that to make comparisons in genetic heterozygosity communities with similar composition must be used. Endogenous heterogeneity may therefore potentially have less of an effect, or certainly a harder one to investigate, compared to exogenous heterogeneity.

Case two: How could genetic diversity affect species diversity?

Case two describes the scenario where genetic diversity of the individuals in a locality could influence the species diversity and cause a SGDC. Case two seems to have the most diverse outcomes depending on the exact circumstances and composition of the community,

though usually relies on genetic diversity having a consequence (either positive or negative) on the individuals' present fitness rather than their evolutionary potential. Vellend and Geber (2005) reflected upon this relationship using a number of scenarios.

Scenario 1 - Population level effects

At the population level, high genetic diversity is thought to increase the population's fitness and hence reduce extinction risk, therefore preserving species diversity. There are three central mechanisms that might lead to a positive correlation between genetic and species diversity in an unstable environment, without a strong selection pressure:

1) If different genotypes utilise different resources, this may encourage co-existence (Antonovics 1978; Vellend and Geber 2005).
2) If genetic diversity can offer protection against natural fluctuations in environmental conditions in time and space (Williams 1975; Vellend and Geber 2005).
3) If genetic diversity acts as a cushion against the undesirable effects of competitors (Burdon 1987; Vellend and Geber 2005).

Scenario 2 – In a stable environment

If however the environment is stable and a strong selection pressure favours a few genotypes, then at first genetic diversity might reduce the fitness of the population, though this 'genetic load' would be swiftly removed from the population as species specialise and niche widths narrow (Crnokrak and Barrett 2002; Vellend and Geber 2005; Zheng *et al.* 2012). Under this scenario species and genetic diversity may show a negative correlation.

Scenario 3 – With a dominant species

Population level effects can transcend to community level effects if there are either only a small number of species in a community, or if there is a dominant species (Whitham *et al.* 2003; Vellend and Geber 2005; LeRoy *et al.* 2012). If there are only a few species and low genetic diversity is the cause of one of these species becoming locally extinct then this could cause a positive relationship between species and genetic diversity (Vellend and Geber

2005). For communities containing a dominant species, Vellend and Geber (2005) put forward two opposing scenarios which might lead to genetic diversity affecting species diversity:

1) If there is a widespread dominant species with different genotypes existing in different localities, then this species, subject to genotype, could differentially affect other species around them, perhaps through competition. This dominant species could therefore be driving its own form of 'spatially varying selection' (Vellend and Geber 2005) and would result in a positive or negative relationship between species and genetic diversity, depending on how it changed the competitive relationship. This hypothesis is Vellend and Geber's (2005) variation on Whittaker (1975) 'Diversity begets Diversity' Hypothesis which states that greater species richness drives higher rates of diversification.
2) If high genetic diversity in the dominant meant they were more productive or efficient at resource use, this could exclude less productive species and therefore reduce species diversity (driving a negative relationship). This is Vellend and Geber's (2005) variation on Elton (1958)'s Theory of Invasibility, where high diversity reduces the likelihood of invasive species from colonising.

Scenario 4 – Without a dominant species

What about communities without a dominant species? If species in a community were independent then they would follow the population-level effects and cause a positive relationship with higher genetic diversity reducing the probability of local extinction. It is however more likely that species within a community will not be independent and that some form of competition will exist between them. Competing populations with high genetic diversity may encourage co-existence and consequently retain high species diversity; genetic diversity is needed for evolution and co-evolution of competitors (Macarthur and Levins 1967; Pimentel 1968; Aarssen 1983; Taper and Case 1992; Vellend and Geber 2005). Though initial genetic diversity can cause high species diversity, in more stable environments genetic diversity can be lost through selection of fitter genotypes via competition, so a negative relationship may be seen (see Scenario 2).

Case three: How could species diversity affect genetic diversity?

Case three describes the theory of how species diversity may influence genetic diversity to cause a SGDC. Species diversity can be an important selection factor for individual populations. As in Case two, the hypotheses Vellend and Geber (2005) proposed for this causal relationship were variants of the Diversity Reduces Invasibility and Diversity Begets Diversity hypotheses. There are two hypotheses suggesting a causal effect of species diversity on genetic diversity, with opposing correlation predictions:

1) The number of competitors within a community could act as a form of 'stabilizing selection'. When there are many diverse competitors, only certain genotypes within the population can compete. With fewer competitors more genotypes could compete and gain access to resources (Vellend and Geber 2005). Such a situation could result in a negative relationship between species and genetic diversity.

Karlin *et al.* (1984) and Xu *et al.* (2016) found a negative correlation with genetic diversity and the number of conspecifics (see Section 1.8.2). Both of these studies support the hypothesis that interactions between species can act as directional selection, decreasing niche width and therefore reducing genetic diversity (Karlin *et al.* 1984; Vellend 2003). This supports Van Valen's 'Niche Variation Hypothesis', which states that niche widths should be greater in species-poor communities and narrower in species-rich communities (Van Valen 1965; Vellend and Geber 2005).

2) Similarly, with genetic diversity in case two (scenario 3 (1)), greater species diversity could act as a cause of 'diversifying selection' (Vellend and Geber 2005). In competition, different species may be able to co-exist with different genotypes of the dominant species (Harper 1977; Vellend and Geber 2005), thus causing a positive SGDC.

Vellend (2005) modelled relationships between species and genetic diversity in plant communities and found that area and immigration drove positive correlations. Although spatial heterogeneity had a positive effect on species diversity, the relationship was not always reciprocated in genetic diversity; in fact, positive, negative and unimodal relationships were found (Vellend 2005). One possible explanation for a negative

relationship is using Hubbell (2001)'s zero-sum assumption (resources are continually saturated therefore the community is at its carrying capacity). In Hubbell's Neutral Theory scenario, if different species immigrate into the community, the population size of each species must decrease, in turn potentially the genetic diversity within populations may decrease because of genetic drift. Hubbell's assumptions have however been criticised for being over simplified (Etienne *et al.* 2007).

1.6.2 Conclusions

The theory of community ecology and population genetics can both be stripped back to their own 'big four' fundamental underlying processes: selection, drift, speciation, and dispersal for community ecology; and selection, drift, mutation, and gene flow for population genetics (Vellend 2010). Analogues between these underlying processes were the reason researchers sought connections between species and genetic diversity. Many researchers feel, however, that such a simple framework could not be used to fully describe the apparently boundless number of processes that could affect a community within seemingly unique environments. There are many reasons why species and genetic diversity would not correlate because although the 'big four' have parallels at both species and genetic level, they are not the only processes to affect them. For example, genetic diversity is greatly influenced by recombination and random mutations (though population size is a factor in this), but species diversity has no obvious counterpart to the effects of recombination (Vellend and Geber 2005).

In all three cases (Figure 1.4), the theories depend heavily on the separate set of assumptions in each scenario. This helps illustrate that much depends on the context of the community and the environment. Whether species are independent; whether they are dominated by a few species or more evenly spread; the number of competitors; whether there is a gradient of environmental heterogeneity; and the stability of the interaction and environment are all factors that play a part in the direction and strength of a relationship between species and genetic diversity.

In some instances, the genetic models are very naïve. For example, Hubbell (2001)'s predictions only include neutral theory processes, which assume that all individuals have the same fitness and neglect any effect of selection. Though theories based on these models are a very useful starting point, inferring the selection constraints requires empirical experiments.

1.7 Empirical studies

As theory suggests, species and genetic diversity can be linked in many different ways, producing positive and negative correlations, and sometimes no correlation at all. The relationship that is observed depends on the exact processes that are controlling the ecosystem and which species are used to investigate genetic diversity. The variability evident in theory is reflected in the empirical literature (Table 1.3), with studies presenting positive, negative and no correlation. For this latter literature search the same technique as in Section 1.5 was used (keywords, 'species diversity' and 'genetic diversity'). As a result, a comprehensive list of articles was established that focus on the suspected correlation between species and genetic diversity in many different species and environments (Table 1.3 focuses on empirical studies therefore some reviews (Vellend *et al.* 2014), models (Laroche *et al.* 2015) and non empirical studies (Marske *et al.* 2013) are excluded). Studies vary greatly in the experimental approach taken to explore this issue; comparing the genetic diversity from five (Evanno *et al.* 2009) or ten (Odat *et al.* 2004) local sites, to 249 sites, spanning several countries (Taberlet *et al.* 2012). The studies vary between having a minimum of three (Taberlet *et al.* 2012) and 24 (Vellend 2004) individuals per species analysed per site. A range of different methods (Table. 1.3) have been used to detect genetic diversity, for instance, using amplified fragment length polymorphism (AFLP) (e.g. Odat *et al.* 2010; Taberlet *et al.* 2012; Frey *et al.* 2016), allozyme and isozyme alleles (e.g. Hosius *et al.* 2001) and microsatellite markers (He *et al.* 2008; Blum *et al.* 2012; Lamy *et al.* 2013; Xu *et al.* 2016), and sometimes more than one method combined (Vellend 2003; Vellend 2004; Fady and Conord 2010). Studies also vary in the number of species used, ranging from one, which is the most common (He *et al.* 2008; Evanno *et al.* 2009; Odat *et al.* 2010; Blum *et al.* 2012), to ten (Wehenkel *et al.* 2006) and 29 (Taberlet *et al.* 2012) species.

1.7.1 Studies showing a positive SGDC

The most common conclusion from these various studies has been that a positive correlation exists between species and genetic diversity (28 studies, Table 1.3). In such a wide range of environments as explored here, many suggestions of explanations for the positive correlation have been proposed. There are some that occur repeatedly:

Island biogeography

Vellend (2003) reported one of the first studies to explore this correlative relationship; compiling 14 datasets ranging from 5 to 14 different islands (see Table 1.3). The study found positive correlations in 13 of these datasets (including birds, reptiles, mammals and plants). Only five datasets yielded results that were significant at probability $P<0.05$ and another three deemed significant at $P<0.1$. This meta-analysis explains the positive correlations as being a consequence of parallel effects of area and isolation (supporting Case one). Area and Isolation are perhaps particularly relevant when investigating populations on islands, considering that area and dispersal limitations are the main elements of island biogeography (Kadmon and Allouche 2007).

Land-use

Vellend *et al.* (2004) and Odat *et al.* (2004) both looked at the effect different land-use had on SGDC and found similar results. Vellend *et al.* (2004) investigated whether this relationship extended beyond island ecosystems, and compared forest herbs between primary and secondary forestry sites. The effects of size and land-use history on the two levels of diversity drove a significant, albeit weak, positive correlation. Isolation and soil pH also had significant effects. Land-use effects were such that secondary forests had lower genetic and species diversity compared to that of primary forests. Higher divergence of genetic diversity within secondary forests was also found. The cause of this was thought likely to be genetic drift through bottleneck and founder effects.

Odat *et al.* (2004) on the other hand, found that meadow buttercups in five 'semi natural' and five 'agriculturally improved' grassland communities also had a positive SGCD. Overall genetic diversity between populations (beta) was significantly correlated with species evenness. The study gave two possible reasons for this:

1) Different communities may create different selection pressures on the individual species within it and that this shapes genetic diversity (supporting Case three).
2) Properties of the individual habitat in localities may create ecological barriers to gene flow and thus enhance beta genetic heterogeneity (supporting Case one).

Genetic diversity was also found to be higher, but not significantly so, in agriculturally-improved sites compared with semi-natural. Considering the relatively low sample size (five), this result is encouraging and perhaps with increased sample size a significant result would have been recorded. Although no substantial gene-flow obstructions were observed between the two land-use types, the correlation was thought to be the result of the two land-uses being very different in abiotic (management intensity and past usage) and biotic (species composition) factors. Therefore this study supports Case one, that both exogenous and endogenous heterogeneity drove a parallel SGDC.

Not all of Odat *et al.* (2004)'s data, however, presented positive correlations; within population (alpha) genetic diversity was not correlated with species evenness or richness. These authors hypothesised that this was potentially due to low statistical power or to the species chosen. Most of the genetic variation occurred within rather than between populations, the chosen plant species (*R. acris*) was perhaps the reason for the high alpha diversity; it is known to be a widespread, long-lived and outbreeding species.

As Odat *at al*. (2004) and Vellend *et al*. (2004) indicated, land-use, and hence the level of disturbance, may produce positive correlations in species and genetic diversity. Cleary *et al.* (2006) and Evanno *et al.* (2009) both looked at the effect of recent habitat disturbance. The former investigated this effect using butterfly communities within rainforests in Indonesia where the habitats are regularly disturbed by fires and logging; while Evanno *et al.* (2009) observed freshwater gastropod species within floodplain pools in France where the habitats

endured large periods of drought. Both studies found that species and genetic diversity had parallel responses to disturbance and supported case one (Figure 1.4).

Blum *et al.* (2012) also looked at the effect of environmental heterogeneity and land-use on the SGDC in stream fishes. They found that allelic richness showed a positive correlation with species richness, and that both declined with deleterious conditions related to agricultural land-use. Species richness and species Shannon diversity was in turn significantly correlated with the following: 1) water chemistry factors (total nitrates, ammonia, chlorides and sulphate), and 2) distance from the mainstream. The study also found that predictors of genetic diversity were: 1) population size of the target species; 2) certain elements in water chemistry (total phosphorous, nitrogen, Kjeldahl nitrogen, suspended solids and volatile solids); 3) substrate measures (embeddedness, % fines, % gravel and % bedrock); and 4) land-use cover. No relationship was found between Shannon diversity and allele Shannon diversity, so again, like Odat *et al*. (2004), not all of the results show positive correlation. From Blum *et al.* (2012)'s study however, many significant correlations were found between population size, genetic and species diversity, and it served to highlight that when environmental gradients are recorded, then predictors of species and genetic diversity can be found.

Disturbance

Cleary *et al.* (2006), Evanno *et al.* (2009), Wei and Jiang (2012) and Frey *et al.* (2016) all found a positive significant SGDC and attributed this to disturbance, though with different consequences according to the level of disturbance. Cleary *et al.* (2006) concluded that a range of factors could drive a correlation in the face of *El Nino* Southern Oscillation-induced disturbance, particularly fires, for example: 1) individual sites' environment; 2) fragmented habitat; 3) dispersal capability of the constituent taxa. Disturbance through fires caused reduced species and genetic richness, and potentially reduced the communities' resilience to future disturbances (Pease *et al.* 1989; Cleary *et al.* 2006). In Cleary *et al.* (2016) species richness was higher in logged rather than pristine areas. This was thought to be because slightly disturbed areas created greater environmental heterogeneity, which in turn opened up new niches and reduced previous dominant competitors (Rosenzweig 1995; Cleary *et al.*

2006). Another possible explanation is that disturbance could lead to source/sink dynamics where the sink (i.e. more disturbed) localities may be sustained, through immigration (of genes and species), by the source localities (i.e. less disturbed) (Zaccarelli *et al.* 2008).

Evanno *et al.* (2009) found that within site (alpha) species and genetic diversity both decreased over a short time period after a disturbance occurred. Species diversity was thought to have reduced because of the decrease in average water levels and the increased predation pressure that resulted when pool volume was reduced. In parallel, genetic diversity was thought to have decreased because of the reduced population size amplifying the effects of drift, and the drought causing reduced rates of migration. On the contrary, among-site (beta) species and genetic diversity increased. This was thought to be partly due to community and genetic drift that caused separated sites to diverge and partly because the five different localities had different intensities of disturbance, and therefore differed in selection pressure.

Frey *et al.* (2016) looked at SGDC along varying levels of disturbance, both natural and anthropogenic (in the form of landscape alteration) in sand dune plant community along 240 km of coastline. They found positive correlation and concluded that anthropogenic disturbance (specifically recreational infrastructure within 10 km of the site) significantly lowered genetic and species diversities.

Wei and Jiang (2012), focusing on one tree species within riparian forests in central China, however, had a contradictory result to other studies concerning the effect of disturbance, finding a positive correlation in undisturbed environments but not disturbed forests. They suggested that disturbance may result in reduced species diversity but unaffected genetic diversity, indicating non-parallel response to disturbance. This study furthermore cautions against generalizations about SGDC following disturbance.

Observational vs manipulated

While most of the empirical studies employ an observational field experiment to investigate SGDCs, some studies have looked at this relationship through manipulated experiments, for example, Booth and Grime (2003) and Fridley and Grime (2009). The former set up a

manipulated experiment to look at 11 long-lived plant species. The experiment consisted of three treatment groups: 1) where all 16 individuals per species were genetically unique; 2) where the 16 individuals were represented by four randomly selected genotypes for each species, and 3) all individuals per species were identical. Genetic diversity within the constituent species of each treatment group was found to reduce the rate at which species diversity declined, giving support to the theory that genetic diversity promotes resilience. Adam and Vellend (2011) corroborate this as they found testable predictions of beneficial effects of species diversity effecting genetic diversity (Case three) in a model on grass-clover dynamics. This study showed that communities with greater species richness of grasses retained greater genotypic richness within the clover population.

Fridley and Grime (2009) tested whether genetic diversity (in the form of one, four and eight genotypes per species) influenced community composition and annual aboveground productivity across different species diversity (communities of one, four and eight species) in a manipulated experiment. Their results support Case two because they found a marginal increase of species diversity with increased genetic diversity in four- and eight-species communities, and genetic diversity altered the performance of genotypes in monospecific communities of *F. ovina* though increased genetic diversity did not effect community productivity.

1.7.2 Studies showing a negative SGDC

Magurran (2005) stated that negative correlations between species and genetic diversity are rarely seen in nature. This does seem to be reflected in the literature (see Table. 1.3), with only a limited number of examples where a negative correlation has been found. However some studies that conclude that there is no correlation, still show negative relationships with certain diversity indices (e.g. Taberlet *et al.* 2012 with genetic and species endemism and rarity, see Section 1.7.3).

Only two studies found a negative relationship, Karlin *et al.* (1984) and Xu *et al.* (2016). Karlin *et al.* (1984) found that allozyme heterozygosity of a species of salamander was negatively correlated with the number of species of the same genus. Marshall and Camp

(2006), however, found the opposite result when they studied salamanders. They even suggested that Karlin *et al.* (1984)'s result could be biased because the samples they recorded did not fully represent the entire area (Marshall and Camp 2006). The most recent study to have recorded a negative relationship between SD and GD was Xu *et al.* (2016) within tropical seasonal rainforest. Silvertown *et al.* (2009) also suggested a type of negative relationship when studying the effect of resource addition on genetic and species diversity. Though they did not find a negative correlation, they did find that in separate plots genetic diversity increased with resource addition, whereas species diversity decreased.

All of the above go some way in supporting the Niche Variation Hypothesis which states that populations with decreased species diversity (hence less interspecific competitors) promote broader niches that are expected to have greater genetic diversity (Van Valen 1965). Adding more resources will increase the availability for the existing species, which may broaden their niches and enable more genotypes to persist (Silvertown *et al.* 2009). Silvertown *et al.* (2009) used the Resource Competition Theory (Tilman 1982) to explain why species diversity would decrease. The theory states that when supplying one nutrient too fast, this might favour one species and enable it to remove its competitor. This experiment however, does have the limitation that different resource elements (N, P, K, and Mg) were added to different plots, so different resources may have had different effects on species and genetic diversity.

Xu *et al.* (2016) had a similar explanation for their negative SGDC, suggesting that increased soil pH and phosphorus availability positively influenced tree diversity, which lead to a reduced niche breadth for their target species *B. roxburghiana*; which in turn, through competition, reduced its genetic diversity.

1.7.3 Studies showing no SGDC

As expected not all studies show just a positive or negative result, some show a mixture of all outcomes or show no correlation at all. There are six studies that found no correlation and six studies that found a mixture of correlations.

Fady and Conord (2010) and Taberlet *et al*. (2012) both presented broad, ambitious meta-

analysis studies that concluded that no correlation existed between species and genetic diversity. The former study found that the species diversity of vascular flora had no apparent spatial structure, however genetic diversity of 27 plant species (mainly woody trees) co-varied with an east to west gradient throughout the Mediterranean Basin. This was thought to be due to the east to west climate gradient during the last glacial maximum, although they acknowledge that this was not reflected in species diversity. The study could have been improved by recording more environmental heterogeneity like Blum *et al.* (2012); this might have led to other predictors of species and genetic diversity being found. Also, species diversity was calculated from presence–absence data, it may be more appropriate to consider species evenness or richness so that patterns could be explored on a smaller scale.

Taberlet *et al.* (2012), one of the largest studies in this subject area, looked at vascular plants in the Alps (spanning France, Italy, Switzerland, Germany, Austria and Slovenia) and the Carpathian Mountains (spanning Austria, Slovakia, Poland, Hungry, Ukraine and Romania). Patterns of species and genetic diversity did not co-vary across the study area and the results highlighted different areas as biodiversity hotspots for genetic and species diversity. A conservation strategy based only on species diversity would therefore not protect the genetic diversity of these vascular plants. This supports a similar previous study; Puşcaş *et al*. (2008) investigated this relationship with just one species in the same European mountain regions as Taberlet *et al*. (2012). Puşcaş *et al*. (2008) also found discrepancies between species and genetic diversity; they even found that the highest genetic diversity was actually in species-poor areas, indicating a negative relationship.

Similarly to the Mediterranean Basin, the Alps and Carpathian Mountains are unique environments, and environmental heterogeneity was not recorded in either study. Furthermore, based on particular selection criteria for localities for this study (>1,500 m in altitude, >50 species, and >10 genetic species), all localities investigated had relatively high species diversity and high altitude, therefore presented less heterogeneity. Taberlet *et al.* (2012) also chose the genetic diversity samples arbitrarily, rather than basing their choice on environmental heterogeneity gradients. Another possible limitation was that to calculate genetic diversity only three individuals were sampled per species per site. Though the

authors attempt to test whether this influenced the results, this appears a very low figure to find significant differences.

Molecular marker, AFLP, were used in Taberlet *et al.* (2012)'s study to find genetic diversity, a cheap, easy and reliable method that was very practical in a study of this size. Mueller and Wolfenbarger (1999), however, suggested that there was a disadvantage in using AFLP to identify homologous markers (alleles). This method, they argued, may be less useful for studies that require precise assignment of allelic states, such as heterozygosity. Perhaps, for this type of study, using microsatellite markers would have identified greater heterozygosity.

Taberlet *et al.* (2012) did, however, find some correlations:

1) Genetic rarity was positively correlated with species rarity. This, they argued, may be because genetic diversity may be linked to the glacial history of the mountains. Small refugia may have been able to harbour numerous different species, but population sizes would be small leading to decreased genetic diversity via drift.
2) Genetic diversity was averaged over all species per cell and this overshadowed some species-specific response: the genetic diversity of two (of 24) species were significantly correlated with total species richness in the Carpathians, and two (of 27) in the Alps. Six Alp species showed significantly negative relationships.
3) Negative correlations were found when comparing genetic endemism and diversity to species rarity and endemism. This was also the case when investigating the legume functional group.

The numerous conflicting results observed in Taberlet *et al.*'s study suggest that a general correlation cannot be applied. Studies at this scale potentially loose 'the devil in the detail'; if both genetic and species diversity are averaged over a large area (Fady and Conord 2010) based on a few point samples then smaller scale heterogeneity within that environment could be overshadowed.

1.8 Limitations/ knowledge gaps

1.8.1 Species

Number of species assessed

Many of the empirical studies discussed above share the same limitation in that they base their correlations on data from just one species within each community. Only 18 of the 42 studies explored the genetic diversity of more than one species within the community (Table 1.3). Patterns of genetic diversity can vary substantially, even amongst relative species (Zhou *et al.* 2016), therefore caution has to be exhibited in extrapolating from one species to whole communities. Even studies that do look at a larger number of species usually use a very low sample size, for example Frey *et al.* (2016) and Baselga *et al.* (2013) both study a range of species within their environment but are restricted to a maximum of 5 samples per site. Most genetic studies would exclude sites with less than 10 individuals (e.g. Keller *et al.* 2012).

Plant bias

A large proportion of the studies directly comparing species and genetic diversity have used a plant or tree as the focus species (Table 1.3). Given that invertebrates constitute a large proportion of the biodiversity of our planet (estimated to be up to 850,000+ species with 80–95% insect species not yet collected, named and described (Stork 2007)), it is surprising that there is such limited knowledge for invertebrates and SGDCs. Considering the global decline in biodiversity and the essential roles that invertebrates have in ecosystem functioning, it is vital that our knowledge of invertebrate diversity distributions is widened and that invertebrates are considered in conservation strategies (Schuldt and Assmann 2010). Determining whether, and to what extent, a relationship exists between species and genetic diversity in invertebrates remains a priority question. It has however already been acknowledged that certain plant communities and invertebrate populations are directly linked through their trophic interactions; Knops *et al.* (1999) found that a reduction in grassland plant diversity led to a reduction of invertebrate community richness. Conclusions

arising from correlations of species diversity to genetic diversity in different plants cannot, however, be automatically transferred to other species.

Freshwater environment and active dispersal

There are just six studies that take place within the freshwater environment. There is one study on fish (Blum *et al.* 2012), the rest use invertebrates: two work on freshwater gastropods (Evanno *et al.* 2009; Lamy *et al.* 2013) two on Crustacea spp. (Derry *et al.* 2009; Sei *et al.* 2009) and one on aquatic beetles, of which Dytiscidae (predaceous diving beetles) was the dominant type of species (Baselga *et al.* 2013). With the possible exception of the predaceous diving beetles (active dispersal of which is greatly dependent on the individual species (Bilton 2014)), none of these species possesses the ability of terrestrial flight. This life history trait has a huge impact on dispersal and connectivity, therefore may have a large effect on whether a SGDC is observed. Even within the marine and terrestrial systems there are only the bat (Struebig *et al.* 2011) and the butterfly (Cleary *et al.* 2006) that have potentially unhindered movement around their environment.

Csergo *et al.* (2014) attributed the positive correlation they found to habitat size, isolation affect, immigration rate and population size. This suggests that an individual's dispersal and corresponding gene-flow may have a huge effect on whether genetic and species diversity correlate. Considering the effects of isolation on genetic diversity and its suggested effect at driving a positive SGDC (Vellend 2003), not investigating species with greater dispersal ability represents a knowledge gap within the study of SGDC.

1.8.2 Genetic analysis

As a result of the effort involved in studies of large-scale genetic analysis, most studies, especially those that look at more than one species, tend to use mitochondrial DNA markers or another form of marker that is less polymorphic than microsatellite markers (Section 1.4). For example, Wehenkel *et al.* (2006) found a significant correlation between 'trans-specific genetic diversity per species' and species diversity. The approach they used to reach this genetic diversity measure however was very different from other studies. Wanting to use a genetic diversity assessment that would cross the species boundary, so that genetic

diversity could be compared across and between different species, they used isozymes and compared the electrophoretic patterns to find matching patterns across species. If patterns were found in different species that matched then Wehenkel *et al.* (2006) argued that this indicated the same functional characteristic. Three functionally similar enzymes were found among the chosen species, and to make the trans-specific genetic diversity measure, frequencies of all variants of each enzyme system were considered units and entered into the same diversity measure. Wehenkel and his colleagues compared 'trans-specific genetic diversity per species' (trans-specific genetic diversity, divided by species diversity, which they used as proxy for how well overall genetic variation matches species variation) and species diversity and found a significant positive correlation. Possible limitations of this study are: 1) it may have been difficult to compare enzyme patterns accurately; they do not state exactly how they decided which enzyme pattern matched in charge and size; and 2) considering species diversity was only recorded for a functional group of ten species, then it could not have varied much between sites.

Unlike Wehenkel *et al*. (2006), who used isozymes, nine studies chose to use microsatellite markers to find heterozygosity (Table 1.3). Although, the results from isozymes and microsatellites should not contradict each other (as shown by both markers frequently being used together within many studies, for example Chavarriaga-Aguirre *et al*. 1999; Vellend 2004), it may be difficult to compare them accurately. Isozymes are limited by a low number of loci and alleles per locus, whereas microsatellites are characterised by high heterozygosity and multiple alleles (Karasawa *et al.* 2012). Microsatellites should give higher resolution, and can differentiate between populations of the same species, whereas when using isozymes it depends which protein is chosen to compare; if it's a non-essential protein then it may vary between species, but if it is an essential protein, the likelihood is that it will not.

There are only two studies that use microsatellites to look at more than one species; Struebig *et al.* (2011) used between 8 and 15 loci to study three bat species and Lamy *et al.* (2013) used between 8 and 10 loci to study two species of gastropod. Both studies found a positive relationship between SD and GD.

1.9 Conclusion

Published studies suggest that overall, positive correlations are more frequent than negative or no correlations; however the results remain confused and controversial. Between, and sometimes within studies, different results arise as a result of a wide range of factors. It is perhaps to be expected that a definitive answer over a range of environments and species is non-existent. Without empirical evidence on a case by case basis, conservation strategies only considering species diversity may not be safe guarding a population's genetic diversity.

Although there is considerable evidence supporting both a theoretical (Vellend and Geber 2005) and empirical (Table 1.3) relationship between species and genetic diversity, caution needs to be exercised in assuming a positive relationship in all cases (as some papers suggest, e.g. Cleary *et al*. 2006). From available evidence, the extent and direction of the correlation is very changeable depending on the focal species and environment, as well as the way in which the data are collected and analysed. Instead of attempting to find a one-size-fits-all approach to conservation assessment, environments and their constituent species should possibly be considered individually. Patterns do however seem to have emerged in the literature, with a number of studies with different species and environments recording a positive correlation after a recent land-use change or disturbance. Potentially, the same conservation assessment blue print could be used in the future on environments undergoing particular situations. This research is however still in its infancy and further investigation is needed to know when a correlation can be assumed and when it cannot, in order to assure the accuracy of biodiversity conservation assessments in the most economical way possible.

Table 1.3. Summary of 42 empirical studies and models whose focus was to compare genetic and species diversity (most recent - oldest). In the *Species used* column, blue indicates when a plant species has been used. In the conclusion column: pink = positive, blue= no correlation; green = mixed results and yellow =negative.

No.	Species used	Environment	Marker used	Conclusion	Reference
1	7 species from a coastal sand dune plant community (max. sample size of 5 individuals)	240 km of coastline, Atlantic shoreline of southwestern France	2-3 primers AFLP (Amplified fragment length polymorphism)	Positive (negative impact of urbanization)	Frey *et al.* (2016)
2	1 common tree species *Beilschmiedia roxburghiana*	Tropical seasonal rainforest south-western China	10 microsatellite loci	Negative (driven by soil properties)	Xu *et al.* (2016)
3	1 habitat specialist plant, *Saponaria bellidifolia*	Limestone outcrops in the Carpathian Mountains, Southeastern Carpathian Mountains, Romania.	2 primers RAPD (Random Amplified Polymorphic DNA)	Parallel patterns (number of plant communities had a significant effect on both diversity levels)	Csergo *et al.* (2014)
4	1 Red Oak tree species Quercus castanea	Transmexican Volcanic Belt (mountain range), Mexico	14 microsatellite (6 from nuclear DNA and 8 from chloroplast origin)	Positive	Valencia-Cuevas *et al.* (2014)
5	1 Dominant grass species, *Andropogon gerardii*	Experiment - Rainfall manipulated plots	AFLP	No correlation	Avolio and Smith (2013)
6	Aquatic beetle communities (max sample size of 5 individuals)	Lakes, ponds and small streams	mtDNA	Positive	Baselga *et al.* (2013)

No.	Species used	Environment	Marker used	Conclusion	Reference
7	Two freshwater snails 1) *Aplexa marmorat* 2) *Drepanotrema depressissimum*)	Pond network, The Guadeloupe archipelago, French West Indies.	1) 10 2) 8 Microsatellite loci	Positive (due to habitat connectivity)	Lamy *et al.* (2013)
8	Central stonerollers (*Campostoma anomalum*)	Stream fishes in Little Miami River Basin	8 Microsatellite loci	Positive	Blum *et al.* (2012)
9	Dominant plant species (*Ammophila breviligulata*)	Great Lakes Sand dune ecosystem in a manipulated field experiment	3 primers of ISSR (Intersimple sequence repeat) markers	Interaction between SD and GD influenced biomass	Crawford and Rudgers (2012)
10	11 coral reef fish species	Pacific Ocean	mtDNA	Positive	Messmer *et al.* (2012)
11	Widespread high mountain species: 27 spp. in 149 cells in the Alps, and 29 spp. in 30 cells in Carpathians	Vascular flora of the Alps and the Carpathians	AFLPs	No correlation	Taberlet *et al.* (2012)
12	Dominant tree species (*Euptelea pleiospermum*)	Riparian forests of the Shennongjia Mountains in central China	14 nuclear microsatellite loci	Positive correlation found in natural forests but not in disturbed	Wei and Jiang (2012)
13	Tenebrionid beetle communities	Aegean archipelago on 15 islands of different sizes	1 mtDNA and 1 nuclear protein coding marker	Positive	Papadopoulou *et al.* (2011)
14	Clover	Grass – clover dynamics model	Simulated data	Positive	Adams and Vellend (2011)

No.	Species used	Environment	Marker used	Conclusion	Reference
15	3 bat species: Blyth's horseshoe bat (*Rhinolophus lepidus*), the trefoil horseshoe bat (*Rhinolophus trifoliatus*) and the papillose woolly bat (*Kerivoula papillosa*)	In a tropical rainforest, in central peninsular Malaysia, undergoing fragmentation due to oil palm conversion	8-15 microsatellite markers	Positive	Struebig *et al.* (2011)
16	12 species of silicicolous alpine plants	The silicicolous flora of the European Alps	AFLP	Positive	Thiel-Egenter *et al.* (2011)
17	27 species of vascular plants mainly Woody trees	Mediterranean basin (25 counties)	Genetic data came from the literature (isozyme, organelle DNA)	No correlation	Fady and Conord (2010)
18	Eight limestone grassland species	Lab experiment synthesised communities propagated from a small (100m^2) limestone Grassland community	Artificial selection. Comparing 1, 4 and 8 genotypes	SD was increased with GD on deep soil but not shallow soil when compared to single genotype mixtures	Fridley and Grime (2009)
19	Nitrogen-fixing plant (*Daviesis triflora*)	Meta-community of the species-rich southwest Australian flora	11 microsatellite markers	Positive within plant functional groups	He and Lamont (2010)
20	Grass species (*Plantago lanceolata*)	15 Grassland communities, central Germany	AFLP	Positive correlation between populations but not within	Odat *et al*. (2010)
21	Eight salt marsh species (e.g. snail, oyster, crab and the barnacle *Spartina alterniflora*)	Marshes of the Georgia coastal ecosystems	mtDNA (CO1)	Positive (but not significant per species)	Robinson *et al.* (2010)
22	Calanoid copepod species, (*Leptodiaptomus minutus*)	Zooplankton in Boreal lakes	mtDNA (CO1)	Did not correlate	Derry *et al.* (2009)

No.	Species used	Environment	Marker used	Conclusion	Reference
23	Gastropod (*Radix balthica*)	Ain river floodplain in south-eastern France. Meta-community comprising 16 Freshwater Gastropods	AFLP	After a drought alpha SD and GD decreased, whereas beta SD and GD increases.	Evanno *et al.* (2009)
24	Amphipods (monophyletic Gammarus spp.) and Pecos gambusia (*Gambusia nobilis*).	Fish and benthic macro-invertebrate communities in Springs of the northern Chihuahuan Desert	Allozyme	Positive (via drift and isolation)	Sei *et al.* (2009)
25	Grass species (*Anthoxanthum odoratum*)	Park grass experiment	AFLP	Negative or no correlation, in response to resource addition	Silvertown *et al.* (2009)
26	Woddy shrub (*Banksia attenuate*)	Sandplain shrublands of South-west Australia in the functional group made up of five species	11 microsatellite markers	Positive	He *et al.* (2008)
27	Sedge (*Carex curvula*)	Siliceous European alpine grasslands	AFLP	No correlation or Negative correlation	Puşcaş *et al.* (2008)
28	Investigating competition and facilitation	Model	Simulated data	Variety of different diversity-diversity relationships	Vellend (2008)
29	Weedy plant (*Brassica nigra*)	Within a functional group (*B. nigra*, heterospecific forbs, grasses, and bare ground)	Artificial selection. Concentration of allelopathic secondary compound, sinigrin's	Positive - mutual feedback	Lankau and Strauss (2007)
30	Butterfly species (*Drupadia theda*)	Rainforest habitats, Indonesia	5 microsatellite markers	Positive	Cleary *et al.* (2006)

No.	Species used	Environment	Marker used	Conclusion	Reference
31	29 species (5 Plants and algae , 4 Bivalves, 7 Crustaceans, 3 other invertebrates, 6 Fishes and 4 Mammals)	low saline Baltic Sea	Mixture of allozymes, microsatellites, mtDNA; RAPD and haemoglobin (Hb) variation. Microsatellite loci only in 3 mammals, 3 plants and 3 fish species (3-12 loci)	Positive	Johannesson and AndrÉ (2006)
32	Salamanders (*D. ochrophaeus* and *P. jordani*)	Salamanders of the family Plethodontidae across North America	Allozyme richness	Positive	Marshall and Camp (2006)
33	10 shrub to tree species (climax and pioneer species: *Abies alba* L.; *Acer pseudoplatanus* L.; *Betula pendula* Roth; *Carpinus betulus* L.; *Fagus sylvatica* L.; *Picea abies* L.; *Pinus sylvestris* L.; *Rhamnus frangula* L.; Sorbus aucuparia L.; *Tilia cordata* M.)	Forest tree communities, Germany	Starch gel electrophoresis	Positive, negative and not correlated results	Wehenkel *et al.* (2006)
34	Plant communities model	Simulated	Simulated	Positive, negative and not correlated. (though most common was weak - moderate positive correlation)	Vellend (2005)
35	Meadow buttercup (*Ranunculus acris*)	Grassland communities, central Germany	AFLP	Positive	Odat *et al.* (2004)

No.	Species used	Environment	Marker used	Conclusion	Reference
36	Forest herbs (*Trillium grandiflorum*)	Forests, New York state	Allozymes, chloroplast DNA RFLP and microsatellites	Positive (weak)	Vellend (2004)
37	11 long-lived herbaceous plant species (*Festuca ovina*, *Koeleria macrantha*, *Helictotrichon pratense* and *Briza media* (grasses), *Carex flacca*, *C. panicea* and *C. caryophyllea* (sedges) and *Leontodon hispidus*, *Succisa pratensis*, *Campanula rotundifolia* and *Viola riviniana* (forbs)	Manipulated field experiment in Ancient limestone pasture, Derbyshire	Artificial selection. Three levels of genetic diversity	Positive	Booth and Grime (2003)
38	14 datasets including Birds, Mammals, Reptiles and plants	Between 5-14 Islands (2 Canary Islands, 3 Galápagos, Gulf of California, Caribbean, Baja California, 2 California channel, British Isles, Gulf of Maine, Great Barrier Reef and West Indies)	Allozyme, minisatellite and mtDNA	Positive	Vellend (2003)
39	Spruce and beech	Norway spruce-beech forest communities	Isozyme loci	No correlation	Hosius *et al.* (2001)
40	Salamander (*Desmognathus fuscus*)	Co-occurring conspecifics North America	unknown	Negative relationship	Karlin *et al.* (1984)
41	Drosophila	Hawaiian Islands	Enzyme polymorphism	No correlation after reanalysis by (Gotelli and Colwell 2001)	Johnson (1973)
42	Iguana (*Uta stansburiana*)	Species of Iguanidae among islands in the Gulf and California	unknown	Positive	Soule and Yang (1973)

Chapter 2 - Methods

Highlights

- Describing the sites where *Amphinemura s*ulcicollis, *Isoperla grammatica* and *Baetis rhodani* were collected.
- Investigating cryptic diversity within each species and designing a diagnostic tool to remove cryptic species within *B. rhodani*.
- Creating bioinformatics library for each species, described in Macdonald *et al.* (2016a) and Appendix B.
- Development of computer program *PrimerPipeline* to find microsatellite loci and design primers from next generation sequencing data (available to download from its website here: http://www.scrufster.com/primerpipeline/default.htm) and the manual is available in Appendix C)
- Developing novel microsatellites for each species, published in Macdonald *et al.* (2016b) and available in Appendix E.

2.0 Abstract

This chapter contains methods that are central to the rest of the thesis as it describes everything that had to be completed before correct samples could be genotyped for the three target species, *Amphinemura s*ulcicollis, *Isoperla grammatica* and *Baetis rhodani*. The flow of the chapter follows the order in which tasks were completed. Firstly samples were collected from a large number of sites, and then final sites for each species were chosen from these for genetic analysis. Afterwards, DNA was extracted from samples from these chosen sites, and mitochondrial barcoding was implemented to verify correct identification and detect cryptic diversity within each species. For *B. rhodani* it was necessary to develop a diagnostic tool using restriction enzymes to differentiate *B. rhodani* between morphologically identical cryptic species. In order to address the aims of subsequent chapters, species-specific microsatellite markers were developed. Here, the bioinformatics pipeline which lead to the identification of these microsatellites is described, involving the development of the software *PrimerPipeline*, microsatellite testing and at last choosing the final microsatellites that would be used in the analysis of subsequent chapters. Finally the completion of the genotyping dataset for each species is described, including how high levels of accuracy were ensured during the scoring and binning stages. It became apparent that all three species are characterised by high levels of heterozygote deficiency than would be expected when compared to populations in Hardy-Weinberg Equilibrium, therefore, also in this chapter the issue of putative null alleles are discussed for these chosen newly developed markers.

2.1 Site Selection

2.1.1 All Sites Visited

Sixty sites were chosen for this study, across 15 different catchments in upland Wales, covering an area of 145 by 90km (Figure 2.1). These sites were deemed characteristic of upland Wales (from past River Habitat Survey data), and represent a broad range of different land-uses and altitudes (ranging from 111m to 440m) (Appendix A: Table A1). Fifty three of the chosen sites were also site-locations of a much more extensive research program funded by the Natural Environment Research Council (NERC): DURESS (Diversity of Upland Rivers for Ecosystem Service Sustainability: http://nerc-duress.org). At these sites DURESS is investigating food webs, fish production and genetic diversity at various trophic levels (biofilm, invertebrates, fish and birds). Of the remaining seven sites, six were in an area surrounding Llyn Brianne where ecological data have been collected over the past 25 years (Durance and Ormerod 2007).

Primary sampling (sites 1 - 60) was completed between 11.05.12 and 05.06.12. In 2013, 27 of these sites were re-visited (sites 93 – 119; though they are the same locations they were given new site numbers to easily distinguish between years), between 9.05.13 and the 21.05.13 (note sites 61-92 were terrestrial sites subsequently dropped from this study). These latter sites were chosen for resampling, not only because of successful initial sampling in 2012, but also as they represented all 15 catchments, a large geographical range and encapsulated the multiple land-uses and altitudes mentioned above. All samples were labelled with the site code number, followed by the initial of genera name (e.g. A = *Amphinemura sulcicollis*, B= *Baetis rhodani* etc.), and the individual sample number. In this way, Sample “93A4” indicated Afon Colwyn (site 93), *A. sulcicollis,* of which this was the fourth individual of the species at that site.

A large number of sites were visited to ensure sufficient samples (Figure 2.1). Sites failed for three main reasons: 1) insufficient number of the required species; 2) misidentification of species in the field; and 3) the presence of cryptic taxa which would have to be dropped after preliminary genetic analysis. The lack of knowledge on the genetic structure of these species (especially for *A. sulcicollis* and *I. grammatica*) throughout Wales also made

selecting a relatively large number of sites necessary. Full information on the sites and the samples collected from them is included in Appendix A, Table A1.

2.1.2 Choosing the final sites used for genetic analysis

It was not possible (due to time and financial constraints) to analyse all sites using genetic methods, therefore a subset of final sites had to be chosen for each species. First of all, failed sites were excluded; the success of collecting sufficient numbers of each species differed considerably between sites. For example, *Baetis rhodani* are known to be acid-sensitive so tended to be absent from more acidic sites where coniferous forests were prevalent (Kowalik and Ormerod 2006).

To gain preliminary knowledge of the genetic structure of these species throughout Wales, 10 sites were analysed using molecular markers (from 2013) for each species, ensuring that the sites represented a wide range of geographical locations and covered many different catchment areas. More than one site was sampled within the Tywi catchment, thus enabling both intra- and inter-catchment studies to be carried out. Following preliminary analysis, additional sites were selected. For example, preliminary data for *Amphinemura sulcicollis* suggested possible isolation in North Wales. As a result, all north Wales sites with an appropriate number of samples were analysed. Example sites for each species that were sampled in 2012 and 2013 were analysed for both years to investigate inter-annual variation. Inter-annual comparisons also provided validation of the sampling method. Sites that contain 10 samples or more were considered 'full' sites; occasionally a site containing less than 10 samples was used (four sites within *A. sulcicollis* and one within *I. grammatica*). When analysis was performed per site, these incomplete sites were not reported however all samples were used when looking at combined datasets or groups within the data (e.g. when running individual-based algorithms implemented in the coalescent software).

2.1.2.1 Amphinemura sulcicollis

Seventeen sites, comprising a total of 278 individuals, were chosen for *Amphinemura sulcicollis*. These sites represented 10 catchments and altitudes, ranging from 190 to 440 m (Table 2.1a). Thirteen of these sites are considered full sites, four of which were used to compare 2012 and 2013 datasets (at Afon Pistyll and Brefi) and three full sites were selected within the Tywi catchment for intra-catchment assessment (Figure 2.2). Four sites comprised eight or fewer samples, and three of these were due to misidentification of *A. sulcicollis* in the field. *Amphinemura sulcicollis* and *A. standfussi* co-existed in the same locations and typically cannot be separated without the aid of a high resolution microscope. In the case of very small-sized specimens such as those collected in the more northern sites (e.g. Afon Fechan and Nant y Gwryd), genetic analysis was necessary to enable identification (e.g. site 4, contained just one correctly identified sample while 21 samples were found to be *A. standfussi*).

2.1.2.2 Isoperla grammatica

Thirteen sites comprising a total of 237 individuals were chosen for *Isoperla grammatica*, representing altitudes ranging from 160 to 310 m. Sites represented eight different catchments including four sites within the Tywi catchment, and two samples comparing 2012 with 2013 within site CI1 (Figure 2.3, Table 2.1b).

2.1.2.3 Baetis rhodani

Baetis rhodani was investigated at 10 sites, yielding a total of 186 individuals (Figure 2.4). The sites represented eight catchments, and altitudes ranging from 210 to 400 m (Table 2.1c). The data also contain two sites within the Tywi catchment and site Brefi was used to compare 2012 and 2013.

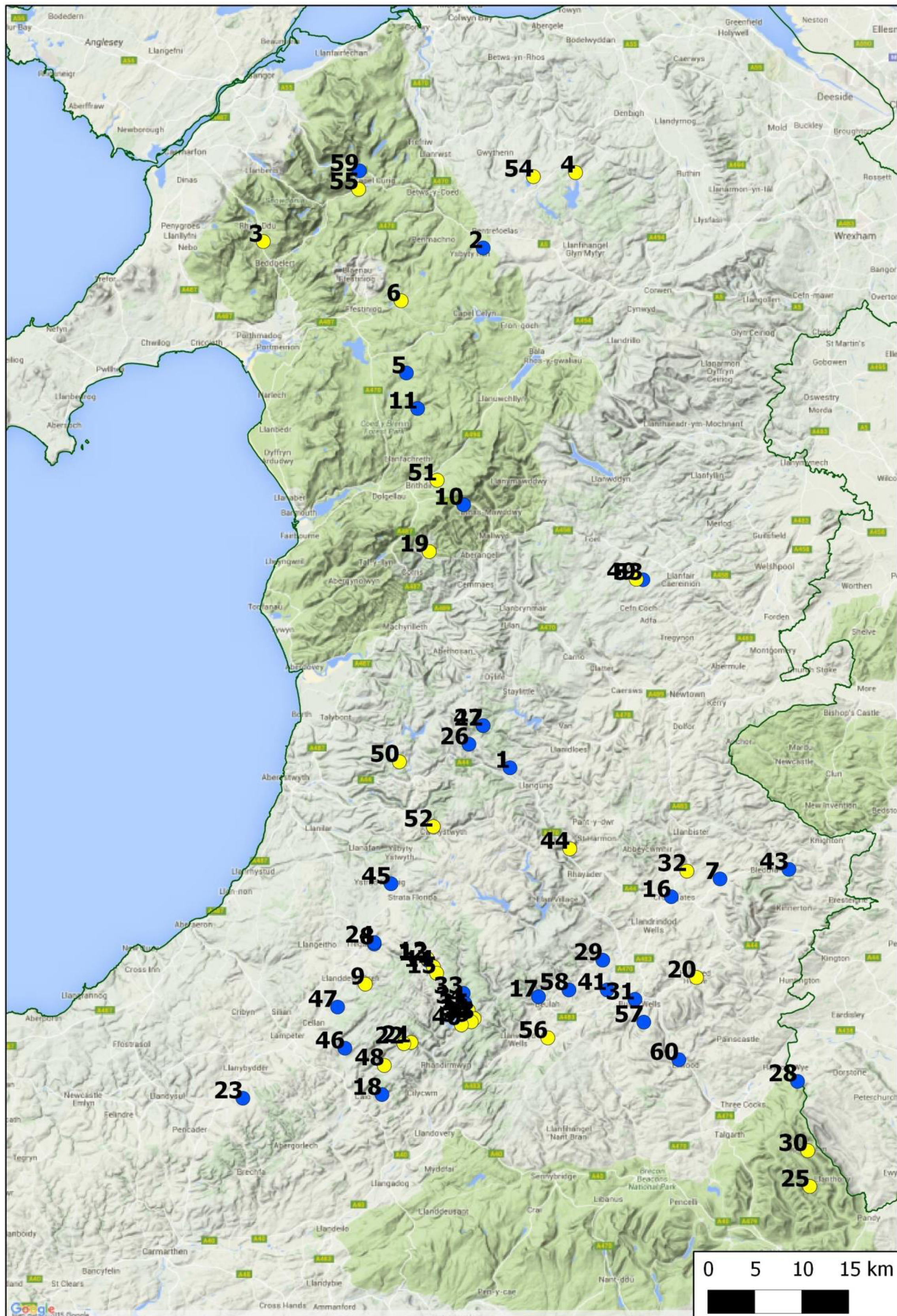


Figure 2.1. The location of all 60 freshwater invertebrate sampling sites throughout upland Wales. Blue circles represent sampling in 2012, yellow circles represent sampling in 2012 and 2013. Site codes numbered 1 to 60 which correspond with Table A1 in Appendix A.

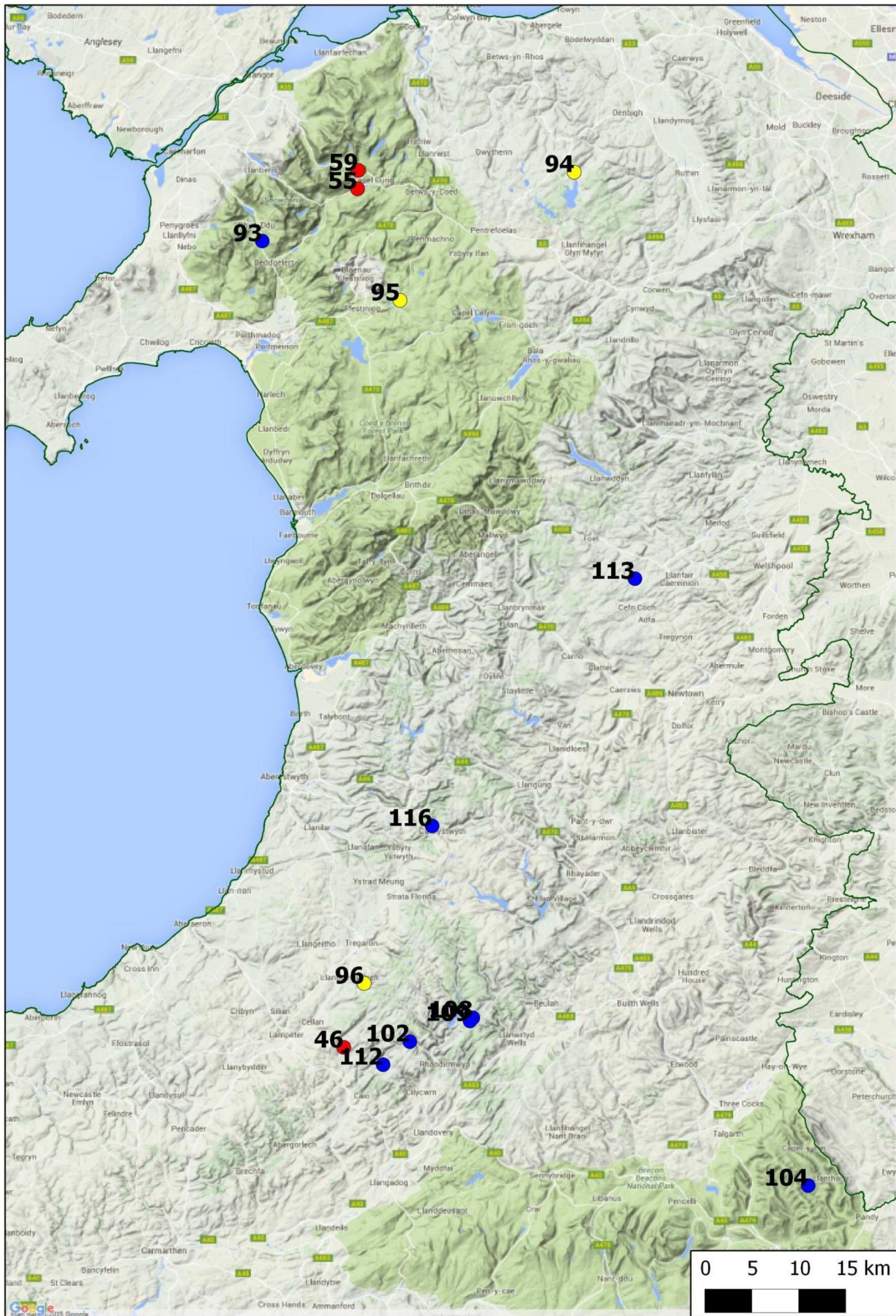


Figure 2.2. The 14 sites from which *Amphinemura sulcicollis* samples were analysed. Blue circles represent sampling in 2013, Red circles represent sampling in 2012, yellow circles represent sampling in 2012 and 2013. All are labelled with their site code (See Table 2.1a).

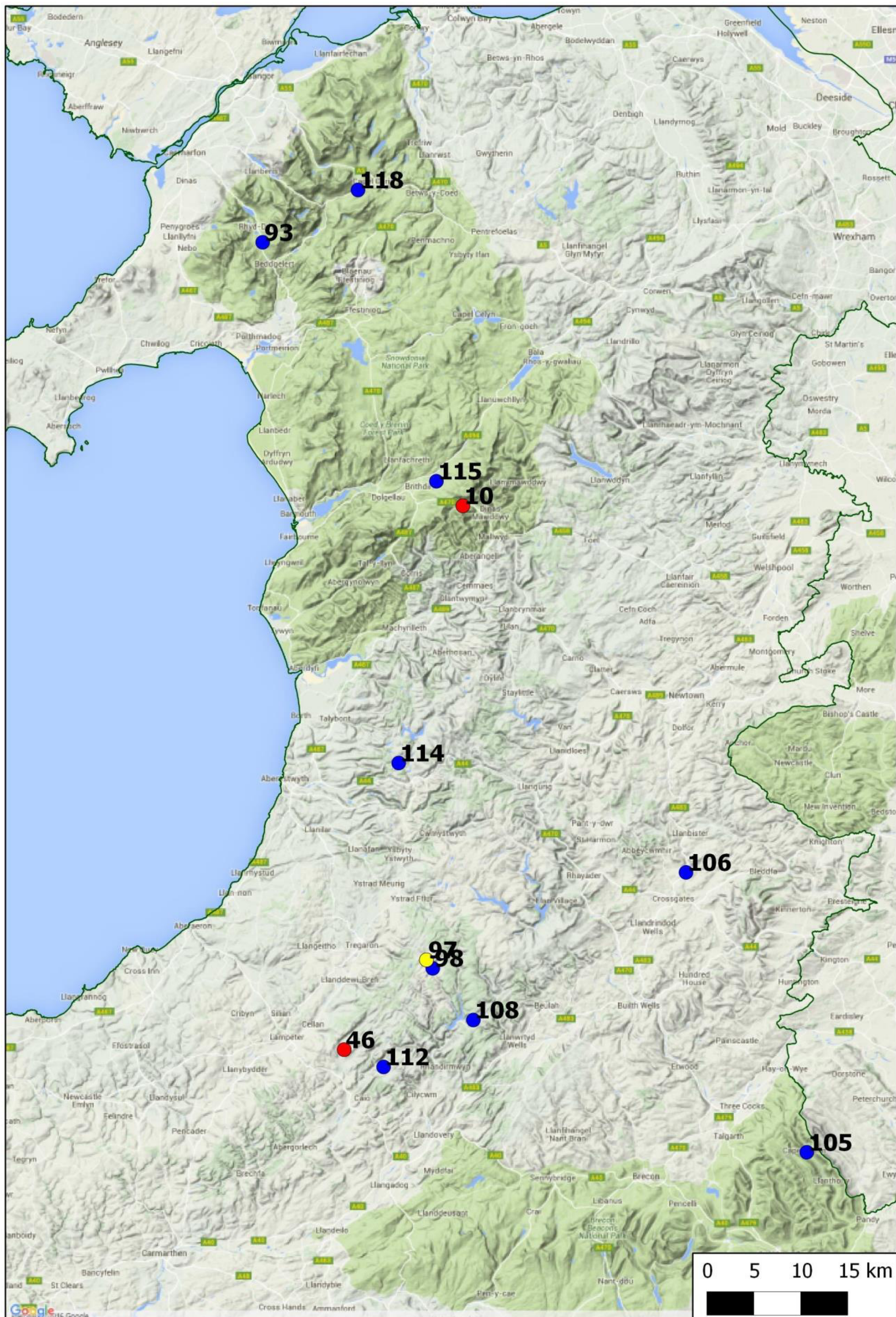


Figure 2.3. The 12 sites from which *Isoperla grammatica* samples were analysed. Blue circles represent sampling in 2013, Red circles represent sampling in 2012, yellow circle represents sampling in 2012 and 2013. All are labelled with their site code (See Table 2.1b).

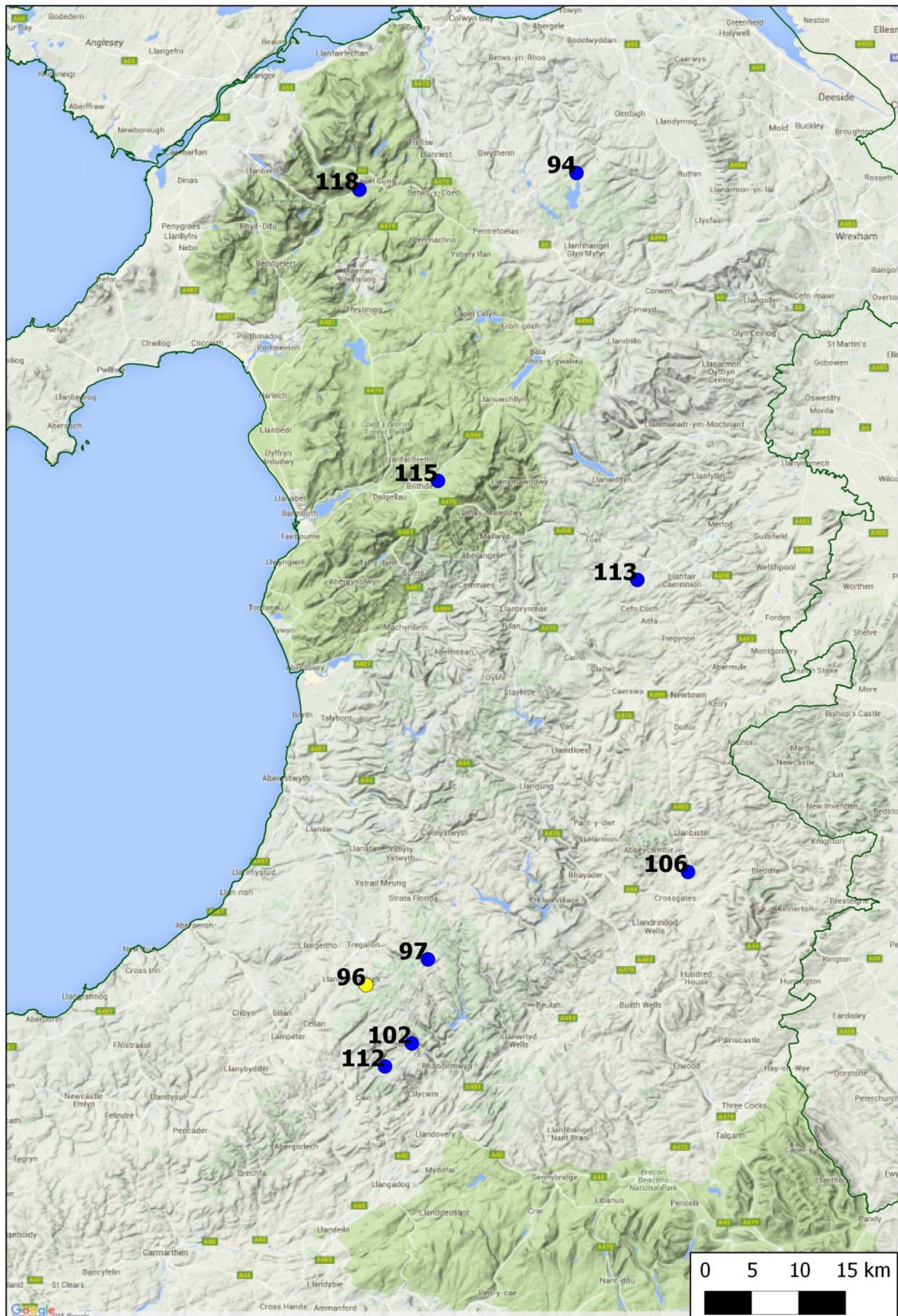


Figure 2.4. The 9 final dataset sites from which *Baetis rhodani* samples were analysed. Blue circles represent sampling in 2013, yellow circle represents sampling in 2012 and 2013 and are labelled with their site code (see Table 2.1c).

Table 2.1a-c. Details of the final sites used for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. Sites are ordered approximately from south to north Wales, and by catchment.

Table 2.1a

N	Site code	Site name	Year	Sample size	Altitude	Catchment
1	104	Grwyne Fawr	2013	20	374	Usk
2	109	Ll7	2013	18	330	Tywi
3	108	Ll6	2013	20	326	Tywi
4	102	Gl1	2013	20	216	Tywi
5	112	Nant Dar	2013	20	220	Cothi
6	46	Nant Clawdd	2012	5	210	Cothi
7	96	Brefi	2013	19	210	Teifi
8	9	Brefi	2012	20	210	Teifi
9	116	Nant Peiran	2013	20	280	Ystwyth
10	113	Nant Gelli Gethin	2013	20	288	Severn
11	95	Afon Pistyll	2013	22	440	Dwyryd
12	6	Afon Pistyll	2012	21	440	Dwyryd
13	94	Afon Fechan	2013	5	400	Dee
14	4	Afon Fechan	2012	1	400	Dee
15	55	Nant y Gwryd	2012	8	210	Conwy
16	59	Upper Llugwy	2012	20	262	Conwy
17	93	Afon Colwyn	2013	19	190	Glaslyn

Table 2.1b

N	Site code	Site name	Year	Sample size	Altitude	Catchment
1	105	Honddu at Capel	2013	15	318	Wye
2	106	Ithon at Llandewi	2013	20	235	Wye
3	108	Ll6	2013	20	326	Tywi
4	12	Cl1	2012	20	370	Tywi
5	97	Cl1	2013	17	370	Tywi
6	98	Cl4	2013	20	364	Tywi
7	112	Nant Dar	2013	20	220	Cothi
8	46	Nant Clawdd	2013	5	210	Cothi
9	114	Nant Glan dwr	2013	20	285	Rheidol
10	115	Nant Helygog	2013	20	215	Mawddach
11	10	Cerist (Afon)	2012	20	160	Dyfi
12	118	Nant y Gwryd	2013	20	210	Conwy
13	93	Afon Colwyn	2013	20	190	Glaslyn

Table 2.1c

N	Site code	Site name	Year	Sample size	Altitude	Catchment
1	112	Nant Dar	2013	19	220	Cothi
2	102	GI1	2013	21	216	Tywi
3	97	CI1	2013	19	370	Tywi
4	96	Brefi	2013	18	210	Teifi
5	9	Brefi	2012	11	210	Teifi
6	106	Ithon at Llandewi	2013	20	235	Wye
7	113	Nant Gelli Gethin	2013	18	288	Severn
8	115	Nant Helygog	2013	19	215	Mawddach
9	94	Afon Fechan	2013	20	400	Dee
10	118	Nant y Gwryd	2013	21	210	Conwy

2.2 Genetic Methods

2.2.1. Sample collection

Larval specimens of *Amphinemura sulcicollis*, *Isoperla grammatica* and *Baetis rhodani* were collected by kick sampling (from sites described in Section 2.1); samples were sorted and identified to genus level on site (e.g. *Baetis* for *Baetis rhodani*) and stored in absolute ethanol. The aim was to collect at least 20 individuals per species, per site, although often many more were collected to anticipate misidentification in the field (please see Appendix A, Table A1 for number of samples collected per site). In the laboratory each larva was identified to species level using Elliott *et al.* (1988) and Hynes (1977) and stored in absolute alcohol (2012 samples were also stored at -80 ^{o}C).

2.2.2 Extraction methods

Four different methods for extracting DNA from tissue samples were used due to cost and availability constraints (see Appendix F, Table F1, F4 and F7, for information on the method used for each sample):

1) The High Pure PCR Template Preparation Kit for blood and tissue (Roche Diagnostics GmbH Mannheim, Germany) was used for preliminary experiments and initial extractions,

following the manufacturer's instructions; 27, 2 and 175 samples were extracted using this method for *A. sulcicollis*, *I. grammatica* and *B. rhodani,* respectively.

2) The DNeasy 96 blood and tissue kit (Qiagen, Hilden, Germany), following the manufacturer's instructions, was used to increase efficiency (as 96 samples can be extracted at one time with this method); 173 and 190 samples were extracted using this method for *A. sulcicollis* and *I. grammatica,* respectively.

3) The Chelex® 100 resin (Bio-Rad, Richmond, CA) (Walsh *et al.* 1991) was tested on a small number of samples because it was a very cost-effective; 13 and 5 samples were extracted using this method for *A. sulcicollis* and *I. grammatica,* respectively).

4) Gentra Puregene Core Kit A, DNA purification from tissue (Qiagen, Hilden, Germany) following the manufacturer's instructions, was used for the last samples extracted because the Chelex method had varying success; 65, 40 and 11 samples were extracted using this method for *A. sulcicollis*, *I. grammatica* and *B. rhodani*, respectively.

To ensure that the extraction method did not affect the results, a number of samples were extracted twice using different methods and their genotypes compared to ensure there was no major difference in DNA quality and that the genotype was identical. Additionally after genotyping was completed, samples were chosen representing different sites for each extraction method (where more than 10 samples were available), for each species and Na, Ho and He were compared using GenAlEx (Peakall and Smouse 2012) (see Section 2.4.2 for methods). *A. sulcicollis* was compared at all four extraction methods (1= 19 samples from 11 sites, 2= 20 samples from 10 sites, 3=13 samples from two sites, 4= 20 samples from four sites). *I. grammatica* and *B. rhodani* were both compared at two different extraction methods (2= 20 samples from 10 sites, 4= 20 samples from two sites for *I. grammatica*; 1= 20 samples from nine sites, 4= 11 samples from one site for *B. rhodani*). All three species showed narrow ranges between different extraction methods, therefore indicating that extraction method had limited impact on the results (Na = 8.3 – 11.2, 19.6 – 20.6, 8.5 – 11.2; H_o = 0.588 – 0.675, 0.683 – 0.713, 0.603 – 0.640; H_e = 0.801 – 0.806, 0.937 – 0.949, 0.812 – 0.814 for *A. sulcicollis*, *I. grammatica* and *B. rhodani* respectively).

Past studies with *Baetis rhodani* (Williams 2003) identified extraction method specific issues with polymerase chain reaction (PCR) inhibitors, however these challenges were not encountered in this study. Therefore, throughout this study, whole (if the individual was small) or partial (usually lower half of the thorax and abdomen if individual was larger, so that the sample could be extracted again, if needed) specimens could be used. Invertebrate legs were not used, as often legs would become detached during storage or confused with the appendages of other organisms.

2.2.3 DNA Barcoding

One of the first challenges before the main genetic analysis was correctly identifying samples to species; especially before any Next Generation Sequencing (NGS) was performed for primer development. Each larva was identified to species level based on morphological features using Elliott *et al.* (1988) and Hynes (1977). However, where a sample was damaged (e.g. legs missing) or very small, misidentifications were expected. Cryptic species (Pfenninger and Schwenk 2007) could also confuse patterns of genetic data.

Cryptic diversity has been well documented in *B. rhodani* (Williams *et al.* 2006) but was unknown in *A. sulcicollis* and *I. grammatica*. The mitochondrial cytochrome c oxidase I gene (mtCOI) has been used in a wide variety of similar studies (Guarnizo *et al.* 2015; Viñas *et al.* 2015) as a DNA barcoding tool. Its interspecific variation has been shown to be much greater than its intraspecific variation for most taxa, and can discriminate between cryptic species (Meyer and Paulay 2005; Viñas *et al.* 2015). The classical barcoding region (Hebert *et al.* 2003) within mtCOI has been very successful within other invertebrate species using consensus PCR primers (Folmer *et al.* 1994; Harper *et al.* 2005; Viñas *et al*. 2015). The aim of this preliminary study was to: 1) validate species recognition based on morphological identification; 2) investigate the diversity within the samples; 3) discover whether cryptic studies might be an issue in any of the species studied; 4) if so, develop a method to separate cryptic species from the rest of the samples; and to 5) choose a sample for each species to perform NGS after ensuring the correct identification has been made.

Table 2.2. PCR reagents for the mtDNA PCR using DNA barcoding primers (LCO1490: GGTCAACAAATCATAAAGATATTGG and HCO2198: TAAACT TCAGGGTGACCAAAAAATCA, Folmer *et al.* (1994)). PCR was carried out in a final volume of 20 µl.

PCR reagent	Final concentrations / quantity	Supplier
Go Taq Buffer x5	4 µl	Promega (Maddison, WI, USA)
$MgCl_2$	3.75 mM	Promega (Maddison, WI, USA)
dNTPs	0.5 mM	Promega (Maddison, WI, USA)
Each Primer	0.5 µM	Sigma (Gillingham, UK)
Bovine Serum Albinum (BSA) x100	0.25 µl	New England Biolabs (Ipswich, MA, USA)
Taq DNA polymerase	0.0625 µl	Promega (Maddison, WI, USA)
DNA	1 µl	

Table 2.3. PCR conditions for the mtDNA PCR using Folmer *et al.* (1994) primers (LCO1490: GGTCAACAAATCATAAAGATATTGG and HCO2198: TAAACT TCAGGGTGACCAAAAAATCA). PCR was carried out in a final volume of 20 µl.

	Temp (°C)	Duration	
Initial denaturation	94	5 minutes	
	94	30 seconds	
	54	30 seconds	x 35
	72	1 minute	
Final extension	72	10 minutes	

2.2.3.1 Sequencing and analysis methods

Standard DNA barcoding primers from Folmer *et al.* (1994), were used (Table 2.2, Table 2.3). Samples were sent to Eurofins (Eurofins MGW Operon, Ebersberg, Germany) for purification and Stanger sequencing. Samples were initially sequenced in the forward and reverse

direction and a consensus used (those samples with 'F&R' in the name), however the sequence quality of the samples was good enough that subsequent samples were sequenced in the forward position only to make the preliminary experiment more cost effective. Sequences were edited by eye using Sequencher v4.9 (Gene Codes) and aligned using ClustalW in Bioedit v7.0.5 (Hall 1999). Sequences were trimmed to 624 bp, 627 bp and 556 bp, for *A. sulcicollis*, *I. grammatica* and *B. rhodani* samples, respectively. The identities of the sequences were checked using a standard nucleotide BLAST (Basic Local Alignment Search Tool) and compared with Genbank's online database (Benson *et al.* 2005). A 1000 bootstrap consensus neighbour joining (NJ) tree was created in Mega v 6.0 (Tamura *et al.* 2013) for each species and a median joining network was then made using Popart v1.7 (Leigh and Bryant 2015) using samples from the main haplogroup only for each species. The NJ tree in MEGA uses a 'Maximum composite likelihood' distance method, estimated under a combination of the Jones–Thornton–Taylor (JTT) model for amino acid sequences and the Tamura and Nei (1993) model (Tamura *et al.* 2011). Median-Joining networks use the 'Hamming distance' which is a simple method which counts the number of character differences between two sequences (Bandelt *et al.* 1999).

Thirty-two *A. sulcicollis* samples from 13 different sites (two individuals per site, apart from sites 109 and 95 which both have five) were used to make the NJ tree. Of these, 31 samples were used to construct the network. Twenty-nine *I. grammatica* samples from 14 sites were used to make the NJ tree and network for this species and one *A. sulcicollis* sample was used as an outgroup. Sixty-five *B. rhodani* individuals were sequenced from 19 different sites. Focus was placed on site 14 because preliminary results showed that cryptic species were present. Twelve sequences from Genbank of other members of the *Baetis* genus were used to compliment the analysis and act as outgroups (*B. tricaudatus, B. macani, B.hudsonicus, B.vernus, B.liebenauae*, as well as *B. rhodani*). Thirty-four of those samples (haplogroup 1) were used to construct the network. Sequences available at Genbank, Accession numbers KU955863-KU955988, see Appendix A, Table A2.

2.2.3.2 Results and discussion

All three species showed similar patterns of diversity, with all or the majority of the samples comprising a single haplogroup. When using the main haplogroup (red circles in Figure 2.5a, blue circles in Figure 2.5b, and green circles in Figure 2.6) to create a median joining network, all three species showed a dominant haplotype, representing the majority of samples (Figure 2.7a-c). The majority of the haplotypes in each species are only separated from each other by one base mutation. However, these results could be due to the barcoding markers used, as they are designed to maximise inter specific diversity, therefore the similarities could be due to ascertainment bias.

2.2.3.2.1 Amphinemura sulcicollis

There was little variation among *A. sulcicollis* samples, with almost all comprising one haplogroup (Figure 2.5a). Although there were only seven *A. sulcicollis* sequences available on Genbank with which to compare, all samples apart from one aligned successfully; suggesting the identification to be correct (apart from 94A1). Sample 94A1, however, aligned with *A. standfussi*, indicating misidentification of this sample. Nine haplotypes were identified within the correctly identified samples; H1 = 20 individuals; H2 = four; with seven haplotypes associated with one sample (H3-H9) (Figure 2.7a). There was no indication of cryptic species within the samples. When the *A. standfussi* sample was genotyped, the microsatellites designed for *A. sulcicollis* did not amplify, excluding the risk of misidentification. Sample 95A1 was chosen for NGS, not only as having been verified as the correct species, but also because of the quality and quantity of its DNA.

2.2.3.2.2 Isoperla grammatica

There was no evidence of the presence of cryptic species in *I. grammatica*. There were no mtCO1 sequences of this species on Genbank (just two mRNA sequences); however, there was no obvious branching in the NJ tree (Figure 2.5b) and, with *I. grammatica* being the most prevalent *Isoperla* species present at the sites visited, no obvious species could be

mistaken for it. The samples contained 11 haplotypes; H1 = eight individuals; H2 = six; H3 = four; H4 = three; H5 = two, with six haplotypes associated with one sample (H6-H11; Figure 2.7b). Individual 96I1 was chosen for NGS as it had the highest quality and concentration of DNA.

2.2.3.2.3 Baetis rhodani

The vast majority of the analysed samples were identified as *B. rhodani* when compared to sequences in Genbank, however, a number of cryptic species were identified. The samples split into three haplogroups, the large majority of the sites were in haplogroup 1 (green circles Figure 2.6), a small cluster branched off, forming haplogroup 2 (pink squares Figure 2.6), and haplogroup 3 was present at site 14 only (CI4 in Llyn Brianne, black triangles Figure 2.6). Morphological examination with a light microscope at x40 magnification allowed identification of marginal spines occurring intermittently between hairs on the gills of all samples sequenced, a feature only present in *B. rhodani* (Elliott *et al.* 1988). When sequenced and blasted in Genbank, however, the samples in haplogroup 3 did not align with other *B. rhodani* sequences. Separated clearly from the other samples (Figure 2.6), they appear more closely related to *Baetis liebenauae*, a species that does not occur in Britain, and thus could represent a previously undetected taxon.

For the samples that were identified as *B. rhodani*, haplogroup 1 (34 samples) comprised 13 haplotypes; H1 =10; H2 = seven; H3 and H4 = three; H5 and H6 = two; and one sample in the other seven haplotypes (H7-13, Figure 2.7c). Individual 102B3 was selected for NGS for its DNA quality and yield.

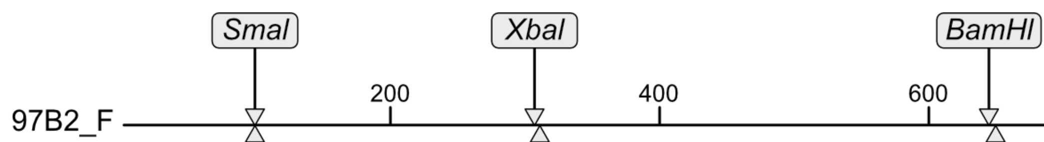
A method of separating haplogroup 1 from the rest of the samples was developed thus avoiding allowing cryptic species into the genotyping dataset for this species. This was only necessary for *B. rhodani* as the other two species consisted of samples of only one haplogroup.

2.2.3.3 Cryptic species discrimination tool for *B. rhodani*

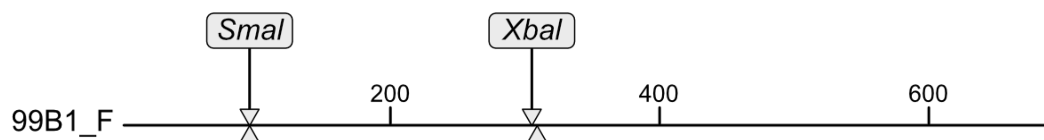
To screen the samples and determine which belong to haplogroup 1 (Figure 2.6 & Figure 2.7c) without having to sequence them, restriction sites (RFLP) within the sequences were examined using CLC Genomics Workbench v6.5.1 (http://www.clcbio.com).

Two restriction enzymes, '*Xba*1' and '*Bam*H1', were used in combination to determine which haplogroup samples belonged to, when visualised on an electrophoresis gel. Enzyme '*Xba*1' (New England BioLabs) was able to cut haplotype 1 and 2, enabling any haplotype 3 samples to be identified and excluded. Enzyme '*Bam*H1' (New England BioLabs) was able to discriminate between haplogroup 1 and haplogroup 2, because it cut the sequences in haplogroup 1 but not haplogroup 2. Therefore, using these two enzymes all cryptic individuals could be removed using the restriction enzyme mix described in Table 2.4.

Example of haplogroup 1:



Example of haplogroup 2:



Example of haplogroup 3:

Table 2.4. Details of reagents of restriction enzyme mix used to determine haplogroup of *Baetis rhodani* sample. Carried out in a final volume of 20 µl and incubated at 37°C for one hour.

Reagent	Quantity	Supplier
PCR product (Table 2.2 and Table 2.3)	10µl	
1X NE buffer 3	2.5 µl	New England Biolabs (Ipswich, MA, USA)
BSA	0.5 µl	New England Biolabs (Ipswich, MA, USA)
BamH1 enzyme	0.5 µl	New England Biolabs (Ipswich, MA, USA)
Xba1 enzyme	0.5 µl	New England Biolabs (Ipswich, MA, USA)

After incubation the products were electrophoresed for 1 hour and 45 minutes at 120V on a 3% ethidium-bromide stained agarose electrophoresis gel before being visualised under UV light (Figure 2.8). Positive controls (known haplogroup 1 sequences from Figure 2.6); negative controls (known sequences from haplogroups 2 or 3, Figure 2.6) and a 100 bp ladder were also ran alongside the products for comparison.

If the sample matched the positive control's pattern, the sample was included; if, however, the sample matched the negative control, it was excluded. Occasionally, the sample had a different pattern; in this case, it was investigated further by examining the restriction sites.

This procedure was completed for all samples; through this screening 20 *B. rhodani* samples were excluded leaving 186 samples on which to carry out microsatellite analysis. The method was verified by sequencing a few samples that had been included and excluded; this yielded supporting results allowing the adopted method to be considered successful.

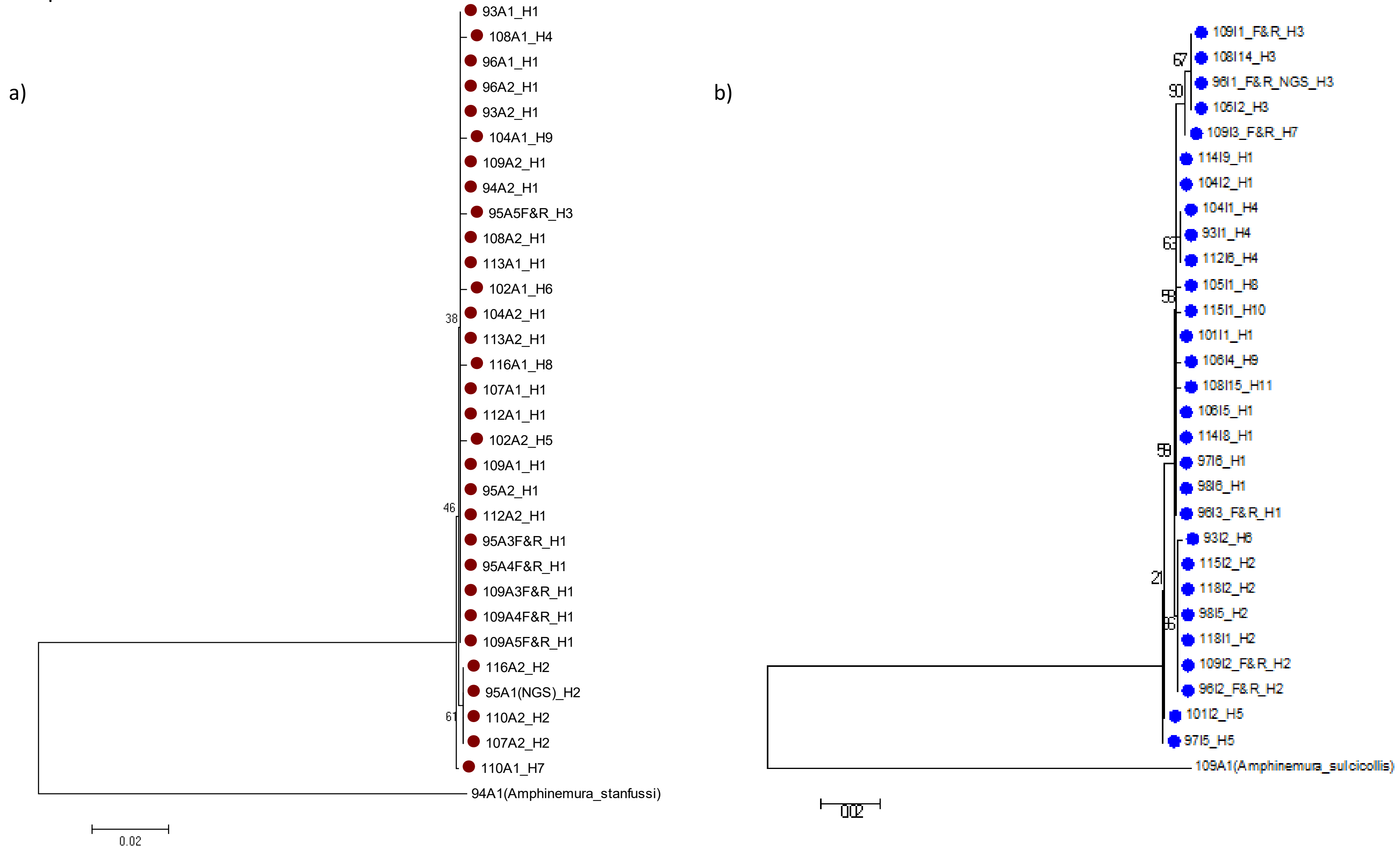


Figure 2.5. A neighbour joining tree of a) *Amphinemura sulcicollis* and b) *Isoperla grammatica* samples from 13 and 14 sites, respectively. 'F&R' = sequenced forward and reverse. 'H1-11' represents the haplotype the sample belongs to. Circle's = haplogroup 1 (used for network diagram)**.** GenBank accession numbers between KU955863-KU955923 (Appendix A, Table A2).

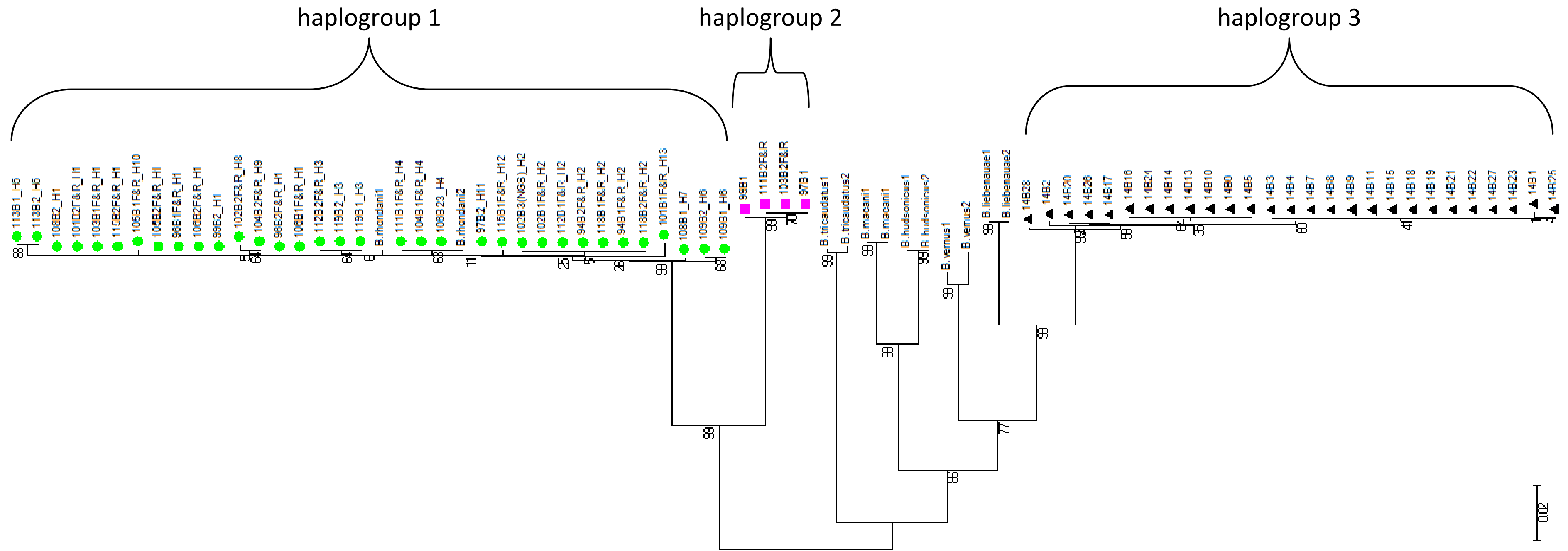


Figure 2.6. A neighbour joining tree containing samples of *Baetis rhodani* from 19 sites. Within the name: 'F&R' = forward and reverse was sequenced, and 'H1-13' represents which haplotype the sample belongs to. Green circle = haplogroup 1 (used for network diagram); Pink square = haplogroup 2, Black triangle = haplogroup 3; GenBank accession numbers: KU955924-KU955988 (Appendix A, Table A2).

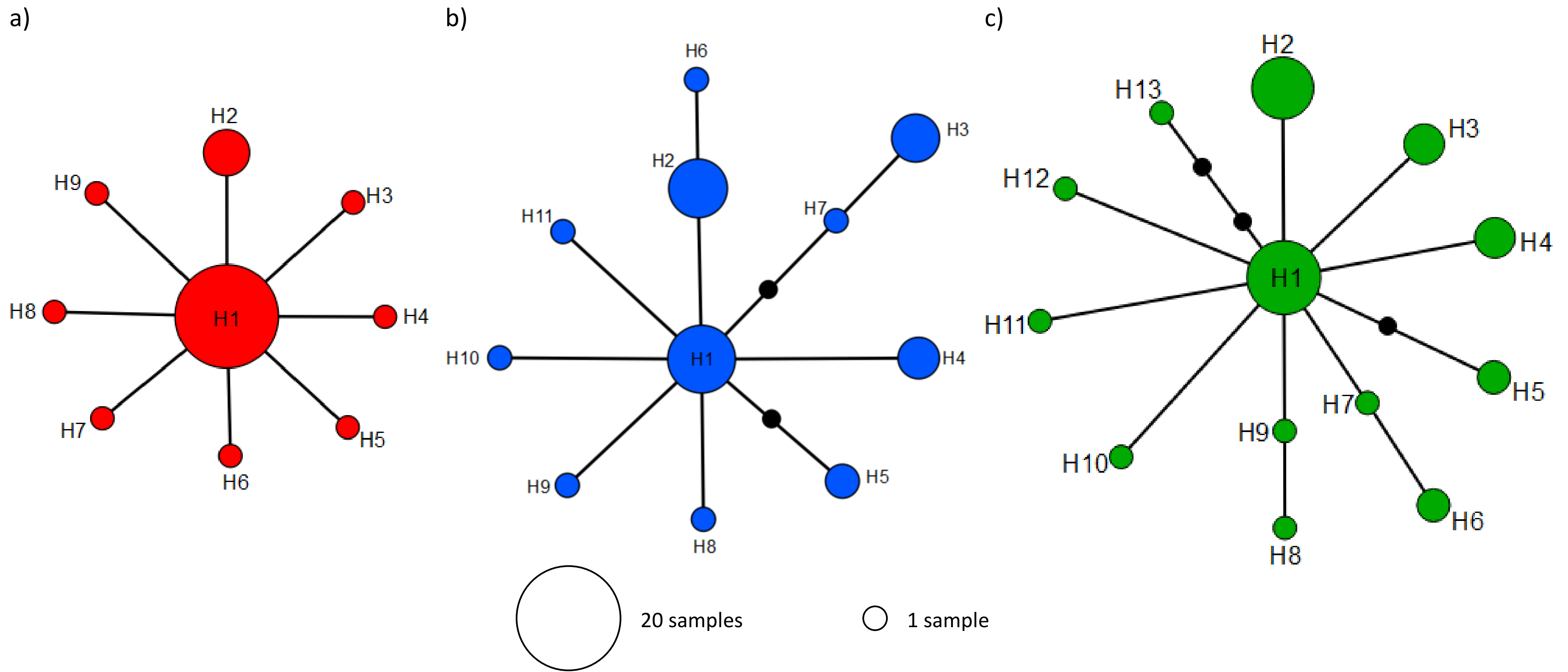


Figure 2.7. Median Joining Network of a) *Amphinemura sulcicollis*, b) *Isoperla grammatica*, and c) *Baetis rhodani*, showing the diversity of haplotypes within the main haplogroup of each species. Black dots represent single mutations between haplotypes.

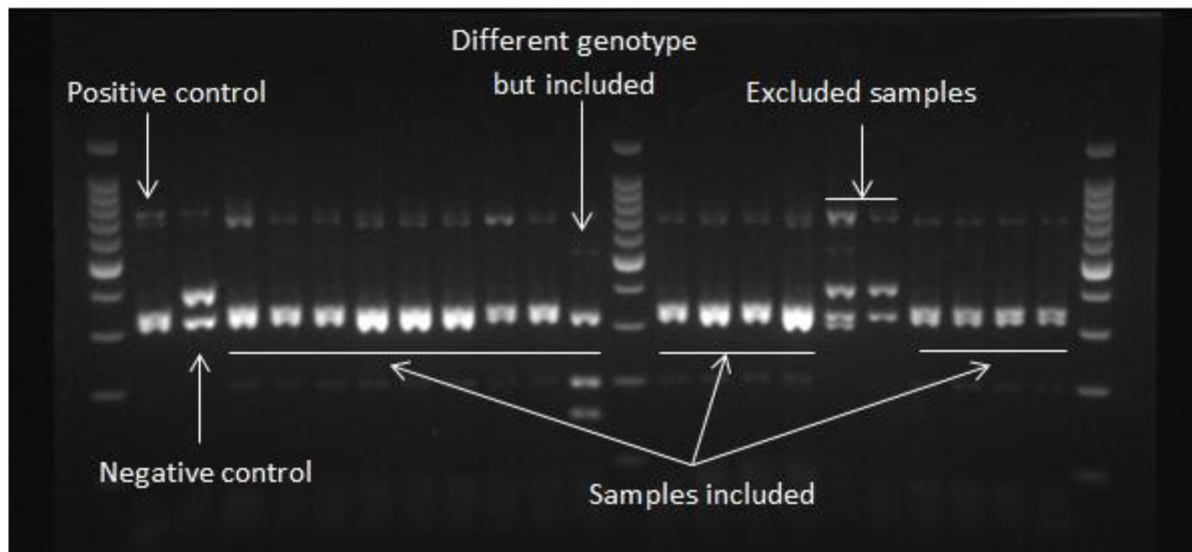


Figure 2.8. Photograph of a 3% ethidium-bromide stained agarose electrophoresis gel under UV light, showing PCR products after incubation with two restriction enzymes (*Bam*H1 and *Xba*1). Samples left to right: 100bp ladder; positive control; negative control; 102B14; 102B15; 102B16; 94B3; 94B4; 94B5; 94B6; 94B7; 94B8; 100bp ladder; 94B9; 94B10; 94B11; 94B12; 94B13; 94B14; 94B15; 94B16; 94B17; 94B18; 100bp ladder.

2.3 Next Generation Sequencing

Identification of microsatellite markers was required for *Isoperla grammatica* and *Amphinemura sulcicollis*, and later also for *Baetis rhodani*. Although seven microsatellites had previously been described by Williams *et al.* (2002), additional microsatellites were required as it is unknown how many unlinked loci are needed to distinguish between recently diverged populations and to increase statistical power (Haasl and Payseur 2011; Putman and Carbone 2014). Ten polymorphic microsatellites loci are usually considered necessary per species (Cervini *et al.* 2006; Östergren *et al.* 2015). The increasing amounts of DNA sequencing reads and decreasing cost of Next Generation Sequencing (NGS) in recent years has considerably accelerated microsatellite locus discovery in other organisms (Yu *et al.* 2011; Fernandez-Silva *et al*. 2013). Therefore NGS was selected as the most time- and cost-effective method to identify novel microsatellites for *Amphinemura sulcicollis*, *Isoperla grammatica*, and *Baetis rhodani*. A complete description of the NGS preparation and subsequent analysis, including assembly, is described in a genomic resource note

(Macdonald *et al.* (2016a), Doi: http://dx.doi.org/10.1101/046227 and contained in Appendix B). Raw NGS data for each species are also available through the Short Read Archive (SRA STUDY: PRJNA315680 (SRP072016), and all assemblies have been deposited at DDBJ/ENA/GenBank (accessions LVVV00000000, LVVW00000000 and LVVX00000000).

2.3.1 *PrimerPipeline* development

In order to design new microsatellite markers for *Amphinemura sulcicollis*, *Isoperla grammatica*, and *Baetis rhodani*, repeat regions had to be located within each dataset and primers designed. This was first done using the programs MISA (http://pgrc.ipk-gatersleben.de/misa/ [accessed 14/01/16]) and Primer3 (Untergasser *et al.* 2012) using a Linux platform. However, challenges were encountered, and these two programs provided incorrect primers and a different approach was needed. In collaboration with a computer programmer (Greg Macdonald) and a colleague (Luis Cunha), a new Windows based program, *PrimerPipeline*, with a graphical user interface was developed. *PrimerPipeline* (http://www.scrufster.com/primerpipeline/), incorporates MISA and Primer3 and evaluates primer-pair outputs. The results display window allows microsatellites, primers and flanking regions to be clearly visualised. This user-friendly program has been thoroughly tested by myself and others around the world, who have provided very positive feedback (see Appendix C for program manual).

2.4 Developing novel microsatellite markers

Microsatellites were developed for all three species and have been published in Macdonald *et al.* (2016b) (accepted 26 May 2016; full manuscript available in Appendix E). For *A. sulcicollis*, 25 primer pairs were tested, four of which were dropped because of poor amplification (Appendix D, Table D1). For *I. grammatica*, 28 primer pairs were tested, 10 of which were rejected due to poor amplification (Appendix D, Table D2). For *B. rhodani*, 28 primer pairs were tested including those developed but not published by Williams (2003); nine primer pairs were rejected due to poor amplification (Appendix D, Table D3). New

primers designed during the course of this study have been submitted to Genbank (See Appendix D and E; Macdonald *et al.* (2016b) for accession numbers).

Table 2.5. PCR reagents for singleplex PCR for microsatellites in Appendix E, Table E1. PCR was carried out in a final volume of 5 µl.

PCR reagent	Final concentrations / quantity	Supplier
Multiplex mix	2.5 µl	Qiagen (Hilden, Germany)
Q solution	0.5 µl	Qiagen (Hilden, Germany)
Each primer	0.5 µM	Sigma (Gillingham, UK)
Fluorescent dye with complimentary M13 tail attached (TAM, HEX or FAM)	0.5 µM	Eurofins (MWG Operon)
DNA template (~50 ng)	1 µL	

Table 2.6. PCR conditions for microsatellites in Appendix E, Table E1. PCR was carried out in a final volume of 5 µl.

	Temp (°C)	Duration	
Initial denaturation	94	3 minutes	
	94	30 seconds	
Annealing Temperature	60	45 seconds	x 35
	72	1 minute	
Final extension	72	10 minutes	

2.4.1 Method

Using the PCR protocols detailed in Table 2.5 and Table 2.6, 28, 25 and 15 primer pairs (*I. grammatica*, *A. sulcicollis* and *B. rhodani*, respectively) were selected for screening using 36-40 samples from two sites in Wales (Appendix E, Table E1, E2). All forward primers were tagged with an M13 tail (5’-AGGGTTTTCCCAGTCACGACGTT-3’) at the 5’ end; this allows, in a singleplex PCR, a fluorescent dye (with a compliment M13 tail) to be attached. The use of

the M13 tail method (Boutin-Ganache *et al.* 2001) negates the need for fluorescently-labelled primers, significantly reducing the cost of genotyping (Schuelke 2000).

2.4.2 Data analysis

The number of alleles, heterozygosity and Hardy–Weinberg equilibrium (HWE) were calculated per locus for two populations of each species in GenAlEx (Peakall and Smouse 2012). Significances for multiple tests were adjusted using Bonferroni correction (Rice 1989). Polymorphism information content (PIC) was calculated per population using Cervus v.3.0.7 (Kalinowski *et al.* 2007). The presence of null alleles, allelic dropout and scoring errors were ascertained using Micro-Checker v2.2.3 (Van Oosterhout *et al*. 2004). Linkage disequilibrium (LD) was estimated using GENEPOP 4.2 (Raymond and Rousset 1995).

2.4.3 Results and discussion

A set of 18, 21 and 13 robust polymorphic microsatellites were identified for *I. grammatica*, *A. sulcicollis* and *B. rhodani*, respectively. High numbers of alleles and PIC were found for all three species when analysed per site (Na = 7-27, 3-16, 5-13, average 17, 9, 9; PIC= 0.67-0.95, 0.36-0.90, 0.42-0.88, average 0.88, 0.73, 0.72; Appendix E, Table. 1) showing that they are a useful resource.

Some loci within each species showed significant departure from HWE for both populations (Iso_11-18, Amp_16-21 and B_11-13; Appendix E, Table E1) along with evidence of null alleles (Iso_7-18, Amp_10-21 and B_9-13; Appendix E, Table E1). Heterozygosity deficit and null alleles are commonly observed within insects (Chapuis and Estoup 2007), including within Baetidae (Alp *et al.* 2012). How this affects the data has been considered within Section 2.6. There was no significant LD or scoring errors found for any of the loci.

2.5 Final Dataset

2.5.1 Choosing Final Microsatellites

Of the microsatellites already published, it was decided that ten microsatellites per species would offer a good resolution (as well as being feasible within both time and budget constraints). These microsatellites were chosen on the basis of preliminary analysis (Appendix E, Table E1) using a combination of three traits: 1) least likely to have null alleles and departure from Hardy Weinberg equilibrium (HWE), 2) easiest to score thus avoiding errors, and 3) evidence of good amplification success to avoid missing data.

For *I. grammatica*, departure from HWE was common, however preliminary analysis showed no consistent pattern of departure from HWE in the first ten microsatellites (Iso_1 - Iso_10; Iso_1-4 were in HWE at both sites whereas with Iso_5-10, only one site was significantly out of HWE, Appendix E, Table E1). These microsatellites also had high success rates and were relatively simple to score, hence Iso_1 - Iso_10 were chosen for ongoing analysis.

Compared to *I. grammatica, A. sulcicollis*, had a greater number of microsatellites that showed inconsistent departure from HWE (Amp_1-Amp_15, Appendix E, Table E1) and the two remaining traits were considered more carefully. Amp_1 and Amp_12 were not chosen because the success rates were lower than the other microsatellites (e.g. only 15 and 11 successful samples out of 20 attempted respectively, Appendix E, Table E1, E2) and Amp_7 was excluded because it was harder to score accurately. Therefore the following ten microsatellites were selected to create the final dataset: Amp_2, Amp_3, Amp_4, Amp_5, Amp_6, Amp_8, Amp_9, Amp_10, Amp_11 and Amp_13.

For *B. rhodani*, seven microsatellites (Brh-1, Brh-2, Brh-3, Brh-4, Brh-6 and Brh-7) had already been developed by Williams (2003). It was decided that six more microsatellites were needed to ensure that there would be at least ten microsatellites as some of the original microsatellites showed variable success rates (Brh-1, Brh-3 and Brh-7). The first ten newly developed microsatellites showed no consistent departure from HWE (Appendix E, Table E1). B_1, B_2, B_3, B_4, B_5 and B_7 were chosen to be added to the original seven

(Appendix E, Table E1). B_6 was not chosen because B_7 was easier to score. This species was genotyped at 13 loci.

2.5.2 Data quality

The microsatellite loci for all three species had very high numbers of alleles (*A. sulcicollis* 11-56; *I. grammatica* 38-72 and *B. rhodani* 11-55). This made scoring difficult and the fact that 26 of the 33 microsatellites had never been used before, meant that extra time and effort had to be invested in ensuring high levels of genotyping accuracy. To attain an accurate and complete a set of genotyping data, the following steps were followed:

- High levels of repeats randomly in the dataset (Table. 2.7) to check and ensure scoring and binning accuracy, for some loci considerably higher than the recommended 10% (Dewoody *et al.* 2006).
- Samples that were in any way ambiguous to score or bin were repeated.
- Unique alleles were repeated to ensure that they were true alleles, and confirm a binning range for that allele.
- To ensure low levels of missing data all failed samples were repeated at least three times.
- In every separate fragment analysis order there were negatives and repeats to ensure that any contamination was found and any differences between orders were rectified.
- Projects were scored multiple times to test scoring accuracy and any discrepancies investigated usually leading to a new scoring rule.
- Results differing on repeat would be re-analysed until a consensus could be found. If no consensus was possible then the sample was excluded.

The final dataset required 12,119 PCR reactions. For *A. sulcicollis* this included 278 samples across 17 different sites (Appendix F, Table F1 and F2). Between 49 and 136 samples were repeated at least twice ranging from 27.4% (Amp_10) to 76.0% (Amp_13) of samples (Table 2.7a, mean 43.4%). This high level of repetition allowed confidence in scoring and in the binning of alleles. It also meant that the level of missing data was very low; only three loci

had any missing data (Amp_10, Amp_11 and Amp_13) with 1 - 22 (0.4%- 7.9%) samples missing per locus (Table 2.7a). No samples had missing data at more than two loci so none were removed.

The *Isoperla grammatica* dataset included 237 individuals from 13 sites (Appendix F, Table F4 and F5). Between 58 (Iso_7, 24.5%) and 134 (Iso_5, 56.5%) individuals were repeated (Table 2.7b; mean 37.5%). Six loci had low levels of missing data (1 – 22 samples; Table 2.7b), but there was only two individuals that had missing data at two loci (98I5 and 108I2). Finally, *B. rhodani* included 186 individuals from 10 sites. The amount of repetition needed varied greatly between loci, with just 20 individuals repeated for B_3 (10.8%) and 102 for Brh-6 (54.8%) based on the amount of ambiguous samples within each loci (Table 2.7c). However, again a very low missing data level was achieved; only two individuals (09B15 and 09B2) had missing data at two loci.

2.5.3 Binning alleles

Usually all microsatellite scores are binned into unique alleles. If there are only a couple of alleles requiring such action, this can usually be done quite simply by hand, however if there are many such alleles, automated binning is typically used (Amos *et al.* 2007; Guichoux *et al.* 2011; Flores-Rentería and Krohn 2013). In this dataset there were 1,112 alleles so a binning system was needed to cope with this number. Firstly, automated binning of microsatellite allele lengths was attempted using TANDEM (Matschiner and Salzburger 2009). Errors were found to occur using this method however, for example with clear heterozygotes binned into homozygotes (i.e two alleles binned into one). It was concluded that the available programs were not fit for the purposes of analysing this dataset and manual binning was therefore used.

Rules were established for each allele, setting a minimum and maximum allele length (in base pairs, bp) and ensuring that no heterozygotes could be binned as homozygotes. A high number of repetitions were necessary to check the accuracy of these limits. The recorded size range within alleles (difference between the minimum and maximum allele length) was noted and the difference between the upper limit of one allele and the lower limit of the

subsequent allele were observed to ensure these ranges were appropriate. The tables of rules for each microsatellite located in Appendix F were completed with the final dataset.

Table 2.7a-c. Summary of final dataset for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhondoni*, per loci and then in total. *No. Repeated* = the number of individual samples that were repeated at least twice; *Repeated as %* = the percentage of total samples that were repeated at least once; *Total successful* = total number of successful samples genotyped, including repeats; *Total failed* = total number failed, including repeats; *Missing data* = the number of individual samples that could not be genotyped; *Missing data as %* = the percentage of data that are missing from the whole dataset per loci; *Not Included* = the number of samples not included in Appendix F final dataset, due to failure to amplify, usually due to species miss-identification; *Negatives* = number of negative samples (blank) sent for fragment analysis, (not included in Appendix F Final dataset); *Total samples* = total number of samples sent for genotyping; and *No. of alleles* = the number of different alleles in the dataset.

Table 2.7a. *Amphinemura sulcicollis* (Summary of Table F1, F2 and F3i-x in Appendix F).

	Amp_2	Amp_3	Amp_4	Amp_5	Amp_6	Amp_8	Amp_9	Amp_10	Amp_11	Amp_13	Total
No. repeated	61	61	73	82	97	62	65	49	90	136	776
Repeated as %	34.1	34.1	40.8	45.8	54.2	34.6	36.3	27.4	50.3	76.0	-
Total successful	341	342	355	367	378	340	348	308	366	433	3,578
Total failed	10	4	8	62	21	8	6	91	55	19	284
Missing data	0	0	0	0	0	0	0	22	11	1	34
Missing data as %	0	0	0	0	0	0	0	7.9	4.0	0.4	-
Not included	13	13	44	20	13	13	45	7	45	20	233
Negatives	13	13	15	17	13	12	12	9	11	16	131
Total samples	377	372	422	466	425	373	411	415	477	488	4,226
No. of alleles	36	18	13	19	11	15	20	22	56	17	227

Table 2.7b. *Isoperla grammatica* (Summary of Table F4, F5 and F6i-x in Appendix F).

	Iso_1	Iso_2	Iso_3	Iso_4	Iso_5	Iso_6	Iso_7	Iso_8	Iso_9	Iso_10	Total
No. repeated	93	97	122	63	134	74	58	73	106	68	888
Repeated as %	39.2	40.9	51.5	26.6	56.5	31.2	24.5	30.8	44.7	28.7	-
Total successful	337	349	375	303	380	316	309	319	336	306	3330
Total failed	6	69	23	29	50	21	8	12	139	5	362
Missing data	0	0	1	1	8	4	2	0	22	0	38
Missing data as %	0.0	0.0	0.4	0.4	3.4	1.7	0.8	0.0	9.3	0.0	-
Not included	0	0	0	0	0	0	0	0	0	0	0
Negatives	14	15	12	10	13	12	13	13	14	14	128
Total samples	357	433	410	342	443	349	330	344	491	326	3825
No. of alleles	43	61	48	65	59	60	64	54	72	38	564

Table 2.7c. *Baetis rhodani* (Summary of Table F7, F8 and F9i-xiii in Appendix F).

	Brh-1	Brh-2	Brh-3	Brh-4	Brh-5	Brh-6	Brh-7	B_1	B_2	B_3	B_4	B_5	B_7	Total
No. repeated	71	32	100	53	36	102	84	32	29	20	37	47	33	676
Repeated as %	38.2	17.2	53.8	28.5	19.4	54.8	45.2	17.2	15.6	10.8	19.9	25.3	17.7	
Total successful	269	232	313	259	236	308	290	220	216	205	223	242	213	3226
Total failed	78	34	48	71	14	28	82	3	3	23	5	5	19	413
Missing data	8	5	1	1	2	3	1	0	0	0	0	1	5	27
Missing data as %	4.3	2.7	1.1	0.5	1.1	1.6	0.5	0.0	0.0	1.1	0.0	0.5	3.8	
Not included	17	17	22	19	18	18	19	25	21	24	24	25	22	271
Negatives	13	12	14	14	10	14	13	11	10	10	10	11	10	152
Total samples	377	295	397	363	278	368	404	259	250	262	264	283	268	4068
No. of alleles	13	17	19	27	27	26	55	19	21	22	11	40	24	321

2.6 Null Alleles

2.6.1. Introduction

Microsatellite null alleles is an issue which results from mutations at primer sites that cause certain alleles not to amplify during PCR, leading to false homozygotes (Shaw *et al.* 1999a; Van Oosterhout *et al*. 2004). Non amplified (null) alleles are a common issue in microsatellite studies but can bias estimates of allele and genotype frequencies decreasing observed heterozygosity and increase the apparent level of inbreeding (Dewoody *et al.* 2006; Van Oosterhout *et al.* 2006).

Although microsatellites were chosen to reduce the instance of microsatellite null alleles within the final dataset, they could potentially remain an issue for all three species. This is not unexpected as many studies have noted the high frequency of null alleles within insects (for example Ephemeroptera (Alp *et al*. 2012); Lepidoptera (Meglecz *et al.* 2004); Diptera (Lehmann *et al.* 1997) and Orthoptera (Chapuis *et al.* 2005; Chapuis and Estoup 2007)). Such observations are thought to be linked to both large effective population sizes and highly variable flanking regions (Chapuis and Estoup 2007), typical of invertebrate populations.

Null allele scoring errors are particularly challenging to detect because, by definition, null alleles fail to produce a visible product (Dakin and Avise 2004; Dewoody *et al*. 2006), however they can be indirectly tested by examining Hardy-Weinberg (HW) proportions (Chapuis and Estoup 2007). Many population genetic models assume panmixia and that the population is in HWE (Chakraborty *et al.* 1992; Brookfield 1996; Van Oosterhout *et al.* 2004). Null alleles can, however, be misidentified if the observed heterozygote deficiencies have another cause (e.g. the Wahlund effect, where more than one population is inadvertently analysed as a single genetic unit).

The aim of this section was to 1) investigate null alleles and other scoring errors in the final dataset; 2) to begin to investigate their effect on the data and 3) to discuss reasons for finding evidence of null alleles and 4) to decide whether any loci should be dropped from further analysis.

2.6.2 Data analysis

For each species, the number of alleles, heterozygosity and HWE were estimated per locus for all sites comprising more than 10 samples using the software GenAlEx (Peakall and Smouse 2012). Significances for multiple tests were adjusted using Bonferroni correction (Rice 1989). Polymorphism information content (PIC) was calculated per site using Cervus v.3.0.7 (Kalinowski *et al.* 2007). The presence of null alleles, as well as scoring errors such as allelic dropout and stuttering were ascertained using Micro-Checker v2.2.3 (Van Oosterhout *et al.* 2004). Allelic dropout describes an error where smaller alleles are preferentially amplified in a heterozygous genotype; this can cause potential scoring error that can bias allele and genotype frequencies (Dewoody *et al*. 2006). Whereas stuttering can be produced by slippage of taq polymerase during PCR which produces stutter bands (multiple peaks for one allele), some loci are prone to stuttering and can contribute to scoring errors (Dewoody *et al.* 2006). For *A. sulcicollis* 13 sites were compared at each loci, *I. grammatica* had 12 sites and *Baetis rhodani* had 10. All individuals inferred to be related (at a probability of ≥ 0.8 for half siblings or more closely related individuals) were removed from the data before this analysis was performed (described in Section 3.3).

2.6.3 Results

Average sample sizes (N) per site ranged from 16-17, 17-19 and 16-17 per loci for *A. sulcicollis*, *I. grammatica* and *B. rhodani* (Table 2.8a-c), respectively, therefore success rate of each locus did not differ greatly between loci or species. All three species were found to be very polymorphic with the mean number of alleles (Na) ranging from 5 (Amp_6) - 15 (Amp_2); 11 (Iso_10) - 23 (Iso_2 and Iso_4) and 6 (Brh-2 and B_4) - 20 (Brh-7). The mean polymorphic information content (PIC) ranged from 0.59 (Amp_3) - 0.87 (Amp_2); 0.82 (Iso_10) - 0.94 (Iso_3 and Iso_4) and 0.55 (B_4) - 0.93 (Brh-7), (the complete descriptive statistics per loci can be found in Appendix G and are summarised per species in Table 2.8a-c).

The percentage of sites that were found to deviate from HWE after Bonferroni correction per locus ranged from 0 - 100% for *A. sulcicollis*. Three loci (Amp_10, Amp_11 and Amp_13), had significant deviation from HWE at more than half of the sites due to heterozygote deficiency (Table 2.8a). Loci for *I. grammatica* ranged from 0 - 100% where results from seven loci were higher than 50% (Iso_4

- Iso_10, Table 2.8b). For *B. rhodani*, the results ranged from 0 - 90% with four loci (Brh-1 - Brh-3 and B_5) out of HWE for more than half of the sites analysed (Table 2.8c).

As expected, the levels of null alleles matched the patterns of HW disequilibrium for all three species. *Isoperla grammatica* had the highest levels both of polymorphism and inferred null alleles. Seven (Iso_4 - Iso_10) of the 10 loci (70%) showed evidence of null alleles in over half of the sites analysed (Table 2.9b). *B. rhodani* had five loci (Brh-1-3, B_4-5) and *A. sulcicollis* had four (Amp_9-11 and Amp_13) with inferred null alleles. However, because *B. rhodani* was analysed for 13 loci, the levels of null alleles within the data are very similar - 38% compared to 40% for A. *sulcicollis* (Table 2.9a and Table 2.9c). For *B. rhodani* evidence of scoring error due to stuttering were found in three loci (Brh-1 in 50% of the sites, Brh-3 and B_4 in 20% of the sites). There were no consistent evidence of scoring errors or allelic dropout was found in any other loci in any of the species.

Table 2.8a-c. Summary of descriptive statistics for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*, per loci, per site (only sites with <10 samples per site were included). N = number of individuals; N_A = number of alleles per locus; H_O = observed heterozygosity; H_E = expected heterozygosity; HWE = p value of Hardy-Weinberg Equilibrium test; no. sig = total number of sites that are significantly out of HWE with and without Bonferroni correction (critical P value p ≤0.005)); % = percentage of sites significantly out of HWE with and without Bonferroni correction, and PIC = polymorphism information content.

Table 2.8a. *Amphinemura sulcicollis* (full data in Appendix G, Table G1).

	Locus	Amp_2	Amp_3	Amp_4	Amp_5	Amp_6	Amp_8	Amp_9	Amp_10	Amp_11	Amp_13
N	Min	15	15	15	15	15	15	15	11	14	15
	Max	19	19	19	19	19	19	19	18	19	19
	Average	17	17	17	17	17	17	17	16	17	17
Na	Min	10	5	5	7	4	8	5	6	9	5
	Max	21	10	9	14	7	12	11	13	18	10
	Average	15	8	7	11	5	9	7	10	13	7
Ho	Min	0.688	0.316	0.625	0.471	0.444	0.438	0.294	0.133	0.250	0.294
	Max	1.000	0.813	1.000	1.000	0.882	0.889	0.588	0.824	0.556	0.563
	Average	0.893	0.644	0.840	0.835	0.707	0.727	0.409	0.474	0.432	0.450
uHe	Min	0.797	0.290	0.774	0.497	0.579	0.802	0.435	0.618	0.800	0.558
	Max	0.952	0.808	0.867	0.925	0.815	0.908	0.827	0.922	0.944	0.889
	Average	0.910	0.637	0.830	0.834	0.687	0.867	0.707	0.813	0.883	0.772
HWE	Min	0.000	0.064	0.039	0.037	0.003	0.000	0.000	0.000	0.000	0.000
	Max	1.000	1.000	1.000	1.000	0.994	0.662	0.065	0.254	0.000	0.014
	Average	0.495	0.589	0.545	0.522	0.461	0.228	0.012	0.026	0.000	0.002
	no. sig (p=0.005)	1	0	0	0	1	3	6	10	13	12
	% with Bonferroni	8	0	0	0	8	23	46	77	100	92
	no. sig (p=0.05)	1	0	1	2	2	5	12	11	13	13
	% without Bonferroni	8	0	8	15	15	38	92	85	100	100
PIC	Min	0.75	0.27	0.71	0.46	0.50	0.75	0.40	0.57	0.75	0.52
	Max	0.92	0.75	0.82	0.89	0.76	0.87	0.79	0.87	0.91	0.85
	Average	0.87	0.59	0.78	0.79	0.61	0.82	0.65	0.76	0.84	0.72

Table 2.8b. *Isoperla grammatica* (full data in Appendix G, Table G2).

		Iso_1	Iso_2	Iso_3	Iso_4	Iso_5	Iso_6	Iso_7	Iso_8	Iso_9	Iso_10
N	Min	15	15	15	15	15	15	15	15	14	15
	Max	20	20	20	20	20	20	20	20	19	20
	average	19	19	19	19	19	19	19	19	17	19
Na	Min	12	19	14	16	16	15	17	16	14	9
	Max	20	25	23	28	22	25	26	23	24	15
	average	16	23	19	23	19	18	20	21	19	11
Ho	Min	0.632	0.647	0.733	0.600	0.235	0.353	0.500	0.412	0.200	0.250
	Max	1.000	0.950	0.947	0.947	0.684	0.947	0.947	0.850	0.579	0.700
	average	0.869	0.850	0.851	0.769	0.526	0.573	0.708	0.664	0.409	0.437
He	Min	0.757	0.942	0.892	0.940	0.888	0.893	0.927	0.910	0.921	0.683
	Max	0.947	0.976	0.964	0.982	0.976	0.974	0.980	0.972	0.976	0.904
	average	0.907	0.965	0.941	0.964	0.943	0.932	0.954	0.959	0.954	0.860
HWE	Min	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	Max	1.000	0.625	0.797	0.688	0.000	0.930	0.425	0.035	0.000	0.081
	average	0.403	0.134	0.189	0.059	0.000	0.078	0.068	0.003	0.000	0.007
	no. sig (p=0.005)	0	5	2	10	12	11	8	11	12	11
	% with Bonferroni	0	42	17	83	100	92	67	92	100	92
	no. sig (p=0.05)	3	7	6	11	12	11	9	12	12	11
	% without Bonferroni	25	58	50	92	100	92	75	100	100	92
PIC	Min	0.72	0.91	0.86	0.91	0.86	0.86	0.90	0.88	0.89	0.65
	Max	0.92	0.95	0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.87
	average	0.87	0.94	0.91	0.94	0.91	0.90	0.92	0.93	0.92	0.82

Table 2.8c. *Baetis rhodani* (full data in Appendix G, Table G3).

		Brh-1	Brh-2	Brh-3	Brh-4	Brh-5	Brh-6	Brh-7	B_1	B_2	B_3	B_4	B_5	B_7
N	Min	10	11	11	10	11	8	11	11	11	11	11	11	11
	Max	19	20	20	20	20	20	20	20	20	20	20	20	20
	Average	16	16	17	17	16	16	17	17	17	17	17	17	16
Na	Min	5	4	4	7	9	6	17	4	9	10	4	7	5
	Max	9	8	11	14	15	12	23	9	13	16	7	14	11
	Average	7	6	7	10	11	10	20	7	11	12	6	10	7
Ho	Min	0.133	0.313	0.125	0.600	0.563	0.438	0.647	0.625	0.611	0.667	0.313	0.313	0.267
	Max	0.500	0.636	0.455	0.938	0.938	0.875	1.000	0.938	1.000	0.941	0.625	0.650	0.750
	Average	0.342	0.471	0.232	0.795	0.768	0.644	0.857	0.732	0.839	0.779	0.412	0.547	0.593
He	Min	0.659	0.683	0.286	0.790	0.805	0.760	0.952	0.683	0.783	0.883	0.518	0.779	0.569
	Max	0.873	0.825	0.818	0.907	0.890	0.917	0.983	0.792	0.911	0.936	0.673	0.911	0.829
	Average	0.780	0.764	0.686	0.858	0.855	0.825	0.967	0.746	0.861	0.909	0.617	0.848	0.703
HWE	Min	0.000	0.001	0.000	0.003	0.000	0.000	0.000	0.130	0.030	0.001	0.000	0.000	0.000
	Max	0.009	0.255	0.091	0.939	0.946	0.361	1.000	0.984	1.000	0.688	0.136	0.025	0.702
	Average	0.001	0.027	0.009	0.332	0.268	0.125	0.248	0.399	0.386	0.121	0.041	0.004	0.320
	No. sig (p=0.005)	9	9	9	1	2	2	3	0	0	3	4	8	2
	% with Bonferroni	90	90	90	10	20	20	30	0	0	30	40	80	20
	No. sig (p=0.05)	10	9	9	4	4	5	5	0	1	5	7	10	3
	% without Bonferroni	100	90	90	40	40	50	50	0	10	50	70	100	30
PIC	Min	0.59	0.60	0.26	0.74	0.75	0.70	0.92	0.61	0.74	0.84	0.44	0.72	0.54
	Max	0.83	0.77	0.77	0.87	0.86	0.85	0.94	0.73	0.87	0.90	0.61	0.87	0.78
	Average	0.72	0.70	0.64	0.81	0.81	0.77	0.93	0.68	0.82	0.87	0.55	0.80	0.65

Table 2.9a-c. Describing results from Micro-Checker for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*, per loci, per site (only sites with greater than 10 samples were included). Where 'yes' = there is evidence of null alleles and 'no' = there is no evidence of null alleles ([1] denotes evidence for scoring error due to stuttering). Total= the total number of sites that show evidence of null alleles per loci.

Table 2.9a. *Amphinemura sulcicollis*.

Locus	Site code													
	104	109	108	102	112	96	9	116	113	95	6	59	93	Total
Amp_2	no	no	no	no	no	no	no	no	no	no	no	no	no	0
Amp_3	no	no	no	no	no	no	no	no	no	no	no	no	no	0
Amp_4	no	no	no	no	no	no	no	no	no	no	no	no	no	0
Amp_5	no	no	no	no	yes	no	no	no	no	no	no	no	no	1
Amp_6	no	no	no	no	no	no	yes	no	no	no	no	no	no	1
Amp_8	yes	yes	no	no	no	no	no	no	yes	no	no	no	no	3
Amp_9	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	no	yes	no	10
Amp_10	no	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	11
Amp_11	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	13
Amp_13	yes	yes	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	yes	12

Table 2.9b. *Isoperla grammatica*.

Locus	Site code												Total
	105	106	108	12	97	98	112	114	115	10	118	93	
Iso_1	no	no	no	no	no	no	no	no	no	no	no	no	0
Iso_2	no	yes	no	no	yes	yes	no	yes	no	no	no	no	4
Iso_3	yes	yes	no	yes	no	no	no	no	no	no	yes	no	4
Iso_4	yes	yes	yes	no	yes	yes	yes	yes	yes	yes	yes	no	10
Iso_5	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	12
Iso_6	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	11
Iso_7	no	yes	no	yes	no	yes	yes	yes	yes	yes	yes	yes	9
Iso_8	yes	no	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	11
Iso_9	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	12
Iso_10	yes	yes	yes	yes[1]	yes	no	yes	yes	yes	yes[1]	yes	yes	11

Table 2.9c. *Baetis rhodani*.

Locus	Site code										Total
	112	102	97	96	9	106	113	115	94	118	
Brh-1	yes	yes[1]	yes[1]	yes[1]	yes	yes	yes[1]	yes[1]	yes	yes	10
Brh-2	yes	yes[1]	yes	yes	no	no	no	yes	yes	yes	7
Brh-3	yes	yes	yes[1]	yes[1]	yes	yes	no	yes	yes	yes	9
Brh-4	no	yes	no	no	yes	no	no	no	no	no	2
Brh-5	yes	yes[1]	no	no	no	no	no	no	no	no	2
Brh-6	no	no	no	yes	no	no	yes[1]	yes	no	yes	4
Brh-7	no	yes	no	yes	no	yes	yes	no	no	no	4
B_1	no	no	no	no	no	no	no	no	no	no	0
B_2	no	no	no	no	no	yes	no	no	no	no	1
B_3	no	no	no	no	no	yes	no	no	no	yes	2
B_4	yes	yes	yes[1]	yes	no	yes[1]	no	no	no	no	5
B_5	yes	yes	yes	yes	no	yes	yes	yes	yes	yes	9
B_7	no	no	no	no	no	yes	no	no	yes	no	2

2.6.4 Further investigation into effect on genetic structure

2.6.4.1 Data analysis

Due to the high level of null alleles in all three species, removing all the affected loci would severely deplete the data and the resolution with which to answer the main aims of this study, especially for *I. grammatica* and *A. sulcicollis*. Attempting to replace the loci that are suspected of null alleles would also be inappropriate as the loci remaining had already been identified as having null alleles in the initial screening with two sites, or were rejected for another reason, for example poor amplification (see Section 2.5.1 and Appendix D).

The presence of null alleles has been shown to overestimate population differentiation (Chapuis and Estoup 2007). To explore this possibility, an investigation was carried out on how the loci suspected of null alleles influenced the structure of *A. sulcicollis*. The approach involved running STRUCTURE (Pritchard *et al.* 2000) with only the loci not suspected of null alleles before adding one locus at a time to determine each locus' influence and to ensure that population structure was not being unduly influenced by particular loci. For *I. grammatica,* removing all loci suspected of null alleles would have only left three loci; too low a number for implementing STRUCTURE, whereas *B. rhodani* showed no population structure so the null alleles present could not be influencing inferred population differentiation.

Using STRUCTURE v2.3.3 (Pritchard *et al.* 2000; Falush *et al.* 2003) all plots were the result of seven iterations using all available sites; and the program Clump (Jakobsson and Rosenberg 2007) was used to average these iterations. All plots were run with 100,000 Burn-in period and 1,000,000 Markov chain Monte Carlo (MCMC) repetitions; all were run with no priors and investigated at a number of different assumed clusters (K=1-5).

To find the most probable number of clusters (K value), as well as using the graphical outputs from STRUCTURE, the mean likelihood distribution (mean LnP(K)), and the Delta K (*ΔK*) were found using STRUCTURE Harvester web v0.6.93 (Earl and vonHoldt 2012). *ΔK* is the second order rate of change in log probability between successive K values, and the mean LnP(K) is the mean log likelihood of the data at each step of the MCMC (Evanno *et al.* 2005), and both can be used to help find the most probable number of clusters with a dataset.

2.6.4.1 Results

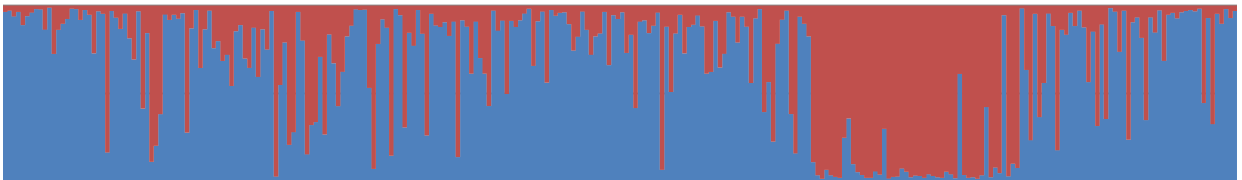
First of all, all ten loci were used to create STRUCTURE plots representing two, three and four clusters (Figure 2.9a-c), so that this could be compared with subsequent plots (where loci exhibiting evidence of null alleles are removed). Using the full set of loci population structure was detected for *A. sulcicollis*, particularly with individuals contained in sites 95 and 6, and to a lesser extent the last site (most northern) 93, and the first site (most southern) 104; the rest of the individuals are a mixture of clusters (Figure 2.9a-c; details of STRUCTURE result explained in Chapter 3).

When, however, the null allele loci are removed, and only the six remaining loci are used to run STRUCTURE, the pattern of structure is almost entirely lost (Figure 2.10). Although there is a slight pattern around individuals contained in sites 95 and 6, it is very different when compared to the full dataset (Figure 2.9).

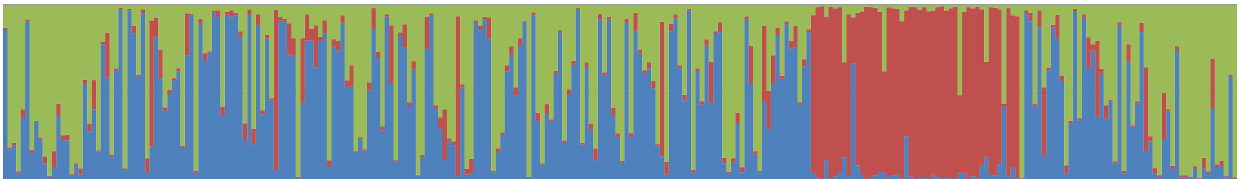
By adding one locus at a time back, the effect of individual loci can be discerned. For example, Figure 2.11 was made with seven loci, all six loci without evidence of null alleles plus loci ‘Amp_9’ which shows evidence of null alleles (Table. 2.9a), therefore Figure 2.11 shows the effect of ‘Amp_9’ individually. Using these loci the STRUCTURE plot shows even less structure than without ‘Amp_9’ included when assuming two clusters (Figure 2.11a),

however shows slightly more structure when assuming three clusters (Figure 2.11b). In this case three clusters was considered to most likely number of clusters using STRUCTURE harvester, whereas in all other subsequent plots (Figure 2.12- 2.14) two clusters was considered the most likely. In turn, the other three loci showing evidence of null alleles were tested by combining them with the six loci with no null alleles. The effect of Amp_10 (Figure 2.12), Amp_11 (Figure 2.13) and Amp_13 (Figure 2.14), gives very similar results, showing slightly more structure than with Amp_9 (Figure 2.11) or then with only six loci (Figure 2.10), but not as much as with all loci combined (Figure 2.9).

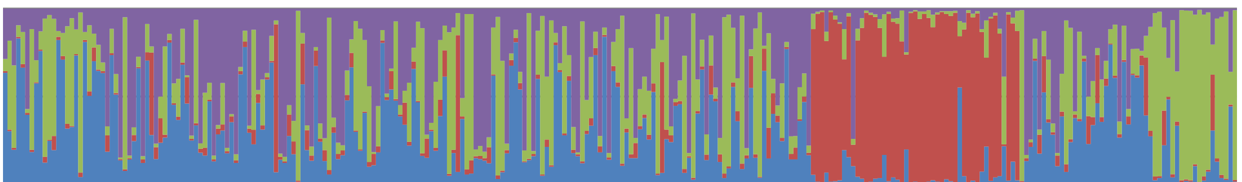
a)



b)



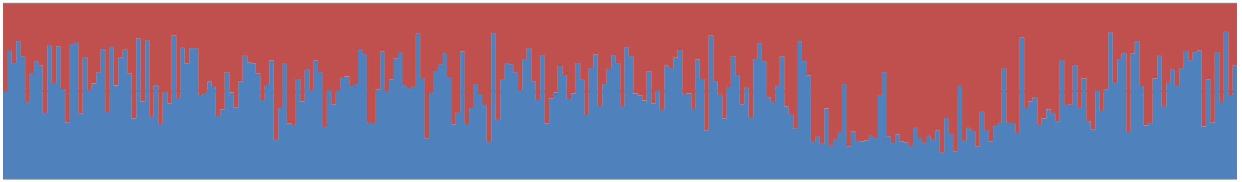
c)



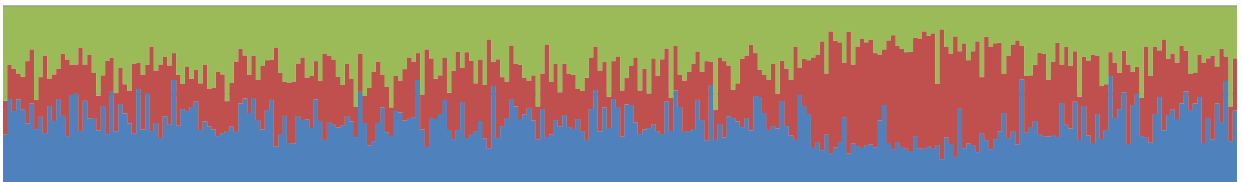
104	109	108	102	112	46	96	9	116	113	95	6	94+4	55	59	93

Figure 2.9. STRUCTURE plots showing results for *Amphinemura sulcicollis* using the whole dataset (10 loci including 4 loci that show evidence of null alleles: Amp_9, 10, 11 and 13) a) assumed two clusters; K=2, b) assumes three clusters; K=3 and c) assumes four clusters K=4. Site numbers shown below (grey shows partial sites), sites ordered from south to north of Wales.

a



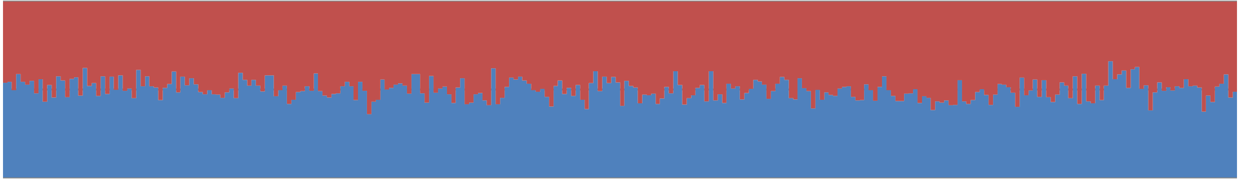
b)



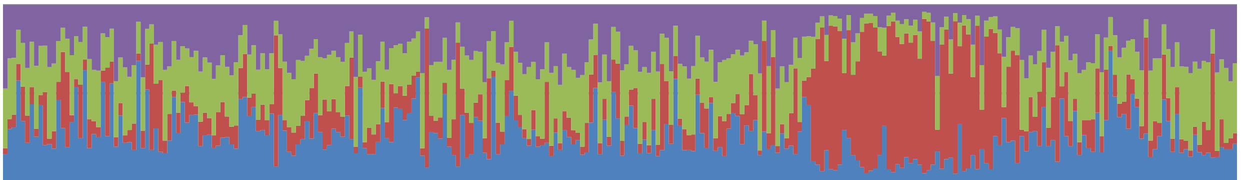
104	109	108	102	112	4 6	96	9	116	113	95	6	9 4 + 4	5 5	59	93

Figure 2.10. STRUCTURE plot showing results for *Amphinemura sulcicollis* using six loci only, not suspected of null alleles, a) assumed two clusters; K=2 and b) assumes three clusters; K=3. Site numbers shown below (grey shows partial sites), sites ordered from south to north of Wales.

a)

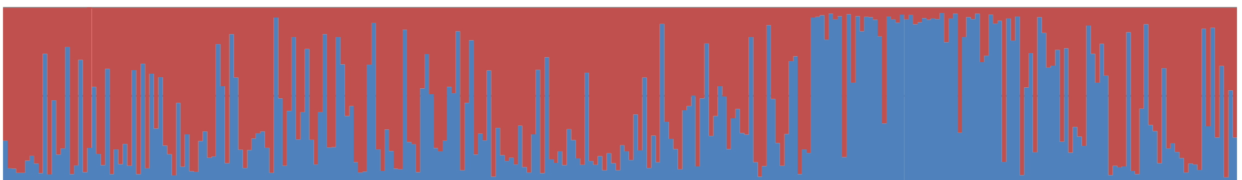


b



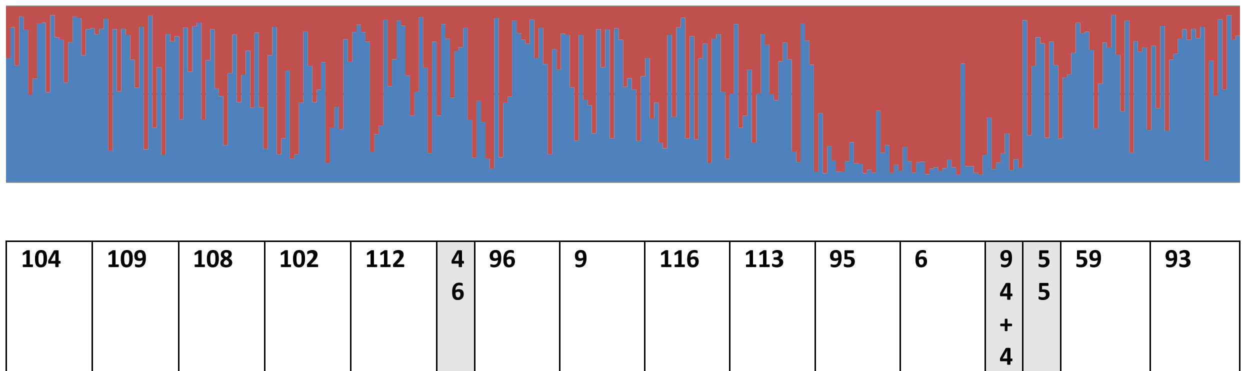
104	109	108	102	112	46	96	9	116	113	95	6	94+4	55	59	93

Figure 2.11. STRUCTURE plot showing results for *Amphinemura sulcicollis* using six loci without evidence of null alleles plus Amp_9 a) assumed two clusters; K=2 and b) assumes three clusters; K=3. Site numbers shown below (grey shows partial sites), sites ordered from south to north of Wales.



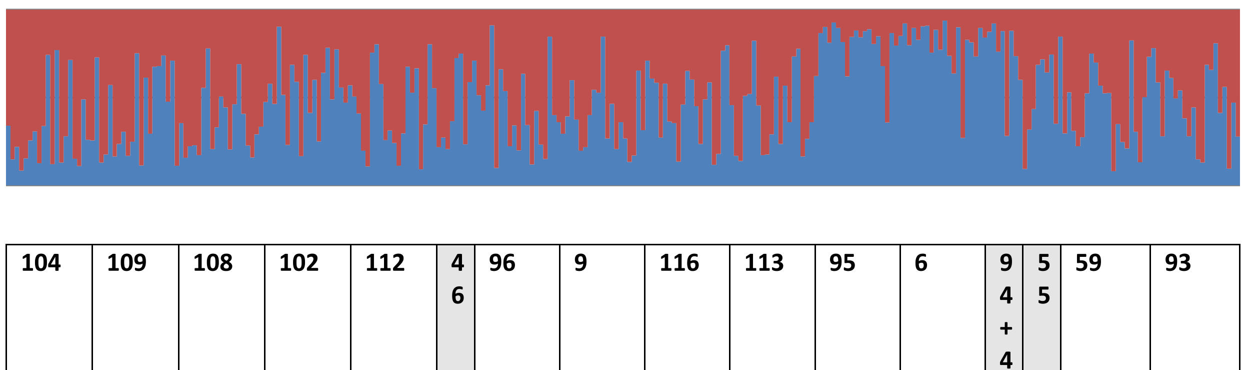
104	109	108	102	112	46	96	9	116	113	95	6	94+4	55	59	93

Figure 2.12. STRUCTURE plot showing results for *Amphinemura sulcicollis* using six loci without evidence of null alleles plus Amp_10 for K=2 (identified by structure harvester to be the most likely value of K).



104	109	108	102	112	46	96	9	116	113	95	6	94+4	55	59	93

Figure 2.13. STRUCTURE plot showing results for *Amphinemura sulcicollis* using six loci without evidence of null alleles plus Amp_11 for K=2 (identified by structure harvester to be the most likely value of K).



104	109	108	102	112	46	96	9	116	113	95	6	94+4	55	59	93

Figure 2.14. STRUCTURE plot showing results for *Amphinemura sulcicollis* using six loci without evidence of null alleles plus Amp_13 for K=2 (identified by structure harvester to be the most likely value of K).

2.6.5 Discussion

As expected, due to their frequency in invertebrates (Addison and Hart 2005; Brownlow *et al.* 2008; Schultheis *et al.* 2008; Postaire *et al.* 2015) evidence of null alleles have been found in a high proportion of loci for *A. sulcicollis*, *I. grammatica* and *B. rhodani*. Evidence of null alleles were found because there was a lower than expected (under Hardy-Weinberg) number of heterozygotes in the data. This heterozygote deficiency also results in significant departure from Hardy-Weinberg, though reasons for this will be discussed in chapter 3.

In *B. rhodani* a small number of loci were identified as having scoring error due to stuttering, however, because of the high level of repetition at these loci (19.2%-53.8%, Table 2.7c) and

the measures taken to ensure scoring accuracy (Section 2.5.2), it could be that stuttering was picked up by Microchecker because these loci have many subsequent alleles (alleles only two base pairs apart, see Appendix F, Table F11i, iii, xi) rather than caused by stuttering at these loci. No other genotyping error, apart from null alleles, was found consistently in any other species.

The further investigation into *A. sulcicollis* showed that running STRUCTURE with just six loci that did not show evidence of null alleles severely reduced the genetic structure, this was expected as STRUCTURE has been identified to need a greater number, especially when populations were closely related (Nelson and Anderson 2013). When adding one loci at a time (that did show null alleles), the amount of structure was very similar, indicating that individual loci do not seem to contribute to overestimating genetic structure.

Within this dataset there is a relationship between polymorphism (number of alleles per loci) and presence of null alleles. On average, *I. grammatica* has the highest total number of alleles per loci (Table 2.7b) and also the highest proportion of null alleles present (Table 2.9b). Comparatively, the other two species have both fewer instances of null alleles and a lower number of total alleles. The fact that loci that are extremely polymorphic tend to not be in HWE and therefore show evidence of null alleles could be due to variability in the flanking region of the microsatellite. Polymorphisms have been found previously, within the flanking region of microsatellites with highly polymorphic species (Blankenship *et al.* 2002; Ablett *et al.* 2006). This could cause a mutation at a priming site which could lead to null alleles.

Conversely, evidence of null alleles could be due to the fact that statistical and biological significance may not match up in highly diverse systems such as these (Hedrick 1999). Perhaps available programs are geared towards species with less polymorphism, such as vertebrates, however, within this data, for example loci 'Iso_9' has 72 unique alleles, perhaps the same assumptions can't be used for very different species.

In conclusion, the loci that have been identified as consistently out of Hardy-Weinberg and have shown evidence of null alleles, the possible effects on the data are still unclear. The experiment with *A. sulcicollis* and STRUCTURE did not show any one loci driving the genetic structure found, however, the high level of null alleles, especially in *I. grammatica*, could have consequences. Therefore, though all loci will be used in subsequent chapters, in

chapter 3 all results will be repeated without the loci out of HWE to further investigate their effect (Appendix I).

Chapter 3 - Genetic Structure and Diversity of Three Aquatic Invertebrates: Response to Environmental Stressors

Contributions

- Chemistry data made available from DURESS (Diversity of Upland Rivers for Ecosystem Service Sustainability: http://nerc-duress.org). Funded by the Natural Environment Research Council (NERC).

Highlights

- The genetic structure and genetic diversity of three freshwater invertebrates were compared. Genetic diversity for each species was then compared to environmental stressors such as pH.
- *Amphinemura sulcicollis* showed more genetic population structure, when compared to *Isoperla grammatica* and *Baetis rhodani.*
- Genetic diversity mirrored genetic population structure: isolated sites had significantly lower genetic diversity within *A. sulcicollis* and *I. grammatica*.
- Within *A. sulcicollis* genetic diversity was significantly positively correlated with pH, possibly suggesting that only certain genotypes can survive at the acidified sites.
- The genetic diversity of *I. grammatica* had a negative correlation with metal presence (aluminium and cadmium) which was driven by one isolated site and was not significant.
- *Baetis rhodani* does not show this pattern because this species is acid sensitive and was not present at study sites that could drive this adaption.

3.0 Abstract

Connectivity and dispersal between populations can have a strong effect on a species' resilience to disturbance and ability to recover or recolonise areas. Traditional assessment methods predict limited dispersal ability for adult aquatic invertebrates; however genetic methods, such as the use of microsatellite markers, have revealed previously unknown gene-flow. Here, the genetic structure and diversity of three freshwater macroinvertebrates (*Amphinemura sulcicollis*, *Isoperla grammatica* and *Baetis rhodani*) was investigated using microsatellite markers (10-13 per species) and the data used to infer dispersal and investigate the role of environmental stressors. It was hypothesised that due to its small body size *A. sulcicollis* may show the most isolation. It was also hypothesised that sensitivity to certain stressors would particularly affect acid sensitive *B. rhodani* to a greater extent than the other, more tolerant species.

There was high genetic polymorphism and significant Hardy-Weinberg disequilibrium in all three species due to heterozygote deficiency, commonly found in large outbreeding populations of invertebrate. As expected, genetic structure was different in each species; results from *B. rhodani* revealed no genetic structure, while *A. sulcicollis* clustered into four distinct populations reflected most strongly at two sites in North Wales. *I. grammatica* showed genetic isolation at one site only.

Genetic diversity mirrored genetic structure in each species, where demographically isolated sites had significantly lower genetic diversity. This structure and corresponding genetic diversity might reflect differing sensitivity to environmental factors such as pH and metals rather than dispersal ability alone. While sites from catchments separated by >100km were not differentiated genetically in all three species, particular acid and metal-rich sites had reduced genetic diversity and local isolation. *B. rhodani* was not present at these sites due to its acid sensitivity, but *A. sulcicollis* and *I. grammatica* were revealed to be genetically differentiated at these sites possibly due to adaption to local conditions. Since aquatic invertebrates are commonly used to investigate ecosystem health, the genetic structure of species present may reveal more detailed local environmental influences on organisms than can be gained from species assemblages.

Key words: population genetic structure, invertebrates, genetic diversity, environmental stressors

3.1 Introduction

Connectivity between populations can have a major effect on genetic diversity and structure, which in turn may influence the likelihood of populations recovering and responding to environmental change (Frankham *et al.* 2009; Alp *et al.* 2012). Within the freshwater environment connectivity is governed by stream network hierarchy, where fish and invertebrates without a terrestrial life stage are restricted to waterways. Dispersal within streams by drift and upstream larval movement have been well studied (Jackson *et al.* 1999; Elliott 2002; Petersen *et al.* 2004). Invertebrates with a terrestrial adult life stage, however, have the capacity of more complex dispersal patterns (Petersen *et al.* 2004). Long range dispersal of adult aquatic insects is particularly vital to recovery after a disturbance such as acidification (Bradley and Ormerod 2002b). Since acidification may affect large areas of a river system (e.g. Feeley *et al.* 2011) removing the potential of recolonization from upstream, recovery must, therefore, rely on dispersal from an unaffected (or less affected) reach.

Many studies have investigated the dispersal capabilities of adult aquatic invertebrates using traditional sampling methods involving, for example, malaise traps. Petersen *et al.* (2004) and Briers *et al.* (2002), investigating the dispersal of stoneflies and mayflies within the same geographic area as the current study, concluded that most adults remain close to the stream channel from which they emerged. Petersen *et al.* (2004), discovered that approximately 50% of stoneflies travelled less than 18 m, while 90% travelled less than 60 m. The authors also found that female mayflies travelled an even shorter distance, with approximately 50% the individuals caught within 7–11 m of the stream channel from which they emerged. Briers *et al.* (2002) also found that the numbers of adults declined sharply with distance from the stream; with 90% of adults caught within 11 m of the stream channel. These studies corroborate many other findings (Svensson 1974; Jackson and Resh 1989; Sode and Wiberg-Larsen 1993; Collier and Smith 1995; Kuusela and Huusko 1996;

Griffith *et al.* 1998; Petersen *et al.* 1999; Delettre and Morvan 2000), all suggesting that adult aquatic insects do not disperse far from their emergence stream.

Studying physical dispersal, however, may not reflect underlying gene-flow between populations. For example, these traditional studies could be overlooking leptokurtic dispersal, where most offspring breed very close to their parents but a small percentage breed a large distance away. The movement of just one migrant per generation into a subpopulation can be sufficient to minimize the loss of polymorphism and heterozygosity; however, the effective dispersal of up to 10 migrants per generation may not cause uniformity of allele frequencies across subpopulations (Mills and Allendorf 1996). Immigrants could have both a positive (i.e. rescue) and negative (i.e. outbreeding depression) effect on recipient populations (Tallmon *et al.* 2004; Turlure *et al.* 2014). Therefore, too many migrants per generation could be unfavourable and hinder the emergence of local adaptations (Éva Kisdi 2002).

There are also other factors which could affect gene flow, meaning it would not nessessarily reflect dispersal, such as if dispersers had reduced survival or reproductive success in the new location, possibly due to factors of the environment, e.g. certain stressors (Bensch *et al.* 1998; Hendry 2004; Nosil et al. 2005; Garant *et al.* 2007). The scientific community and regulatory agencies have become increasingly aware of the long-term impact of environmental stressors on the sustainability of ecosystems (Bickham *et al*. 2000). The effects of stressors on biodiversity can vary depending on the level of evolutionary adaption and phenotypic plasticity in the species studied (Buchwalter *et al*. 2008; Garbuz *et al.* 2008; Gutiérrez-Cánovas *et al.* 2013; Gutierrez-Canovas *et al.* 2015). This implies that some effect of these stressors is asserted at the genetic level (i.e has a genetic response), where stress can cause micro-evolutionary responses (Bradshaw and Holzapfel 2006; Higgins *et al.* 2014) leading to tolerance (Buchwalter *et al*. 2008; Garbuz *et al*. 2008; Gutierrez-Canovas *et al.* 2015). Environmental stressors, such as acidity and metals could therefore influence genetic structure and diversity. Despite the importance of genetic diversity on populations resilience however, few studies have addressed the effects of chemical contamination on population genetics (Bickham *et al.* 2000). Advances in genetic techniques can offer new insights on gene-flow, dispersal capabilities and biodiversity of aquatic invertebrates (Saura *et al.* 2014; Marchant *et al.* 2015). Many previous studies (e.g. Smith and Collier 2001;

Schultheis *et al.* 2008; Saito and Tojo 2016), have utilised allozyme and mtDNA markers; however, when investigating fine-scale genetic structure these markers have been shown to be insufficiently polymorphic to provide appropriate levels of resolution (Wilcock *et al.* 2001a; Teske *et al.* 2015). Studies using hypervariable molecular markers such as microsatellites, can however be used to assess fine scale genetic structure of invertebrates, and have in the past shown species to have both higher (Alp *et al.* 2012) and lower (Teske *et al.* 2015) dispersal potential than traditional methods or less informative genetic markers previously suggested.

The number of studies that have used microsatellites to assess dispersal and biodiversty in invertebrates is still relatively low, particularly when focusing on freshwater Insecta (Table 1.2). Microsatellites have been utilised to describe limited dispersal in a freshwater crustacean belonging to the *Parabathynellidae* family (Asmyhr *et al.* 2014); to reveal population structure within the Spring amphipod, *Wangiannachiltonia guzikae* (Robertson *et al.* 2014) and to investigate pelagic dispersal in marine invertebrate, *Pyura doppelgangera* (Teske *et al*. 2015). This latter study successfully showed that the species had limited dispersal ability despite poorly resolving genetic markers predicting limited genetic structuring along environmentally homogeneous coastlines.

There are also very few studies that compare the dispersal ability and genetic structure of more than one freshwater invertebrate in the same community using microsatellites. Wilcock *et al.* (2007) compared two caddisfly species, *Plectrocnemia conspersa* and *Plectrocnemia flavomaculatus*. Though both species have a terrestrial life stage, P. conspersa was found to disperse strongly, facilitating gene flow within and between catchments, whereas P. flavomaculatus was found to have strong genetic differentiation. Alp *et al*. (2012) studied the amphipod *Gammarus fossarum* (no adult flight stage) and the mayfly *Baetis rhodani* to compare the dispersal abilities between a life cycle that is purely aquatic to an insect that also has a terrestrial life stage. Alp *et al.* (2012) and Williams (2003) found that *B. rhodani* had very weak genetic structure across the catchment studied, despite traditional methods suggesting limited dispersal.

Here, the genetic structure and genetic diversity of three freshwater macroinvertebrates, *Amphinemura sulcicollis* (Nemouridae; Plecoptera), *Isoperla grammatica* (Perlodidae;

Plecoptera) and *Baetis rhodani* (Baetidae; Ephemeroptera), and their response to environmental stressors were compared. It is thought that this is the first study to investigate the genetic structure of three freshwater invertebrates simultaneously over a large spatial scale (>100 km) including multiple freshwater catchments. The aims of this study were to investigate the role of genetically effective dispersal in influencing population structure and how certain environmental stressors influence genetic diversity.

Amphinemura sulcicollis and *I. grammatica* have not previously been studied using microsatellite markers, whereas *B. rhodani* has been shown to have very weak genetic structure (Williams 2003; Rebora *et al.* 2005; Alp *et al.* 2012). In this study newly developed microsatellites, in addition to seven previously described markers for *B. rhodani* (Williams *et al.* 2002) were used. A larger number of loci was needed as it was unknown at the outset how many markers would be required to distinguish between recently diverged populations, as well as the fact that extra loci increase statistical power (Haasl and Payseur 2011; Putman and Carbone 2014). Many studies choose a minimum of ten polymorphic loci when investigating population structure (Cervini *et al.* 2006; Östergren *et al.* 2015).

All three species used in this study have aquatic larval and terrestrial adult life stages however they differ in body size, functional group and trophic level (Table 1.1), which may affect their dispersal ability and response to stressors, as in Wilcock *et al.* (2007), hence their ability to recolonise areas after a disturbance. *Amphinemura sulcicollis* is predominantly a shredder with faculty also for grazing and detritivory; *B. rhodani* is a generalist detritivore or scraper/grazer whereas *I. grammatica* is a predator. The species studied also differ in their sensitivity to salt (Kefford *et al.* 2003), metals (for example cadmium (Buchwalter and Luoma 2005; Buchwalter *et al.* 2008)), and acidity, where Ephemeroptera (particularly Baetid) species have been shown to be more sensitive than Plecoptera (Kowalik and Ormerod 2006; Murphy *et al.* 2013; Murphy *et al.* 2014). This could affect their relative survival at any given location therefore their genetic structure.

It is hypothesised that differences in size, functional group and sensitively to stressors between the species studied will affect their corresponding genetic structure, as is often observed even between closely related species (Wilcock *et al.* 2007). Specifically, as body size may affect dispersal potential (Gutiérrez *et al.* 1997) it is hypothesised that the smallest

species *A. sulcicollis* could have lower dispersal capability and corresponding gene flow. Dispersal capability is only one factor effecting gene flow however, therefore it is hypothesised that *B. rhodani's* sensitivity to acidic environments will affect this species structure as it may not be able to persist at acidic sites, as opposed to *A. sulcicollis* and *I. grammatica*, which are thought to be more tolerant (Kowalik and Ormerod 2006; Murphy *et al.* 2013; Murphy *et al.* 2014), leading to differences among the species studied.

3.2 Methods

3.2.1 Sampling and genotyping

Individuals of the three target species, *Amphinemura sulcicollis*, *Isoperla grammatica* and *Baetis rhodani* were collected from sites in upland Wales and DNA was extracted as described in Chapter 2, sections 2.2.1 and 2.2.2 (Table 2.1a-c, Figure 2.2, 2.3 and 2.4). Samples were genotyped at 10, 10 and 13 loci (*A. sulcicollis*, *I. grammatica* and *B. rhodani* respectively, see section 2.5.1, following PCR protocols detailed in Table 2.5 and Table 2.6). Prior to this, *B. rhodani* individuals were DNA barcoded to remove cryptic species (as detailed in section 2.2.3.3). Samples were sent for fragment analysis (Dundee Biosciences, UK) using ROX500 as the internal size standard. Fragment analysis products were visualised and scored using Genemarker v 1.91 (Holland and Parson 2011) and allele sizes were binned by hand (section 2.5.3). The complete genotype dataset analysed in this chapter for each species can be found in Appendix F.

3.2.2 Stressors

Chemistry data, collected in 2012 and 2013 during the DURESS project, was made available to compare with this study's genetic data (Appendix H, Table H1). Four stressors were chosen: aluminium (Al, µg/l), cadmium (Cd, µg/l), pH (pH-units) and total oxidised nitrogen (TON, mg/l). Each is an important stressor in the aquatic environment and key drivers of biological effects (Kowalik *et al.* 2007). Acidification is a well-known, well studied, stressor

for aquatic invertebrates (Ormerod and Jenkins 1994; Kowalik *et al.* 2007), particularly *Baetis rhodani* (Kowalik and Ormerod 2006). Acidification can be naturally occurring (Gutierrez-Canovas *et al*. 2015) or streams could be acidified by anthropogenic factors such as air-borne pollution and land-use change i.e. conifer plantations (Feeley *et al.* 2011). Metal-sensitivity has often been observed in aquatic invertebrates to the extent that it can be used to predict species composition, and has been known to differ between Plecoptera and Ephemeroptera (Clements *et al.* 2000; Buchwalter *et al.* 2008). Elevated levels of TON (total oxidised nitrogen, predominantly nitrate) can be toxic to aquatic invertebrates (Camargo *et al.* 2005). Nitrate toxicity may differ depending on body size, effecting those with a smaller body size more than larger. Effects are also dependent on water salinity, and environmental adaptation (Camargo *et al.* 2005) and could be a consequence of land-use intensification (i.e. fertiliser run off).

3.3 Statistical analysis

To reduce the risk of overestimating population structure, each dataset were tested for relatedness (hence genetic non-independence) among individuals within sites prior to the main analysis. This was performed using COLONY v 2.0.5.8 (Jones and Wang 2010) assuming male and female polygamy, with inbreeding and without parthenogenesis, using a pairwise, full-likelihood analysis with a very high likelihood precision (4 threads). Within each site the 'best most likely full sibling family', 'full sibling' and 'half sibling' were investigated. While in general low levels of relatedness were found, all individuals shown to be half siblings or more closely related with a p-value of 0.8 or greater were removed. This resulted in 39 genetically non-independent (i.e. related) individuals being removed from 13 sites (up to five per site) for *A. sulcicollis*; two individuals from two sites, (1 per site) were removed from *I. grammatica*, and 20 individuals removed from 9 sites (up to 4 individuals per site) from *B. rhodani* dataset.

Subsequent analysis was performed 'per site' as well as 'per cluster'. For all 'per site' analysis only sites that contained more than 10 individuals were used (Chapter 2, Table 2.1(a-c)). Analysis was repeated 'per cluster' for Hardy-Weinberg analysis, descriptive statistics, F_{IS} and genetic diversity indices. This was because if population structure is found

within the species, the 'per site' approach might potentially be unrepresentative of true populations or demes (Dharmarajan *et al.* 2013). The inadvertent combination of distinct gene-pools in population genetic analysis can cause high levels of Hardy-Weinberg disequilibrium due to Wahlund Effect (Johnson and Black 1984; Waples 1990), therefore the analysis was repeated instead analysing Bayesian clusters of individuals as the population unit. These clusters corresponded to results of each species population genetic structure (see Section 3.4.2).

The analysis described above was then repeated removing all loci that were consistently out of Hardy-Weinberg equilibrium (HWE) to establish the extent of their effect on the results. In chapter 2 it was identified that certain loci were consistently out of HWE (as summarised in Table 2.10a-c) and those showing departure from HWE in more than 50% of the sites tested were removed. For *A. sulcicollis* Amp_10, Amp_11 and Amp_13 were removed. Within *I. grammatica*, there were many loci consistently out of HWE (Iso_4, Iso_5, Iso_6, Iso_7, Iso_8, Iso_9 and Iso_10), leaving the reduced dataset with only three loci that were in HWE for more than 50% of sites, and for *B. rhodani* four of thirteen loci were removed (Brh-1, Brh-2, Brh-3 and B_5; see section 2.6.3 Table 2.10(a-c)).

3.3.1 HWE and descriptive statistics

Descriptive statistics per site (averaged across loci) are reported, including number of alleles (Na) and observed and expected heterozygosity (Ho and uHe), calculated in GenAIEx v. 6.5 (Peakall and Smouse 2006; Peakall and Smouse 2012). Deviations from Hardy–Weinberg equilibrium (HWE) (specifically the global test for heterozygosity deficient) and tests for linkage disequilibrium (LD) were estimated using GENEPOP v. 4.0.1 (Rousset 2008). Tests for significance were based on 10,000 de-memorization steps, 100 batches and 5,000 iterations per batch for LD and 10,000 de-memorization steps, 20 batches and 5,000 iterations per batch for HWE. For HWE and LD, significance levels were adjusted using Bonferroni correction (Rice 1989). Inbreeding coefficients (F_{IS}) and their associated p values (deviation from zero) per site, were calculated in ARLEQUIN v 3.1 (Excoffier and Lischer 2010). Per loci, per site descriptive statistics are reported in Appendix G and summarised in Table 2.10(a-c).

3.3.2 Genetic structure

To investigate genetic structure and barriers to dispersal between study sites, analysis of isolation by distance and Bayesian clustering approaches were adopted. Genetic differentiation (F_{ST}) was estimated between all pairs of locations using ARLEQUIN v 3.1 (Excoffier and Lischer 2010). Isolation by distance was tested in GenAlEx (Peakall and Smouse 2006; Peakall and Smouse 2012) using a Mantel test to evaluate the correlation between genetic differentiation (F_{ST}), calculated in ARLEQUIN, and geographic distance, calculated in GenAIEx, with significance determined by 999 permutations. Also in ARLEQUIN an Analysis of Molecular Variance (AMOVA) (Excoffier *et al.* 1992) was preformed to investigate where the greatest variation lay, within individuals, among individuals or among populations.

Bayesian clustering analysis was performed using the program STRUCTURE v. 2.3.4 (Pritchard *et al.* 2000). The number of clusters (K) in each dataset were determined by implementing the programme with K values varying from 1 to 10, with ten independent runs for each K value, all with flat prior probability distributions. To determine most appropriate K value, burn-in Markov Chain Monte Carlo (MCMC) (Karandikar 2006) replication was set to 100,000 and data were collected over 1,000,000 MCMC replications in each run. All STRUCTURE plots presented are the result of the average of the ten replicates, using all available sites (including partial sites), using the program CLUMP v 1.1.2 (Jakobsson and Rosenberg 2007). Finally, STRUCTURE HARVESTER (Earl and vonHoldt 2012) was used to identify the most likely number of clusters (value of K) using an ad hoc statistic ΔK based on the rate of change in the log probability of data between successive K values (Evanno *et al.* 2005).

3.3.3 Genetic diversity correlations

3.3.3.1 Genetic diversity

Allelic richness (accounting for differences in sample sizes among sites) was estimated for each locus per site using the program FSTAT (version 2.9.3.2, http://en.bio-

soft.net/tree/FSTAT.html, accessed 14 March 2016). Values were compared among sites using a linear mixed model (controlling for locus) using the program RStudio (R version 3.0.2, https://www.rstudio.com/products/rstudio/, accessed 14 March 2016). The library 'nlme' was used to generate p-values. The model was re-levelled so each separate site was run as the reference location, therefore comparing every site to each other (R script in Appendix J). In order to visualise genetic diversity across the region, interpolation maps were created using QGIS v. 2.14.2 (QGIS Development Team 2016) for each species. Mean allelic richness values were used per site and for sites that had values for 2012 and 2013, an average was used so there was one value per site.

3.3.3.2 Stressors

To test whether genetic diversity is influenced by environmental stressors, correlations between measures of genetic diversity and each stressor were calculated using a linear regression in RStudio. Genetic diversity estimates comprised mean allelic richness, as calculated above (averaged across all loci); and expected heterozygosity (as calculated in section 3.3.1). These measures were chosen because they differ in their predicted sensitivity to population decline, with expected heterozygosity expected to take longer to reflect a population bottleneck than allelic diversity (Hoban *et al.* 2014), therefore using both was expected to provide different levels of information. Average expected heterozygosity (H_E) as opposed to observed heterozygosity (H_o) was used since this parameter is less affected by departures from Hardy–Weinberg equilibrium and sample size (Nei 1987; Johannesson and AndrÉ 2006). Unfortunately, chemistry data were not available for all sites (data available for which site is shown in Appendix H, Table H1). Where data were available at the same site for 2012 and 2013, an average was used.

3. 4 Results

3.4.1 HWE and descriptive statistics

Amphinemura sulcicollis, *I.* grammatica and *B. rhodani* were all characterised by very high genetic diversity, particularly *I. grammatica* (for per locus analysis see section 2.6.3, Table 2.10a-c and Appendix G). Mean number of alleles (Na) per site, across loci, ranged from 7.4 - 10.8, 16.3 - 20.6, and 8.5 - 10.8 for *A. sulcicollis*, *I.* grammatica and *B. rhodani*, respectively (Table 3.1a-c).

All species featured a very strong pattern of HW disequilibrium, with all sites departing significantly from HWE and having significant F_{IS} values for all species (Table 3.1a-c). This was due to high levels of heterozygote deficiency, which could have been driven by loci that were consistently out of HWE. However, when the analysis was repeated without these loci, similar patterns were found (Appendix I), for *A. sulcicollis*, though four sites were in equilibrium (109, 108, 112 and 116), all other sites continued to be out of HWE (Appendix I, Table I1a). A similar pattern for F_{IS} was found, though six sites had non-significant values when the problem loci were removed (109, 108, 112, 116, 95 and 6), all other sites still showed departure from equilibrium.

Similarly for *I. grammatica*, though four sites changed (108, 115, 118 and 93), all other sites continued to be significantly out of HWE, and only two sites (108 and 93) had non-significant F_{IS} values when using the reduced dataset (Appendix I, Table I1b). For *B. rhodani* using a reduced loci dataset influenced two sites (97 and 115) but all others were still significantly out of HWE, and site 115 was the only site to not have a significant F_{IS} value (Appendix I, Table I1c). No linkage disequilibrium was observed in any species.

When samples were re-analysed according to Bayesian clustering results (see section 3.3.2 for details), this had no effect on the pattern of HW disequilibrium or F_{IS} in any species (Table 3.1a-c). Neither did performing analysis according to Bayesian clustering and with a reduced dataset (Appendix I, Table I1a-c).

AMOVA for all three species revealed as expected that much of the variance was partitioned within individuals (80.23% p-value = 0.000+-0.000, 69.21% p-value = 0.000+-0.000, 74.82% p-value = 0.000+-0.000 for *A. sulcicollis*, *I.* grammatica and *B. rhodani*, respectively),

compared to among individuals (17.1% p-value = 0.000+-0.000, 30.04% p-value = 0.000+-0.000, 24.59% p-value = 0.000+-0.000) and among sites (2.67% p-value = 0.013+-0.004, 0.75% p-value 1.000+-0.000, 0.6% p-value = 1.000+-0.000). This reflects each species high polymorphism rates and comparatively low levels of differentiation among populations.

Table 3.1a-c. Microsatellite diversity in a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*, using the full dataset. Mean (across loci) and standard error (SE) values calculated per site. N = number of individuals; Na = number of alleles; Ho = observed heterozygosity; uHe = expected heterozygosity; all calculated in GenAIEx. HWE = p-value of Hardy-Weinberg Equilibrium test calculated per site using Genepop, F_{IS} = the inbreeding co-efficient and p = the F_{IS} p value. **Bold** represents significance (critical p-value for HWE after Bonferroni correction = ≤0.005, ≤0.005, ≤0.0038, for *A. sulcicollis*, *I. grammatica* and *B. rhodani*, respectively).

Table 3.1a. *Amphinemura sulcicollis*.

Site		N	Na	Ho	uHe	HWE	F_{IS}	p
104 (Cluster 1)	Mean	17.0	10.8	0.676	0.842	**0.000**	0.222	**0.000**
	SE	0.0	1.1	0.090	0.030			
109	Mean	15.9	9.5	0.677	0.837	**0.000**	0.141	**0.000**
	SE	0.1	0.8	0.080	0.022			
108	Mean	15.8	10.1	0.696	0.806	**0.000**	0.123	**0.000**
	SE	0.2	1.1	0.061	0.037			
102	Mean	17.0	8.5	0.612	0.793	**0.000**	0.208	**0.000**
	SE	0.0	1.1	0.064	0.034			
112	Mean	17.0	10.4	0.647	0.792	**0.000**	0.158	**0.000**
	SE	0.0	1.5	0.071	0.045			
96	Mean	14.9	8.3	0.616	0.786	**0.000**	0.163	**0.000**
	SE	0.1	1.0	0.077	0.030			
9	Mean	17.9	9.5	0.630	0.790	**0.000**	0.173	**0.000**
	SE	0.1	1.3	0.064	0.037			
116	Mean	18.2	9.5	0.631	0.768	**0.000**	0.123	**0.002**
	SE	0.8	1.4	0.079	0.059			
113	Mean	15.9	9.5	0.666	0.823	**0.000**	0.174	**0.000**
	SE	0.1	0.9	0.061	0.029			
95	Mean	15.9	8.4	0.660	0.796	**0.000**	0.166	**0.001**
	SE	0.1	0.7	0.075	0.020			
6	Mean	16.9	7.4	0.573	0.720	**0.000**	0.174	**0.000**
	SE	0.1	0.7	0.059	0.040			
59	Mean	18.0	10.8	0.650	0.822	**0.000**	0.193	**0.000**
	SE	0.0	1.1	0.076	0.031			
93 (Cluster 4)	Mean	16.0	8.3	0.600	0.748	**0.000**	0.197	**0.000**
	SE	0.0	0.9	0.064	0.049			
Cluster 2	Mean	163.5	20.0	0.639	0.817	**0.000**	0.174	**0.000**

Site		N	Na	Ho	uHe	HWE	F_{IS}	p
	SE	1.4	4.0	0.059	0.030			
Cluster 3	Mean	38.4	14.0	0.622	0.819	**0.000**	0.190	**0.000**
	SE	0.3	1.7	0.065	0.034			

Table 3.1b. *Isoperla grammatica*.

Site		N	Na	Ho	uHe	HWE	F_{IS}	p
105	Mean	15	16.3	0.663	0.943	**0.000**	0.308	**0.000**
	SE	0	1.0	0.055	0.009			
106	Mean	20	19.4	0.640	0.939	**0.000**	0.329	**0.000**
	SE	0	0.9	0.069	0.010			
108	Mean	19	19.4	0.692	0.945	**0.000**	0.291	**0.000**
	SE	0	1.9	0.084	0.012			
12	Mean	20	19.2	0.628	0.936	**0.000**	0.338	**0.000**
	SE	0	1.4	0.075	0.015			
97	Mean	17	18.0	0.585	0.948	**0.000**	0.393	**0.000**
	SE	0	1.0	0.089	0.007			
98	Mean	20	19.4	0.688	0.946	**0.000**	0.291	**0.000**
	SE	0	1.2	0.057	0.009			
112	Mean	20	20.0	0.666	0.949	**0.000**	0.305	**0.000**
	SE	0	1.2	0.066	0.008			
114	Mean	19	18.5	0.671	0.939	**0.000**	0.306	**0.000**
	SE	0	1.0	0.052	0.009			
115	Mean	20	20.6	0.700	0.948	**0.000**	0.278	**0.000**
	SE	0	1.4	0.057	0.009			
10	Mean	20	19.9	0.705	0.942	**0.000**	0.268	**0.000**
	SE	0	1.5	0.054	0.011			
118	Mean	20	19.1	0.641	0.937	**0.000**	0.328	**0.000**
	SE	0	1.6	0.062	0.011			
93 (Cluster 2)	Mean	19	17.3	0.710	0.883	**0.000**	0.213	**0.000**
	SE	0	1.5	0.064	0.029			
Cluster 1	Mean	212.7	55.5	0.663	0.952	**0.000**	0.265	**0.000**
	SE	1.9	3.3	0.057	0.008			

Table 3.1c. *Baetis rhodani*.

Site		N	Na	Ho	uHe	HWE	F_{IS}	p
112	Mean	16.0	8.8	0.572	0.805	**0.000**	0.296	**0.000**
	SE	0.0	0.8	0.064	0.029			
102	Mean	19.6	10.8	0.616	0.815	**0.000**	0.266	**0.000**
	SE	0.2	1.1	0.054	0.024			
97	Mean	15.8	9.4	0.612	0.781	**0.000**	0.229	**0.000**
	SE	0.2	1.4	0.083	0.034			
96	Mean	16.9	10.6	0.618	0.818	**0.000**	0.255	**0.000**
	SE	0.1	1.2	0.062	0.029			
9	Mean	10.6	8.5	0.640	0.812	**0.000**	0.253	**0.000**
	SE	0.2	0.9	0.049	0.028			
106	Mean	17.8	9.1	0.576	0.786	**0.000**	0.281	**0.000**
	SE	0.1	1.2	0.053	0.027			
113	Mean	15.9	9.4	0.640	0.778	**0.000**	0.186	**0.000**
	SE	0.1	1.2	0.070	0.047			
115	Mean	15.8	10.2	0.640	0.813	**0.000**	0.235	**0.000**
	SE	0.2	1.2	0.072	0.031			
94	Mean	15.9	9.4	0.607	0.798	**0.000**	0.249	**0.000**
	SE	0.1	1.2	0.063	0.032			
118	Mean	19.8	9.8	0.640	0.807	**0.000**	0.219	**0.000**
	SE	0.1	1.0	0.063	0.029			
Cluster 1	Mean	164.3	23.5	0.616	0.811	**0.000**	0.233	**0.000**
	SE	0.5	3.1	0.056	0.026			

3.4.2 Genetic structure

Significant genetic differentiation was evident for some site comparisons when pairwise F_{ST} was investigated. Within *A. sulcicollis* significant differentiation was found between site 93 and all other sites, and sites 95 and 6 were also significantly different from all other sites (but not each other, as expected because they represent temporal samples from the same location, Table 2.1a). Site 104 was significantly different to eight other sites, and site 116 was significantly different to all sites except 9 (Table 3.2a). These patterns were recapitulated using the reduced dataset, although there were fewer significant differences, consistent with a lower number of markers limiting statistical power (Appendix I, Table I2a). Within *I. grammatica*, only the northern site 93 was significantly differentiated from all other sites (Table 3.2b), and using only three loci (the reduced dataset) this pattern remained (Appendix I, Table I2b). *B. rhodani* showed no sites that were consistently significantly differentiated from all other sites; however sites 113, and 94 were significantly different from five other sites, and site 106 was significantly different to three (Table 3.2c). When a reduced dataset were used this pattern was strengthened in sites 94 and 106 but reduced in site 113 (Appendix I, Table I2c).

A Mantel test for isolation-by-distance identified a significant, although weak, correlation between genetic and geographic distances for *A. sulcicollis* ($R^2 = 0.09$, $p = 0.030$), but was not significant for *I. grammatica* ($R^2 = 0.10$, $p = 0.119$) or *B. rhodani* ($R^2 = 0.01$, $p = 0.227$).

Bayesian clustering supported the F_{ST} results for each species. Though STRUCTURE identifies clusters of individuals not pertaining to sites as flat priors were used, results tended to group individuals from the same site. Figure 3.1a-c shows STRUCTURE plots assuming two, three and four clusters (K = 2 - 4) for *A. sulcicollis*. Three (Figure 3.1.b) was identified as the most probable number of clusters, indicated by three clusters having the largest rate of change (Delta K, ΔK) and the highest mean likelihood (mean LnP(K)) (Evanno *et al.* 2005) using the program STRUCTURE HARVESTER (Earl and vonHoldt 2012). The most clearly defined cluster for *A. sulcicollis* comprised all individuals from sites 95 and 6 as well as individuals from sites 94 and 4 (henceforth Cluster 3). Other clusters well supported with F_{ST} values but more subtle within STRUCTURE comprised individuals from site 104 (Cluster 1), individuals from site 93 (Cluster 4); despite the STRUCTURE results it was decided not to

group these two sites but to analyse them as separate clusters based on their significant F_{st} results and on their respective genetic diversity (see section 3.4.3.1) which marks these two sites as unique. All remaining individuals were admixed and grouped in to an *ad hoc* grouping, Cluster 2 (Figure 3.1a-c).

The F_{ST} results for *I. grammatica* suggested that site 93 (Cluster 2) was genetically distinct from the other sites; this was mirrored in the STRUCTURE results, individuals from all other sites were admixed and grouped into Cluster 1. In line with this, two was identified to be the most likely number of clusters using STRUCTURE HARVESTER (Figure 3.2a-c). In contrast, within *B. rhodani* no difference in cluster proportions could be reliably determined for any of the geographic regions analysed (Figure 3.3a-c).

Based on these results the data were re-analysed according to Bayesian population structure. All individuals were included since there was no significant effect comparing results including all individuals, to results using only individuals that had a cluster assignment (q-value) of greater than 0.8.

F_{ST} analysis showed, as expected, that all genetic clusters in *A. sulcicollis* were significantly differentiated from each other (Table 3.3a). However, for the reduced loci dataset, Cluster 1 and 2 were no longer significantly differentiated, which is supported by the observation that this was the least well supported cluster (Appendix I, Table I3a). For *I. grammatica*, whether using the full or reduced dataset Cluster 1 and 2 remained significantly differentiated (Table 3.3b, Appendix I, Table I3b).

Table 3.2a-c. Pairwise genetic differentiation (F_{ST}, distance method) of a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*, between sites, using the full dataset. Values with a p value of <0.05 in **bold**.

Table 3.2a. *Amphinemura sulcicollis*.

	104	109	108	102	112	96	9	116	113	95	6	59	93
104	0												
109	0.011	0											
108	**0.020**	0.009	0										
102	**0.026**	0.008	**0.031**	0									
112	**0.018**	0.012	0.007	**0.026**	0								
96	0.014	0.002	0.017	0.010	0.009	0							
9	**0.027**	0.007	0.012	**0.028**	0.009	0.011	0						
116	**0.037**	**0.026**	**0.035**	**0.030**	**0.019**	**0.027**	0.011	0					
113	0.007	0.001	0.007	0.010	-0.004	-0.004	0.008	**0.020**	0				
95	**0.064**	**0.051**	**0.056**	**0.055**	**0.067**	**0.049**	**0.052**	**0.076**	**0.047**	0			
6	**0.088**	**0.074**	**0.069**	**0.077**	**0.082**	**0.070**	**0.059**	**0.089**	**0.072**	0.002	0		
59	**0.019**	0.005	**0.025**	0.013	0.014	0.002	0.015	**0.027**	0.005	**0.043**	**0.072**	0	
93	**0.047**	**0.050**	**0.057**	**0.042**	**0.028**	**0.031**	**0.040**	**0.032**	**0.020**	**0.083**	**0.102**	**0.030**	0

Table 3.2b. *Isoperla grammatica*.

	105	106	108	12	97	98	112	114	115	10	118	93
105	0											
106	0.010	0										
108	0.011	0.012	0									
12	0.009	0.011	**0.015**	0								
97	0.008	**0.016**	0.013	0.015	0							
98	0.006	0.011	0.005	0.010	0.011	0						
112	0.009	0.010	0.009	0.009	0.004	0.007	0					
114	0.010	**0.015**	**0.014**	0.011	0.014	0.012	0.008	0				
115	0.014	0.008	0.011	**0.014**	0.009	0.010	0.006	0.012	0			
10	0.011	0.011	0.010	**0.014**	**0.015**	0.008	0.007	0.013	0.010	0		
118	0.006	0.004	0.011	0.009	0.009	0.009	0.008	0.010	0.007	0.008	0	
93	**0.041**	**0.038**	**0.04**	**0.048**	**0.041**	**0.043**	**0.040**	**0.046**	**0.039**	**0.030**	**0.031**	0

Table 3.2c. *Baetis rhodani*.

	112	102	97	96	9	106	113	115	94	118
112	0									
102	0.004	0								
97	0.010	0.008	0							
96	0.006	0.012	0.007	0						
9	0.018	0.014	0.021	0.014	0					
106	0.018	**0.018**	0.017	0.011	**0.025**	0				
113	0.016	**0.031**	**0.021**	0.014	**0.024**	-0.001	0			
115	0.004	0.008	0.004	0.004	0.012	**0.019**	**0.019**	0		
94	**0.027**	0.016	**0.026**	0.021	**0.026**	-0.004	0.009	**0.023**	0	
118	0.005	0.006	0.015	0.002	**0.021**	0.016	**0.017**	0.010	**0.022**	0

Table 3.3a-b. Pairwise genetic differentiation (F_{ST}, distance method) of a) *Amphinemura sulcicollis*, b) *Isoperla grammatica*, between clusters, using the full set of loci. Values with a p value of <0.05 (*) in **bold**.

Table 3.3a. *Amphinemura sulcicollis*.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1	0			
Cluster 2	**0.016**	0		
Cluster 3	**0.068**	**0.047**	0	
Cluster 4	**0.047**	**0.028**	**0.080**	0

Table 3.3b. *Isoperla grammatica*.

	Cluster 1	Cluster 2
Cluster 1	0	
Cluster 2	**0.036**	0

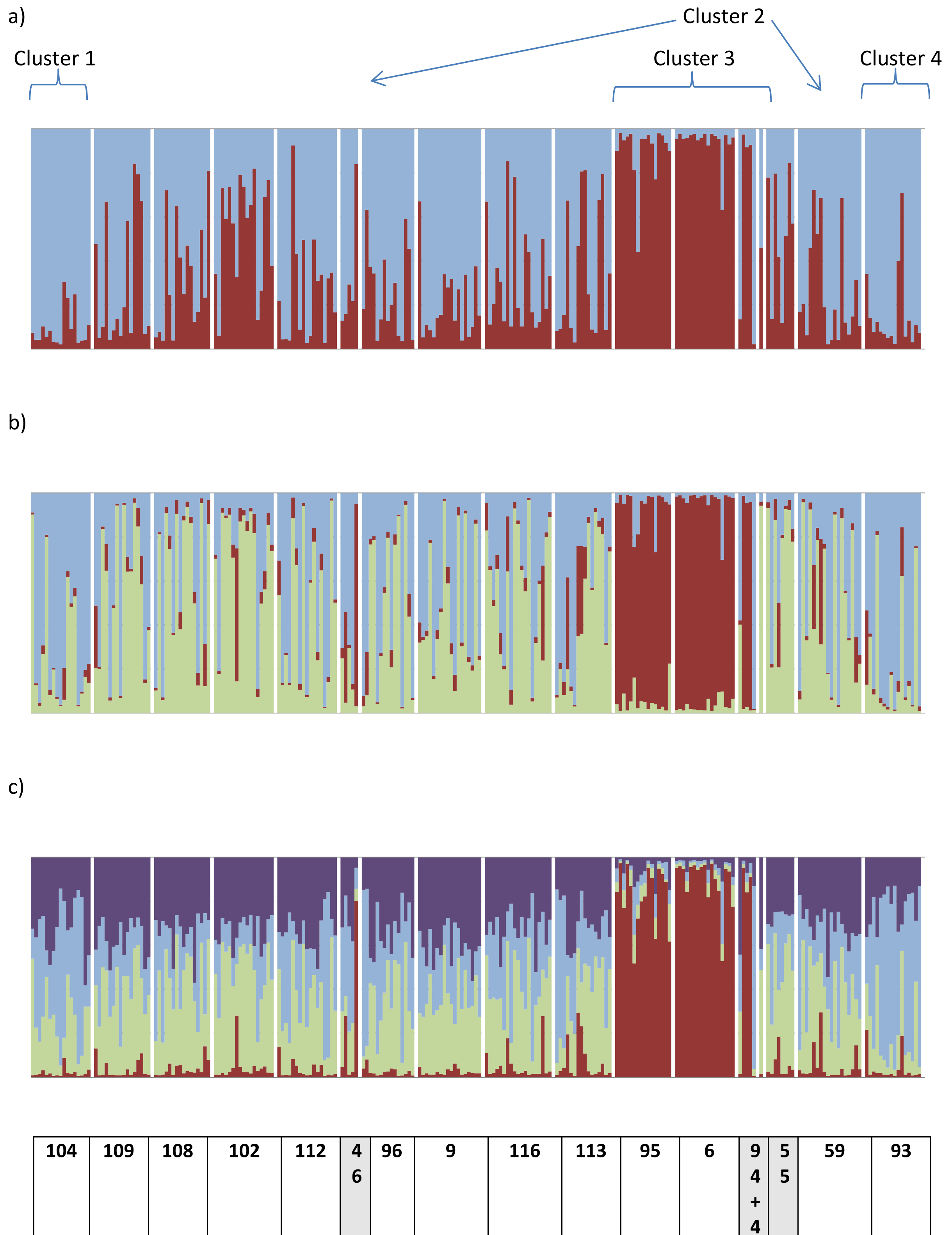


Figure 3.1. *Amphinemura sulcicollis* STRUCTURE plots showing a) two clusters (K=2), b) three clusters (K=3) and c) four clusters (K=4). White bars within plot divide different sites, site numbers shown below (grey shows partial sites), sites ordered from south to north of Wales.

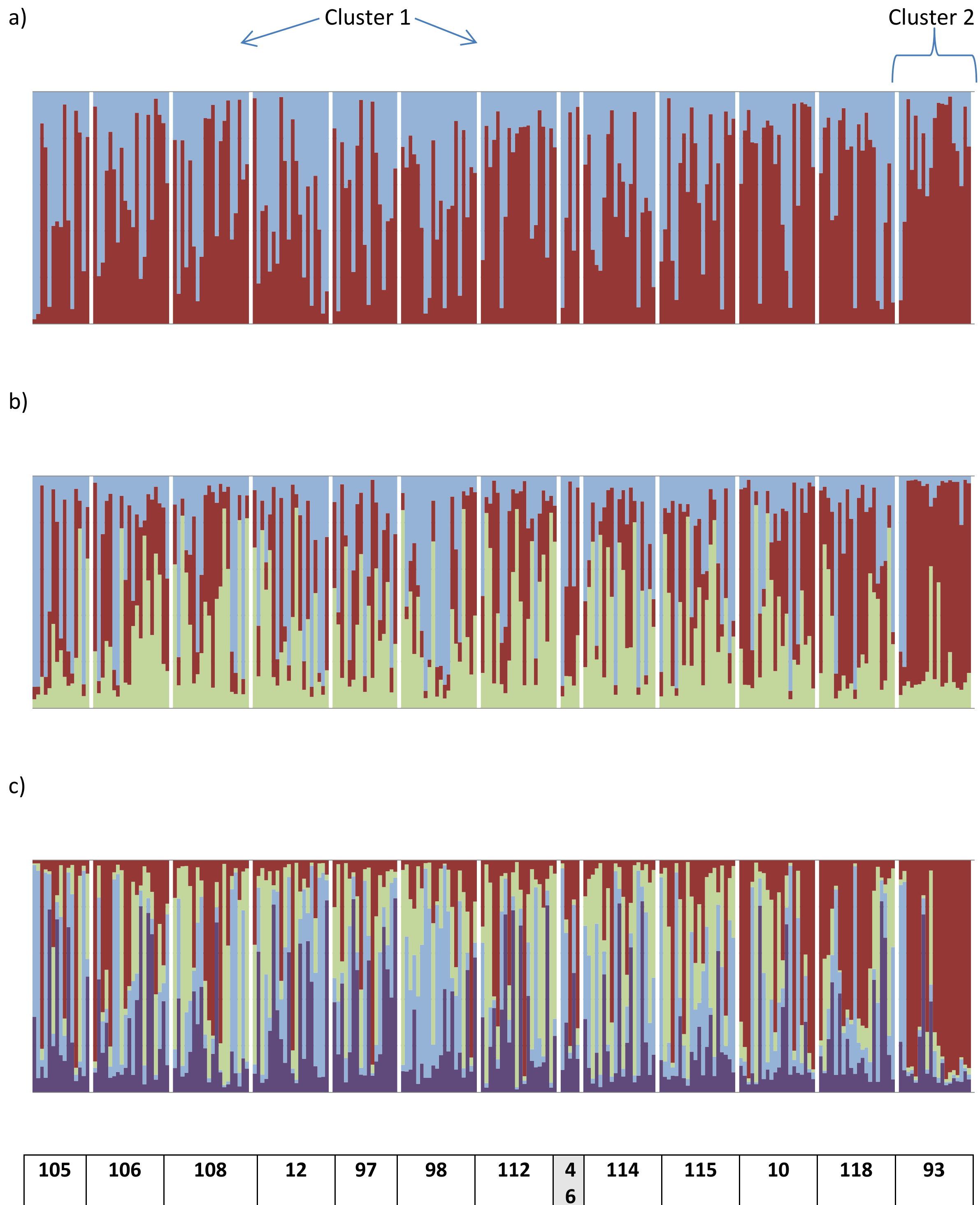


Figure 3.2. *Isoperla grammatica* STRUCTURE plots showing a) two clusters (K=2), b) three clusters (K=3) and c) four clusters (K=4). White bars within plot divide different sites, site numbers shown below (grey shows partial sites), sites ordered from south to north of Wales.

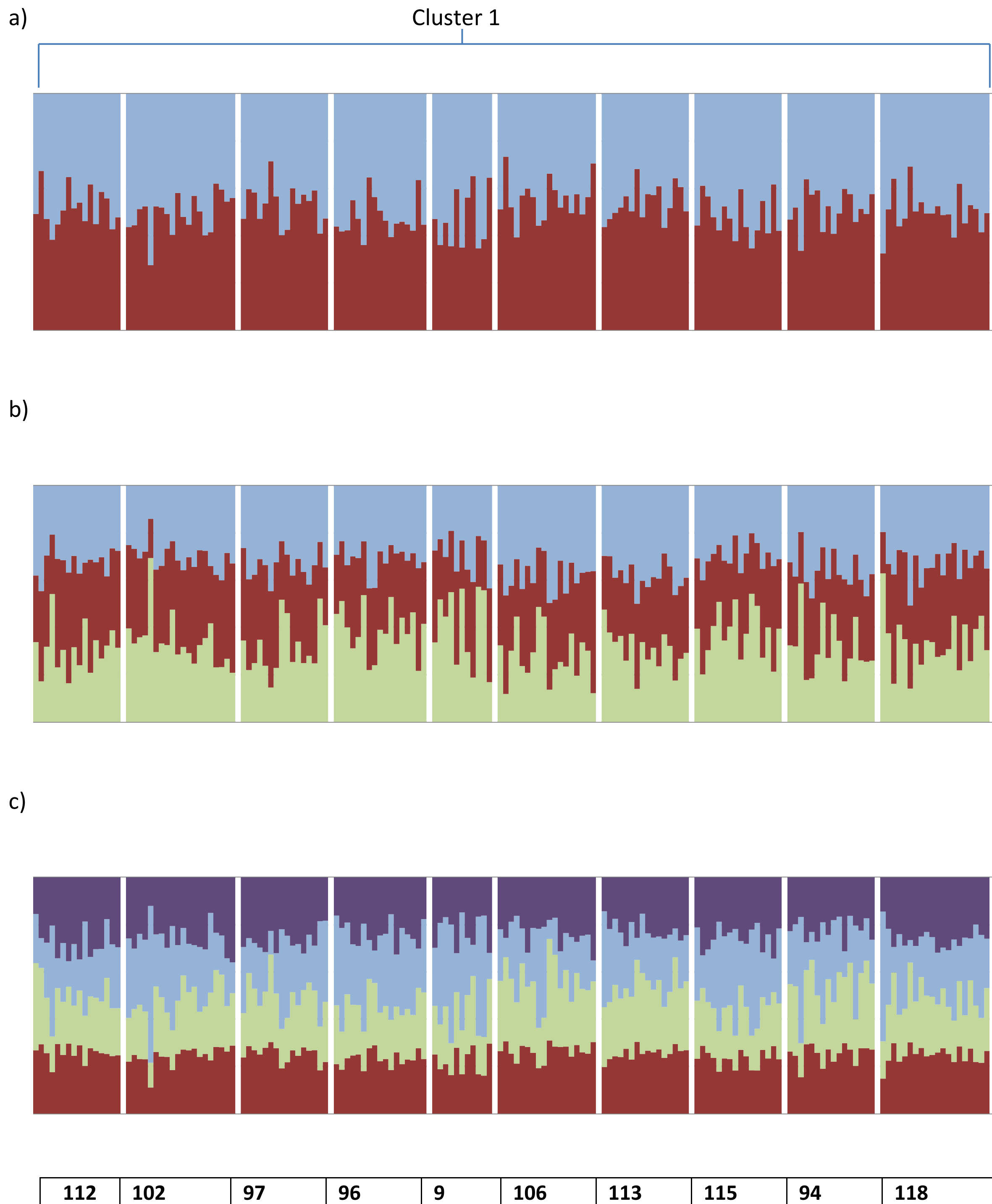


Figure 3.3. *Baetis rhodani* STRUCTURE plots showing a) two clusters (K=2), b) three clusters (K=3) and c) four clusters (K=4). White bars within plot divide different sites, site numbers shown below (grey shows partial sites), sites ordered from south to north of Wales.

3.4.3 Genetic diversity correlations

3.4.3.1 Genetic diversity

Allelic richness (accounting for different sample sizes; Appendix H, Table H2), estimated for each species had a mean value that ranged between 6.3 – 9.0, 14.6 - 17.0, and 6.5 - 7.4, with a minimum sample size of 11, 14, and 8 for *A. sulcicollis*, *I.* grammatica and *B. rhodani*, respectively. Although *I. grammatica* has the highest mean estimates, both *I. grammatica* and *A. sulcicollis* had similar ranges of mean allelic richness across their respective sites, whereas *B. rhodani* had the least variation among sites. Figure 3.4(a-c) demonstrates how allelic richness of each species varies over the geographical range. Results were very similar with the reduced dataset, ranging between 7.2 - 9.8, 14.4 - 19.2 and 6.7 - 8.5 with a minimum sample size of 15, 15, and 8 for *A. sulcicollis*, *I.* grammatica and *B. rhodani*, respectively (Appendix I, Table I4).

When mean allelic richness was compared between sites, significant differences were found, particularly for *A. sulcicollis* and *I. grammatica* (Table 3.4a-c). For *A. sulcicollis*, site 104 (Cluster 1) had the highest allelic richness, significantly higher than seven other sites (Table 3.4a, Figure 3.4a), whereas with the exception of site 59, the northern sites 93 (Cluster 4), 95 and 6 (95 and 6 representing the same site at different years, and forming part of Cluster 3) had the lowest allelic diversity, site 6 being significantly lower than nine other sites (Table 3.4a, Figure 3.4a). These patterns are maintained when repeated with the reduced dataset, although there were fewer significant differences, the same patterns were seen for site 104 which remained the highest (significantly higher than five other sites); site 6 (significantly lower than six other sites) and site 93 (significantly lower than three sites instead of five) had the lowest genetic diversity. However there was a marked difference when comparing site 95 to others because with the reduced dataset it was only significantly lower than site 104 whereas is was significantly different to four other sites using the full set of loci (Appendix I, Table I5a). When analysing the data as Bayesian clusters, the genetic diversity of Cluster 1 (site 104) remained the highest, although it was not significantly different to the *ad hoc* Cluster 2 (Figure 3.1), whereas Northern clusters, Cluster 3 (individuals from sites 95, 6, 94 and 4) and Cluster 4 (site 93) featured significantly lower genetic diversity than Cluster 1 and *ad hoc* Cluster 2 (Table 3.5a). This pattern was retained

when using the reduced dataset, except that Cluster 2 (the *ad hoc* cluster) no longer showed significantly greater allelic richness than Cluster 3 (Appendix I, Table I6a).

Within *I. grammatica*, the northern site 93 had the lowest mean allelic richness, significantly lower compared to six other sites (Table 3.4b, Figure 3.4b), this was maintained when loci out of HWE are removed, as site 93 still possessed significantly lower allelic richness than four other sites (Appendix I, Table I5b). When comparing clusters, Cluster 2 (site 93) possessed significantly higher allelic richness than Cluster 1 (mixed) whether using full (Table 3.5b) or reduced (Appendix I, Table I6b) dataset.

Within *B. rhodani* there were, as expected, fewer significant differences in allelic richness, however site 106 had significantly lower values than the four other sites, with site 96 possessing the highest value, however it was only significantly higher than two other sites (Table 3.4c, Figure 3.4c). When the reduced dataset was used more significant differences were found, for example, site 96 was significantly higher than five other sites and site 106 still had the lowest allelic richness and was significantly higher than five other sites (Appendix I, Table I5c).

Within all three species there was no significant difference between sites comparing inter annual variation (i.e. 96 & 9 and 95 & 6 for *A. sulcicollis*; 97 & 12 for *I. grammatica*; and 96 & 9 for *B. rhodani*) whether using the full or reduced datasets. For complete results from mixed models preformed in R please see Appendix H, Table H3(a-c) (and for reduced loci dataset see Appendix I, Table I7(a-c)).

3.4.3.2 Stressors

Environmental stressors varied substantially over the sites studied; concentrations of aluminium ranged from 4 – 212.5 µg/l, and values of pH ranged from 4.6 – 8.34. Concentrations of cadmium (0 - 0.2 µg/l) remained relatively low compared to Buchwalter *et al.* (2008), who subjected their study invertebrates to a cadmium concentration of 4.6 nM (0.5 µg/l). Total oxidised nitrogen (0.01 – 0.76 mg/l) remained low in all streams considering a level of under 2 mg NO_3-N/l has been suggesting as appropriate to protect most sensitive freshwater species (Camargo *et al.* 2005). Aluminium and pH, on the other hand, reach

harmful levels within the sites studies. Andrén and Wiklund (2013) concluded that aluminium levels should be < 15–20 µg/l to sustain invertebrate populations, and specifically the mortality for target species *B. rhodani* was shown to increased dramatically at pH < 5.7 and aluminium levels > 20 µg/l (Andren and Wiklund, 2013).

Mean allelic richness (Appendix H, Table H2a-c) and expected heterozygosity (Table 3.1a-c) were then correlated with environmental stressors (Appendix H, Table H1). Table 3.6 shows a summary of all pairwise linear mixed models and Figure 3.5 show example plots of these correlations. The most significant correlation was for *A. sulcicollis*, when comparing both indices of genetic diversity with pH (Figure 3.5a and b). At the sites with the lowest pH, individuals tend to have lower genetic diversity. Though not significant, there was also a weak positive correlation between expected heterozygosity and pH in *I. grammatica*, however there were no correlations between pH and mean allelic richness in *I. grammatica* or between both genetic diversity measures and pH in *B. rhodani*.

There were no significant correlations between both metals (cadmium, Cd and aluminium, Al) and genetic diversity in any species, however some patterns were suggested. Within *A. sulcicollis*, expected heterozygosity and both Cd and Al had a weak negative correlation, though this was not reflected in mean allelic richness (Figure 3.5 c and d). For *I. grammatica* both genetic diversity measures had a weak negative correlation with metal concentration (Table 3.6, Figure 3.5 e and f). Though this relationship is driven by an outlier, site 93, which is the only site to have significantly lower genetic diversity (Table 3.4b) and shows evidence of demographic isolation (Table 3.2b). It also possesses the highest concentrations of Al (91.5 µg/l) and Cd (0.2 µg/l) (Appendix H, Table H1). *Baetis rhodani* did not show any correlation between metal presence and genetic diversity, though there was a weak negative correlation between cadmium concentration and allelic richness.

Considering total oxidised nitrogen (TON) was low throughout all sites studied, it was surprising to find a significant negative correlation between both measures of genetic diversity and TON for *B. rhodani* (Table 3.6, Figure 3.5 g and h). This relationship was again defined by an outlier, in this case, site 106, which has the highest concentration of TON (0.76 mg/l) and also had the lowest genetic diversity within *B. rhodani* (Figure 3.4c). Unfortunately, for this site other chemistry data were not available (Appendix H, Table H1)

so it could be that other unknown factors are driving this relationship. The likelihood of this is increased because of conflicting patterns seen in *A. sulcicollis* (showing a very weak positive correlation with TON and genetic diversity) and *I. grammatica* (showing no correlation with TON) (Table 3.6).

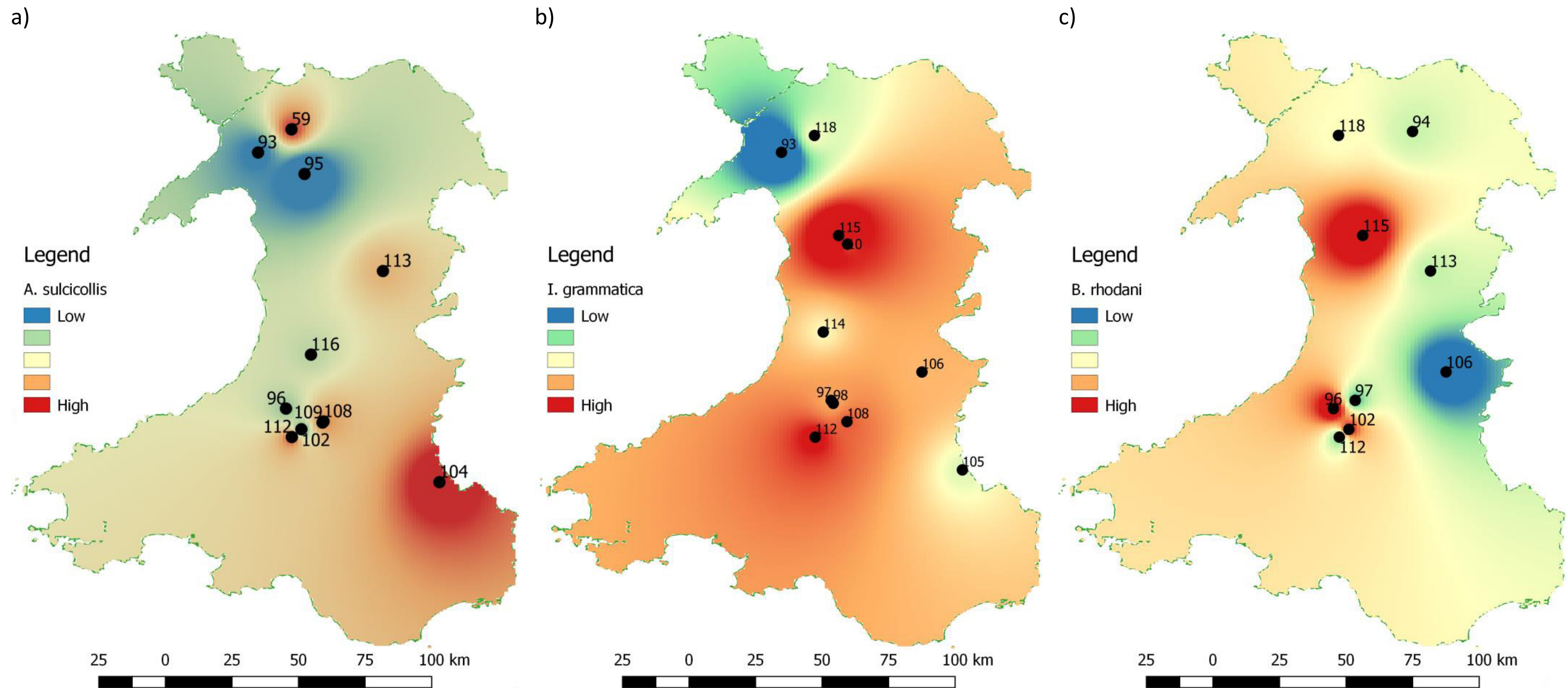


Figure 3.4. Interpolation maps creating using QGIS showing mean allelic richness across sites sampled in Wales, UK, for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. If data were available at a site for more than one year, a mean value was used.

Table 3.4. Comparing pairwise allelic richness between sites, using full datasets for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. Only significant differences in allelic richness are presented: * p-value = <0.05, ** p-value = <0.01, *** p-value = <0.001. Blue = positive coefficient value, Red = negative coefficient value. The horizontal axis is the reference site, if the regression coefficient value was positive (blue) this means that site on the vertical axis has higher mean allelic richness than the reference site above. If the value is negative (red) the site to the left of the value has a lower mean allelic richness compared to the reference site above.

Table 3.4a. *Amphinemura sulcicollis*.

	104	109	108	102	112	96	9	116	113	95	6	59	93
104				**		**	*	*		**	***		**
109											***		
108				*		*				*	***		*
102	**		*		*						*	*	
112				*		*				*	***		**
96	**		*		*							*	
9	*										**		
116	*										**		
113											***		*
95	**		*		*							**	
6	***	***	***	*	***		**	**	***			***	
59				*		*				**	***		**
93	**		*		**				*			**	

Table 3.4b. *Isoperla grammatica*.

	105	106	108	12	97	98	112	114	115	10	118	93
105												
106												
108												*
12												
97												*
98												*
112												*
114												
115												**
10												*
118												
93			*		*	*	*		**	*		

Table 3.4c. *Baetis rhodani*.

	112	102	97	96	9	106	113	115	94	118
112										
102						*				
97				*						
96			*			*				
9						*				
106		*		*	*			*		
113										
115						*				
94										
118										

Table 3.5a-b. Comparing the allelic richness between each cluster, using full datasets for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. Only significant differences in allelic richness are presented: * p-value <0.05, ** p-value <0.01, *** p-value <0.001. Blue = positive coefficient value, Red = negative coefficient value. The horizontal axis is the reference cluster, if the regression co-efficient value was positive (represented here with blue asterisks) this means that site on the vertical axis has higher mean allelic richness than the reference cluster above. If the value is negative (represented here with red asterisks) the cluster to the left of the value has a lower mean allelic richness compared to the reference cluster above.

Table 3.5a. *Amphinemura sulcicollis*.

	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1			***	***
Cluster 2			**	**
Cluster 3	***	**		
Cluster 4	***	**		

Table 3.5b. *Isoperla grammatica*.

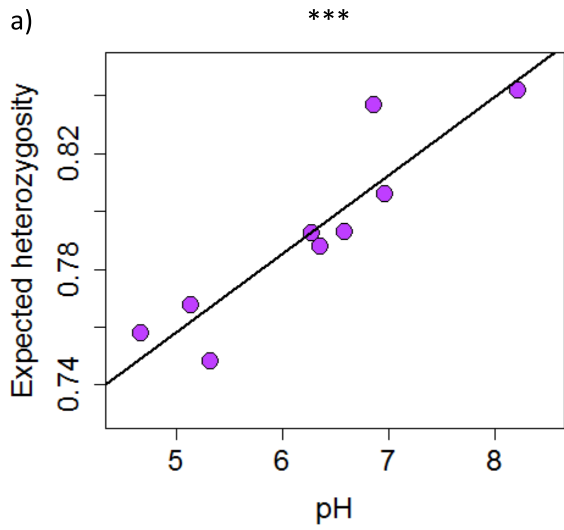
	Cluster 1	Cluster 2
Cluster 1		**
Cluster 2	**	

Table 3.6. Summary of all linear mixed models preformed between the two indices of genetic diversity; (always on the y-axis) H_e = expected heterozygosity; AR = mean allelic richness) and four environmental stressors (on the x-axis); aluminium (Al, µg/l), cadmium (Cd, µg/l), pH (pH-units) and total oxidised nitrogen (TON, mg/l). Showing adjusted R-squared (R^2), p value (p) and number of sites (n) for each regression. Blue = a positive correlation (i.e. as the environmental stressor value increases the genetic diversity indices value increases); Green = a negative correlation (i.e. as the environmental stressor value increases the genetic diversity indices value decreases); Black = no correlation. Significant values in bold: * = p-value <0.05, ** = p-value <0.01, *** = p-value <0.001.

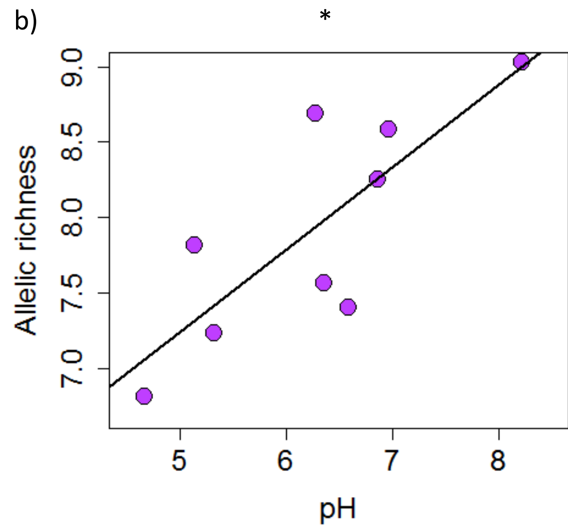
		A. sulcicollis		*I. grammatica*		*B. rhodani*	
		H_e	AR	H_e	AR	H_e	AR
pH	R^2	0.81	0.58	0.26	0.00	-0.14	-0.23
	p	**0.001*** **	**0.010***	0.094	0.351	0.568	0.831
	n	9	9	9	9	6	6
Cd	R^2	0.41	-0.20	0.42	0.45	-0.02	0.36
	p	0.074	0.989	0.096	0.086	0.394	0.124
	n	7	7	6	6	6	6
Al	R^2	0.29	-0.10	0.40	0.17	-0.25	-0.22
	p	0.125	0.532	0.076	0.198	0.978	0.772
	n	7	7	7	7	6	6
TON	R^2	0.15	0.23	-0.08	-0.12	0.59	0.49
	p	0.143	0.090	0.562	0.885	**0.026***	**0.049***
	n	10	10	10	10	7	7

Amphinemura sulcicollis

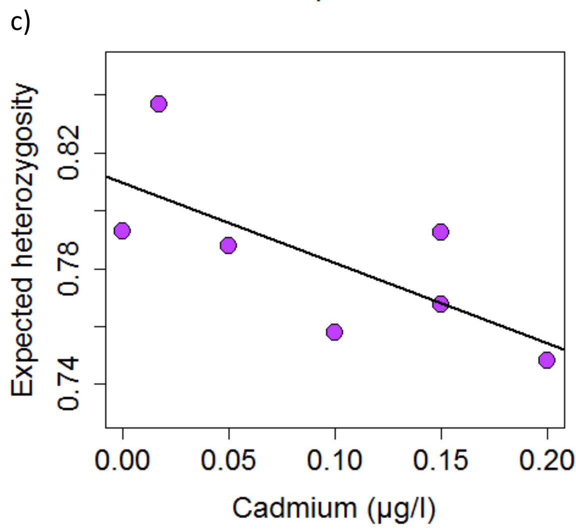
a) ***



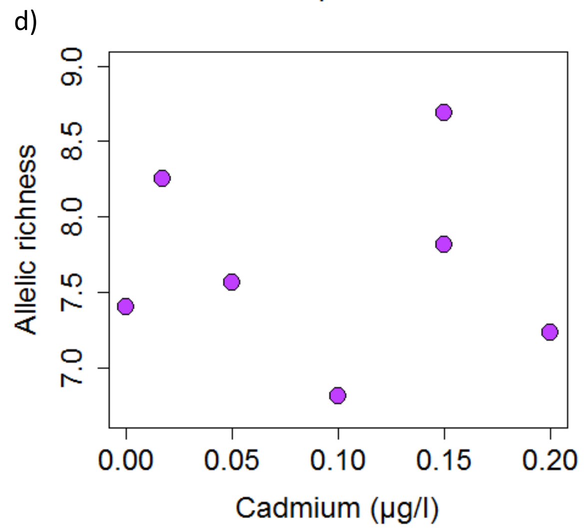
b) *



c)

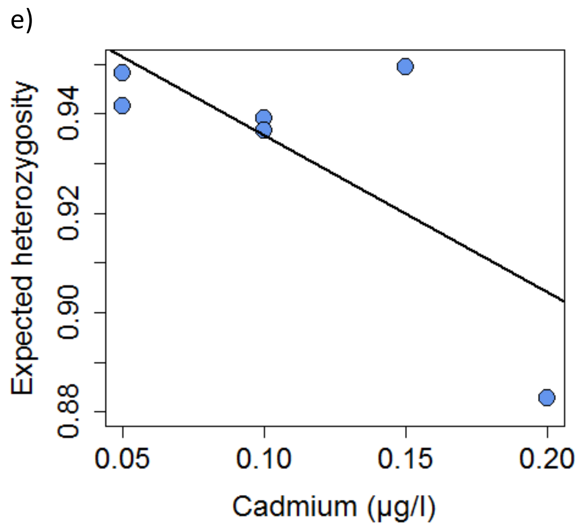


d)

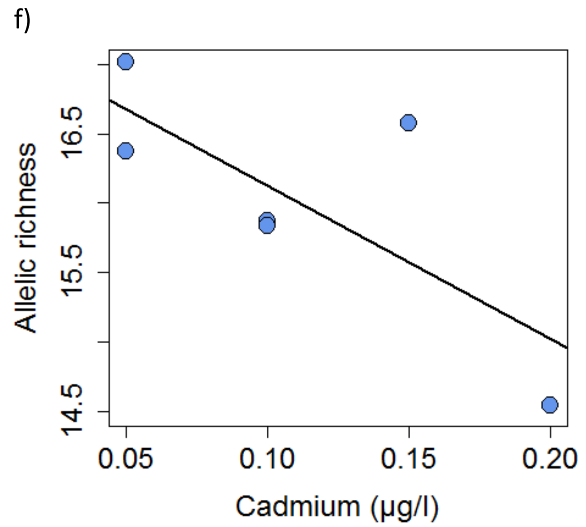


Isoperla grammatica

e)



f)



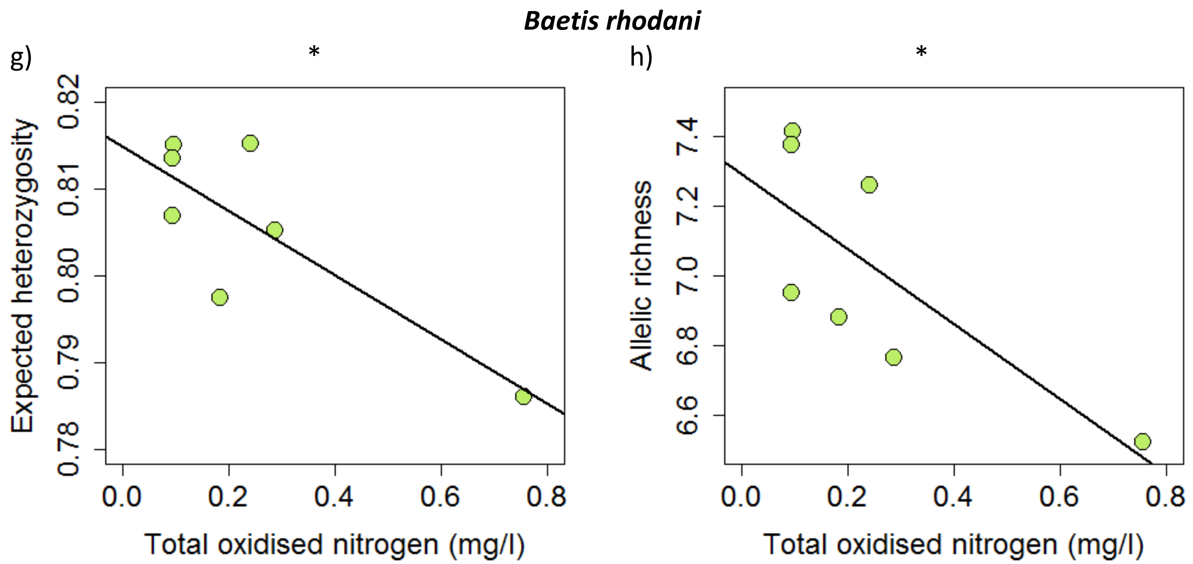


Figure 3.5. Correlations between genetic diversity indices (expected heterozygosity and allelic richness) and environmental stressors for: a) - d) *Amphinemura sulcicollis* showing correlations with pH and cadmium*;* e) - f) *Isoperla grammatica* showing correlations with cadmium and g) - h) *Baetis rhodani* showing correlations with total oxidised nitrogen. Where correlation exists a line of best fit has been added, created from a linear mixed model in RStudio. Stars above graphs denote significances: * = p-value <0.05, ** = p-value <0.01, *** = p-value <0.001.

3.5. Discussion

3.5.1 HWE and descriptive statistics

Amphinemura sulcicollis, *I.* grammatica and *B. rhodani* represent highly polymorphic macro-invertebrate species, particularly *I. grammatica*, which has a high allelic richness even compared to other Plecoptera species (Theissinger *et al.* 2009; Geismar and Nowak 2013; Theissinger *et al.* 2013; Elbrecht *et al.* 2014). A striking pattern for all three species in this study was the non-conformity to Hardy-Weinberg Equilibrium (HWE) due to heterozygosity deficiency, also shown by the high proportion of significant F_{IS} values. As all three species represent large outbreeding populations, characterised by high polymorphism, the likelihood of F_{IS} value reflecting inbreeding within populations is therefore low, especially as few related individuals were found within the data. The significance of F_{IS} values among populations, therefore, is likely to reflect deviations from HWE genotype frequencies and relative population heterozygosity (Addison and Hart 2005). Possible reasons for this departure can be divided into two categories, natural and artificial. Examples of artificial departure from HWE include: poor primer design and optimisation, null alleles and genotyping errors (such as stuttering or large allele dropout; discussed further in section 2.6.5; Addison and Hart 2005; Brownlow *et al.* 2008). Whereas natural reasons include mutation, natural selection acting on the genetic markers, inbreeding effects (i.e. mating among relatives), non-random mating and population admixture (unrecognized spatial or temporal structure within samples known as the Wahlund effect) (Addison and Hart 2005; Brownlow *et al*. 2008).

Difficulties arise, however, when separating poor primer design from natural causes of HW disequilibrium. This was addressed first by having a rigorous genotyping experimental design with high level of repeats (up to 76%) so that accuracy of microsatellite amplification could be assured (section 2.5.2); and second by identifying the loci that were consistently out of HWE (in more than half of the populations tested (Table 2.10 (a-c) and Appendix G), and removing them from the analysis to establish their effect on the results (Appendix I). Though it is difficult to completely rule out these reasons for departure from HWE, it is unlikely in this case because replication proved that primer design and PCR conditions were stable and reliable, and did not identify any null homozygotes. Previous analysis found no consistent evidence of allelic dropout or scoring errors in any of the loci (section 2.6.3) and

removing loci that were out of HWE in 50% of sites or higher (which also removed loci that showed evidence of null alleles in greater than 77%, 33.3% and 50% in *A. sulcicollis*, *I. grammatica* and *B. rhodani* respectively, Table 2.11 (a-c)) had little effect on the general patterns seen in the data, suggesting that these patterns are not being led by problem loci but are instead inherent in the data. There is also the fact that HW disequilibrium was prevalent in such a large percentage of the loci tested (Table 2.7), if the problem was artificial it is expected to be locus-specific, and therefore feature for a smaller number of loci (Dewoody *et al.* 2006; Aguilar and Jones 2009; Dharmarajan *et al.* 2013).

HW disequilibrium due to heterozygosity deficiency is a common theme in widespread invertebrates both in marine and freshwater systems (Addison and Hart 2005; Brownlow et al. 2008; Schultheis et al. 2008; da Silva-Méndez *et al.* 2013; Elbrecht et al. 2014; Postaire *et al.* 2015), therefore it seems likely that this issue is a natural phenomenon. Addison and Hart (2005) performed a literature review of 124 marine invertebrates, and although they find that spawning invertebrates have the highest F_{IS} values, positive F_{IS} values were found in most studies. Though the species studied here are not spawning, they are dominant, widespread, outbreeding populations and therefore might share similar traits, for example, large variance in reproductive success (Purser 1966). Most studies suggest mating among related individuals and Wahlund effect as the cause of HW disequilibrium in invertebrates (Brownlow *et al.* 2008). This study has attempted to test these factors by investigating relatedness and analysing the data based on the structure found (per cluster). Therefore these factors seem unlikely, and as in Brownlow *et al.* (2008) the precise nature of the overall heterozygote deficiency still remains unclear. The high number of alleles and high variance within individuals present in these species and many other invertebrate species may mean that statistical and biological significances may not coincide, as suggested by Hedrick (1999).

3.5.2 Population genetic structuring

The genetic structure of the three described freshwater invertebrates was investigated over an area of upland Wales, 145 x 90km, encompassing 10, 8 and 8 (*A. sulcicollis*, *I.* grammatica and *B. rhodani*) catchments, with the aim of investigating dispersal and inferring gene-flow. *A. sulcicollis* showed the highest genetic structure, suggesting more limited dispersal

compared to *I. grammatica* and *B. rhodani*. It was hypothesised that *A. sulcicollis* might have limited dispersal ability compared to the other species due to its comparatively small body size. As suggested by previous studies, *B. rhodani* has the least structure, suggesting high gene flow between sites and a high level of effective dispersal (Williams 2003; Rebora *et al*. 2005; Alp *et al.* 2012).

As with *Plectrocnemia conspersa* in Wilcock *et al*. (2007), all three study species (with the exception of *A. sulcicollis* at site 104) showed limited genetic structure, hence high inferred levels of gene flow, and this was observed among all southern sites. All three species rely on their terrestrial adult life stage to disperse over land: sites over 100km away from each other, in separate catchments, display limited structure. This was supported by a lack of significant allele frequency differences, comparing sites within the same catchment (the Tywi) and those in adjacent catchments. Further, only *A. sulcicollis* showed a genetic signature of isolation by distance. Analysis at the same site (sampling in 2012 and 2013) showed no temporal differentiation in any species, suggesting that allele frequencies do not change from one generation to the next, allowing us to conclude that genetic data were comparable when sampling a year apart.

The greatest evidence for genetic isolation was found in the north of Wales, although a simple north-south divide was not evident, as there were northern sites that clustered together with southern sites (for *A. sulcicollis* sites 59 and 55; *I. grammatica* and *B. rhodani* site 118, see Figure 3.4 and Appendix H, Figure H1). It was found that particular sites, as opposed to whole regions, showed isolation from the rest of the meta-population suggesting these sites possess unique characteristics impeding gene-flow. For example, at site 93, both *A. sulcicollis* and *I. grammatica* showed significant genetic differentiation. *B. rhodani* was not found at this site in 2013, neither had it been found there on previous sampling expeditions in 1984, 1995 or 2012 (unpublished DURESS data). As *B. rhodani* has a weak genetic structure suggesting high dispersal ability, its non-presence at this site is more likely to be due to the species not surviving there, rather than an inability to disperse. Increased elevation has been known to cause restricted gene flow (Funk *et al.* 2016), however, although the isolated sites are within Snowdonia National Park, in this case the isolated sites are no higher in elevation than sites showing panmixa (Table 2.1(a-c), Appendix H, Figure H2). Therefore further examination into the chemistry of the sites was explored below.

3.5.3 Genetic diversity and correlations with stressors

Differences in genetic diversity correlated with patterns of genetic differentiation, meaning that the most isolated sites were more likely to have significantly different genetic diversity. At sites 93 (for *A. sulcicollis* and *I. grammatica*) and temporal sites 95 and 6 for (*A. sulcicollis*) there was significantly lower genetic diversity than the surrounding sites; however southern site 104 (for *A. sulcicollis*) showed significantly higher genetic diversity. It is common that small isolated populations have lower genetic diversity due to genetic drift (Ellingson and Krug 2016). A reduction in genetic variation may decrease fitness; posing an immediate threat from demographic events or in reducing the long-term capacity to respond to environmental change (Westemeier *et al*. 1998).

The isolated northern sites with reduced genetic diversity (site 93 and 95 in *A. sulcicollis* and *I. grammatica*, Figure 3.4) did not stand out in terms of distance, elevation or surrounding land-use. They do, however, stand out in terms of their chemistry. They have the lowest pH of all sites sampled (4.7 – 5.3 pH) and amongst the highest levels of cadmium (0.1 – 0.2 μg/l) and aluminium (91.5 – 143.3 μg/l). This was consistent temporally with sites that were investigated in two consecutive years showing similar results (Appendix H, Table H1). Genetic diversity was then correlated with the environmental stressors recorded (pH, Cadium and Aluminium concentrations and TON) to see if species tolerance to stressors had any effect on genetic diversity. It was expected that *A. sulcicollis* was a more tolerant compared to baetidea spp., however a genetic response to stressors may still have been shown. A highly significant relationship between increased acidity and reduced genetic diversity was found for this species, and although no other significant correlations were found, this could be due to the fact that low pH and high metal concentrations are only seen in a small percentage of sites. There was, however, a negative correlation between genetic diversity and metal concentrations in *I. grammatica* driven by site 93. The same patterns of population structure or variation in genetic diversity were not seen in *B. rhodani*, because the well-known acid-sensitive species was absent from these sites. The results adhere to the hypothesis that differences in sensitivity between these three species would influence the genetic structure and diversity found; moreover the results revealed that more tolerant Plecoptora species still show a genetic response to acidity.

Loss of genetic diversity at the above sites appears to be caused by a combination of isolation and environmental stressors. It is unlikely simply to be caused by geographical isolation limiting dispersal in the North of Wales, since if this was the case in *A. sulcicollis*, northern site 59 would also show significantly lower genetic diversity (Figure 3.4a). On the other hand, the loss of genetic diversity cannot only be caused by environmental stressors, such as pH, because southern site 116 has a lower pH than site 93 (and high levels of both metals), and though it has slightly reduced genetic diversity, it is only significantly lower than site 104 (which had the highest genetic diversity recorded; Figure 3.4a). This suggests that a combination of geographic isolation and acidity is the likely cause of the loss of genetic diversity at these specific northern sites. It is firstly worth noting however that interactions between the stressors recorded and other stressors not recorded could potentially have influenced the genetic diversity and structure of these species. Interaction effects could not be statistically analysed in this study due to the sample size, however, Jackson *et al.* (2016) found significant interaction effects. In Jackson *et al.* (2016), the cumulative mean effect size of pairs of stressors was less than the sum of their single effects. It is difficult to predict the exact effect of other stressors due to a limited number of studies that investigate the effect of environmental stressors on genetic diversity (Bickham *et al*. 2000), particularly within the freshwater environment. However, stressors typically affecting the freshwater environment such as global change, physical habitat alteration, and invasive species are all likely to harm ecosystem functioning and decrease biodiversity (Navarro-Ortega *et al*. 2015). Dissolved oxygen is another important factor in water quality for aquatic fauna which is well studied in ecology (Perna and Burrows, 2005) however Crispo and Chapman (2008) investigated the effect of dissolved oxygen on the genetic structure of an Afian cichlid fish, Pseudocrenilabrus multicolor victoriae, and found no effect.

The effects of environment stressors present at these sites supports the theory that it is not just distance and dispersal ability that leads to genetic differentiation and differences in genetic diversity in aquatic macroinvertebrates. Previous studies have also shown that dispersal ability is not a limiting factor in terms of stream recovery (Williams 2003; Masters *et al.* 2007). Environmental stressors, however, have been known to cause local adaption in related species. For example, Trichoptera species distribution and composition in the lakes of the Pyrenees have been found to be predominantly governed by local environmental factors, rather than by dispersal constraints (de Mendoza *et al.* 2015). One caddisfly species

provided strong evidence of local adaptation to annual maximum water temperature (Nukazawa *et al.* 2015) and Alp *et al.* (2012) explained genetic differentiation of the species *Gammarus fossarum* to be the cause of habitat specialisation and potentially local adaptation. Macher *et al.* (2016) described a mayfly species response to multiple stressors, and expressed the need for species level data of response to stressors. For example, toxicants are an important yet understudied driver of biodiversity (De Laender *et al.* 2014). Therefore it is possible that isolated sites show reduced genetic diversity because of adaption to environmental conditions such as acidity. When examining long term natural acidification and stonefly ecological traits, Petrin (2011) found smaller body size, greater reproductive output and faster life cycles, suggesting that these stoneflies had adapted to this acidic environment. Perhaps *A. sulcicollis*, as the smallest species investigated, is the most suited to these acidic sites as shown in Petrin (2011). Other studies have found that *I. grammatica* is present at sites with pH <6 where other similar species (*I. difformis*) are absent suggesting acid tolerance in this species (Malmqvist and Sjöström 1989). *Isoperla grammatica* and related species, *Amphinemura stanfussi* abundances were compared between rivers of varying pH in Lock and Goethals (2008)'s study, and they found that *I. grammatica* was one of the most acid tolerant species, and although both species persisted in low pH streams, *I. grammatica* was recorded at lower pH than *A. standfussi*. Eriksen and Pettersen (2016) recorded invertebrate response to aluminium sulphate. They found that Baetis spp (*Baetis alpinus* and *Baetis rhodani*) are the most sensitive but *I. grammatica* was also effected but not as severely. It is suggested that the decline in population sizes could be a behavioural response related to lower prey densities (Baetis, Simuliidae, and Chironomidae) rather than sensitivity. Petrin *et al.* (2007) also found that Plecoptera richness, unlike Ephemeroptera richness, did not correlate with pH. These studies highlight that *I. grammatica* and *A. sulcicollis* are mostly acid tolerant, so perhaps here we show that these species do have a response that has not previously been seen because of the lack of genetic investigation.

Acidification inhibits microbial decomposition of leaf-litter, which in turn reduces food quality and availability for shredder assemblages in the stream (Larrañaga *et al.* 2010; Pye *et al.* 2012). However, several shredder species have been known to consume algae under acid conditions where specialist grazers cannot persist (Ledger and Hildrew 2005; Feeley and

Kelly-Quinn 2015). *Amphinemura sulcicollis*, as an opportunistic feeder (known as a shedder, grazer and collector-gather, see section 1.2.1) might therefore need to locally adapt to these conditions and may thus become demographically isolated at these sites. The "multifarious selection" hypothesis of Rice and Hostert (1993) presents a situation in which speciation is promoted by a multitude of different selection pressures acting on many genes/traits, this could be happening at these unique sites. Buchwalter and Luoma (2005) found that finer morphological features affected species sensitivity to dissolved metal, specifically, the relative numbers of ionoregulatory cells (chloride cells), but also different functional traits could lead to differential exposure to cadmium (or other stressors) in the same habitat, for example seasonal aspects of development or food preference, which will affect bioaccumulation (Buchwalter *et al.* 2008). As *I. grammatica* is a predator it could be more sensitive to bioaccumulation of harmful metals (Jardine *et al.* 2012), which may be why it is absent from site 95 and why site 93 is genetically distinct.

Studies into the effect of acidification usually examine species composition to establish the effect of reduced pH, and many conclude that stoneflies are less susceptible, focusing instead on mayfly species such as *B. rhodani* (Kowalik and Ormerod 2006). This study, however, shows that while individuals are able to persist in these conditions there is still a genetic response to acidification in *A. sulcicollis*. Examining the genetic diversity of these species as well as species compositions may give extra information about the health of the system. For example, many studies have tried to explain why biological recovery does not always follow chemical recovery of streams (Yan *et al.* 2003; Ledger and Hildrew 2005; Kowalik and Ormerod 2006; Kowalik *et al*. 2007). Perhaps a factor limiting recovery could be reduced genetic diversity in some species following acidification which can make populations more vulnerable to other disturbances.

As widely used indicators of ecosystem health (Buchwalter and Luoma 2005), here we show the value of adding genetic methods into the study of aquatic insect ecology, providing further insight into ecosystem health than gained by examining purely the presence or absence of species in certain environments.

Chapter 4 - The Relationship between Species and Genetic Diversity in Freshwater Invertebrates

Contributions

- All species abundance data (collection and identification) and environmental stressor data (as used in Chapter3) were created by DURESS (Diversity of Upland Rivers for Ecosystem Service Sustainability: http://nerc-duress.org). Funded by the Natural Environment Research Council (NERC).
- Species diversity (taxonomic and functional diversity indices) were calculated in collaboration with Tano Gutiérrez-Cánovas (Cardiff University, Post-doctoral Research Associate) using the same R script and data (Appendix K).

Highlights

- A significant correlation was found between species diversity and the genetic diversity of *Amphinemura sulcicollis*, which could be explained by that fact that both genetic diversity for this species and species diversity had a significant positive correlation with pH.
- This observation supports the hypothesis where the characteristics of the locality, in this case acidity, is hypothesised to affect both levels of biodiversity (Case one, see Section 1.5.1.1.1).
- It is suggested that where there is genetic isolation, and a driver which affects both levels of diversity, a positive species-genetic diversity correlation can be driven but that it is not universal.

4.0 Abstract

A positive species-genetic diversity correlation (SGDC) has been observed in numerous taxa although it is not universally supported. A common explanation for a SGDC is that environmental features drive a parallel response in these two fundamental aspects of biodiversity. Additionally, positive SGDC has been most often observed when comparing discrete sampling units such as lakes or islands, suggesting that genetic isolation may have some influence.

SGDC was investigated in three freshwater macroinvertebrates, *Amphinemura sulcicollis*, *Isoperla grammatica* and *Baetis rhodani,* providing the first investigation of this phenomenon in three species using microsatellite markers. The aim was to investigate factors that may lead to a SGDC and it was hypothesised that isolation and a common driver may be needed. This study will help fill knowledge gaps within the SGDC literature on non-plant species, the freshwater environment and species with high dispersal potential. These species differ in their genetic structure and sensitivity to certain environmental stressors. Species diversity (species richness, Shannon and functional diversity) from entire macroinvertebrate assemblages were compared with genetic diversity (mean allelic richness and expected heterozygosity) to investigate whether a correlation existed between these two classes of biological parameter. Both species and genetic diversity were also compared with environmental stressors (pH, aluminium, cadmium and total oxidised nitrogen) to assess potential underlying mechanisms for any SGDC found.

Due to species-specific differences in isolation and acid sensitivity, SGDC patterns differed among species. Genetic diversity in *A. sulcicollis*'s correlated positively with species diversity; the underlying driver of this correlation was thought to be acidity, because both species diversity and genetic diversity for *A. sulcicollis* had a significant positive correlation with pH across locations. However, as hypothesised pH was not the only cause as one low pH site had reduced species diversity but not significantly reduced genetic diversity, possibly due to higher gene-flow in this area. It is suggested that isolation interacts with environmental influences to create SGDCs. SGDC in *Isoperla grammatica* was only moderately positive, due to one outlying site. The acid sensitive *Baetis rhodani* was absent

from acid sites, and there was no significant SGDC. Finding species-specific differences highlights that positive SGDC, though often found, can never be assumed.

Keywords: Species diversity, genetic diversity, SGDC, pH, freshwater, macroinvertebrates

4.1 Introduction

To protect global biodiversity a strong foundation of knowledge is essential (Brooks *et al.* 2006; Baselga *et al.* 2013; Hoban *et al.* 2013). Biodiversity includes ecosystem, species and genetic diversity (Vellend *et al.* 2014). As introduced in Chapter 1 (Section 1.5) investigating possible links between species and genetic diversity (the species-genetic diversity correlation, SGDC) has been increasing in popularity over recent years. Identifying spatial patterns of biodiversity has been a dominant area of research in ecology and evolution, this has been mainly in order to identify biodiversity hotspots and to study evolutionary adaption (Vellend 2014). The idea of a correlation between levels of biodiversity is appealing because it means that a conservation area designed to protect species diversity (Rodrigues *et al.* 2004; Naveda-Rodríguez *et al.* 2016) might also be conserving genetic diversity, simplifying biodiversity assessments and management schemes. In previous studies a positive SGDC has most often been found (Table 1.3), however, negative relationships (Xu *et al.* 2016) and no correlation (Avolio and Smith 2013) have also been observed. Those that find a correlation also aim to understand the underlining mechanisms and positive SGDCs are usually explained by heterogeneous environments driving parallel patterns (Case one, see Section 1.5.1.1.1) (e.g. Frey *et al.* 2016; Csergo *et al.* 2014).

As well as characteristics of the locality[1], which would be different in each environment, isolation is another common factor of SGDCs. Studies that take place across discrete sampling units (whether islands (e.g. Vellend 2003); lakes (e.g. Baselga *et al.* 2013) or forest fragments (e.g. Struebig *et al.* 2011)), are more likely to show a significant positive SGDC than those that study continuous habitat areas (e.g. Taberlet *et al.* 2012; Vellend 2010).

This is the first study to investigate SGDC with three different species, *Amphinemura sulcicollis*, *Isoperla grammatica* and *Baetis rhodani*, using hypervariable microsatellite

[1] a locality being defined as an area where the community may be found

markers. Many SGDC studies use markers that are less polymorphic than microsatellites (Table 1.3), and could potentially miss fine scale genetic differentiation. In Chapter 3, microsatellite markers (10, 10 and 13 loci for *A. sulcicollis*, *I. grammatica* and *B. rhodani,* respectively) were used to investigate genetic structure and diversity. In this Chapter, we compare this genetic diversity to species diversity data at the same sites. As others (Silvertown *et al.* 2009; Blum *et al.* 2012) have also noted, although experiments that manipulate environmental conditions, area and isolation are possible in theory, the time frame to complete this would be unfeasible for most communities. Observational field studies which monitor environmental variation and gradients can, however, help decipher what drives SGDCs and under what circumstances correlations are, and are not, observed (Vellend and Geber 2005; Lankau and Strauss 2007; Silvertown et al. 2009; Struebig *et al.* 2011).

The theory of SGDC is built upon a mixture of traditional models which provide explanations on how selection, drift, speciation and dispersal might interact (Vellend 2010), and which are used to explain why these two levels of biodiversity might co-vary. From MacArthur and Wilson (1967)'s Neutral Island Biogeography Model which incorporates just drift and dispersal to more recent extensions of the concept (Lomolino 2000; Whittaker 2000), models and empirical studies have highlighted the importance of dispersal, gene flow and isolation (Csergo *et al.* 2014). Dispersal can have a range of effects on possible SGDCs; for example, following a mainland-island model, in the absence of selection or speciation, increased dispersal will increase both species richness and genetic diversity, by potentially adding new species and genotypes and counteracting genetic drift. However, if selection is playing a role, dispersal from different localities, each possibly representing a different selective environment, means that different outcomes for SGDC are potentially vast (Vellend 2010).

In Chapter 1 certain knowledge gaps were identified, namely that most published studies directly comparing species and genetic diversity use just one species to assess the genetic diversity, and most often this is a plant or tree species. The freshwater environment has few

studies (six, Table 1.3), and only one previous study using invertebrates with a terrestrial life stage (the predaceous diving beetles; Baselga *et al.* 2012), allowing long-range active dispersal between catchments. Blum *et al.* (2012) called for more studies reflecting different life history traits to improve the understanding of what causes a SGDC. It has been shown that invertebrates without a terrestrial life history stage have greater genetic differentiation and isolation because of restricted gene-flow (e.g. Alp *et al.* 2012). Investigating species such as plants with no active dispersal could explain why so many positive correlations have been found within the SGDC literature, because as Vellend (2014) suggests, SGDCs should be strongest when size and isolation of sampling units exert a strong influence on diversity patterns.

One of the difficulties of studying species and genetic diversity simultaneously is that species diversity is usually studied comparing different geographically distinct sites, for example different streams, where the population of each species is defined as those living in that location (e.g. De Castro-Català *et al.* 2015; Feld *et al.* 2016). Genetic studies on the other hand, which usually focus on one particular species, define the population based on their genetic structure, often studying large geographical areas as one unit if high gene flow is found (e.g. Paetkau *et al.* 1999; Palsbøll *et al.* 2007). Therefore, depending on the genetic structure the two different levels of biodiversity may be occurring at different scales. It is therefore hypothesised that it is only when there is genetic isolation between sites that both species and genetic diversity will be subject to the same features of the locality, possibly leading to a SGDC. However, with high gene-flow between sites, species diversity may be more affected by environmental heterogeneity than genetic diversity. A large proportion of the SGDC literature do not describe the genetic structure of species they are studying, however this study, as it uses three species with greater dispersal potential and differing patterns of isolation, will help to investigate this theme.

In Chapter 3 it was found that genetic diversity mirrored genetic isolation in the species investigated, and that in isolated sites, environmental stressors such as pH drove a decline in genetic diversity. The primary aim of this current chapter is to test the hypothesis that the same drivers could create a parallel response in species diversity. Specifically, we tested

whether isolation in conjunction with environmental stressors drove parallel patterns of genetic diversity for *A. sulcicollis*, *I. grammatica* and *B. rhodani*. We hypothesise that correlations will be more likely where there is genetic isolation and a driver, in this case acidity. *A. sulcicollis*, therefore, is possibly more likely to present a positive SGDC, as this species showed the most genetic structure in chapter 3, whereas within *B.rhodani,* it is hypothesied that it is unlikely to have a SGDC because no genetic isolation was found within the samples studied.

4.2 Methods

4.2.1 Invertebrate sample collection

Genotyping data were collected and genetic diversity was analysed as described in Chapter 3. Species data were made available by the DURESS project. Unfortunately no species data were available for sites 104, 105, 106 and 113 (Grwyne Fawr, Honddu at Capel, Ithon at Llandewi, and Nant Gelli Gethin).

At all other sites that were analysed for genetic diversity, invertebrates were collected by standardised kick-sampling methods to assess species diversity (see Appendix K). Invertebrates were collected in riffles using 2 minute semi-quantitative kick samples, during spring of 2012 and 2013 (Appendix K) from all sites, using a 1 mm mesh hand net. This strategy sampled most major habitats and is a well-calibrated method sufficient to detect differences between sites (Weatherley and Ormerod 1987; Bradley and Ormerod 2002a).

Samples were preserved on-site using 100 % industrial methylated spirit (IMS). In the laboratory, samples were hand-sorted and preserved in 70 % IMS. Major groups were identified and counted to species or genus for most taxa, or to family in cases where taxonomy was difficult or larvae were insufficiently well developed (e.g. Diptera species). Trait-level information was generally at genus or family level. Some taxa were identified at a very coarse taxonomic level (e.g. Oligochaeta and Tricladida).

4.2.2 Species diversity indices

Three indices of species diversity were used:

1) Species richness. As the simplest way of describing community diversity, it is frequently used in biomonitoring studies (Smith and van Belle 1984; Compin and Céréghino 2003).

2) Shannon (1948) diversity index. This is also a commonly used figure to reflect species diversity (Spellerberg and Fedor 2003). Shannon diversity takes into account the relative abundance of the species present, reducing the influence of rare species.

3) Functional diversity. This accounts for both the variability and distribution of biological traits across species (Mason *et al.* 2005). Functional measures are positively related with ecosystem function and stability, showing non-random responses to increased environmental stress (Hooper *et al.* 2005; Gutierrez-Canovas *et al.* 2015). Functional diversity has not been used previously in SGDC literature so here it was used as well as the traditional taxonomic diversity indices to explore how it varied with genetic diversity, also Bady *et al.* (2005) suggest that functional diversity is more precise and informative than other measures of species diversity.

To describe functional diversity patterns, functional dispersion was calculated as explained in Laliberté and Legendre (2010), based on a Gower dissimilarity matrix derived from seven biological traits (Tachet *et al.* 2002). These traits reflect morphology, life history, aquatic stage, dispersal, resistance forms, respiration, and locomotion (Appendix K, Table K3). The information available in Tachet *et al*.'s database contains the affinity of each genus to each trait category (i.e. fuzzy coding approach; Chevene *et al.* (1994)). This procedure is frequently used to account for within-genus variation (e.g. Gutierrez-Canovas et al. 2015; Manfrin *et al.* 2016; Pelosi *et al.* 2016).

Species diversity indices were estimated from the data available in Appendix K (Table K1 = Species data, Table K2 = Genus data and Table K3 = Trait data) using RStudio (R script available in Appendix J, Section 3).

4.3 Statistical analysis

4.3.1 Species genetic diversity correlation

Species diversity results are shown in Appendix K (Table K4), and the genetic diversity indices, expected heterozygosity (uH_e) and mean allelic richness (rarefied for differences in sample sizes) are given in Chapter 3 (Table 3.1a-c) and Appendix H (Table H2a-c), respectively.

First, in order to visualise species diversity over the geographical area, interpolation maps were created using QGIS v. 2.14.2 (QGIS Development Team 2016) for each species diversity index. If data were available for more than one year, a mean value was used, yielding one value per site.

Each species diversity index was correlated with each genetic diversity index, per species, using a linear regression in RStudio (as was used for comparing environmental stressors to genetic diversity indices in Chapter 3 (Section 3.3.3.2; R script in Appendix J, Section 2).

If species and genetic data were available for 2012 and 2013, both values were averaged so there was only one datum point per site (this only occurred once within *I. grammatica* where species and genetic data were available for sites 12 and 97, which represents temporal sites at the same location). Where possible, species and genetic data were compared from the same sampling year, however it was sometimes necessary, for example, to compare species diversity data from 2012 to genetic diversity data from 2013. Within Chapter 3, it was found that genetic diversity was not significantly different between these two different sampling years (Table 3.4a: site 9 *vs* 96 and site 6 *vs* 95 for *A. sulcicollis*; Table 3.4b: site 12 *vs* 97 for *I. grammatica* and Table 3.4c: 9 *vs* 96 for *B. rhodani*). Long-term changes in species diversity have been studied over 25 years within the study area (Tywi catchment) and found to be gradual by Ormerod and Durance (2009), using the information from this study it could be inferred that the difference in species diversity between just two years would not influence the result.

4.3.2 Species diversity *vs* environmental stressor

The same environmental stressors, aluminium (Al, µg/l), cadmium (Cd, µg/l), pH (pH-units) and total oxidised nitrogen (TON, mg/l) (Appendix H, Table H1) that were compared with genetic diversity data in Chapter 3 are compared with species diversity data in the present Chapter. This was to assess whether there were any mutual drivers of species and genetic diversity. Correlations were assessed between each stressor and species diversity index using a linear mixed model in RStudio (as in Section 3.3.3.2; R script in Appendix J, Section 2). If stressor and species diversity data were available for 2012 and 2013, both values were averaged so there was only one datum point per site (occurring for site codes 12 and 97; and site codes 39 and 109), all other comparisons of species data and stressor data were available from the same sampling year.

4.4 Results

4.4.1 Species genetic diversity correlation

The range of species diversity within the sites studied varied for each index; with species richness having the widest range, 11 – 41, with northern site 6 having the lowest value and southern site 97 having the highest. Shannon diversity indices ranged from 1.53 - 2.86 (lowest was again in the north, site 55, and the highest again within the southern Tywi catchment, site 39), whereas functional diversity did not vary a great deal across the sites, from 0.10 – 0.17 (site 6 - site 4, both in northern Wales). Figure 4.1a-c shows that each species diversity index shows generally the same pattern, with the highest species diversity being found in the south of Wales, usually the Llyn Brianne area (Tywi catchment), and the lowest in north of Wales, especially around sites 93, 118/55, 95/6. When the data in Figure 4.1a-c are compared with allelic richness (see Chapter 3, Figure 3.4 a-c), a very similar pattern is seen (also see Appendix K, Figure K1).

Two measures of genetic diversity, mean allelic richness (Appendix H, Table H2a-c) and expected heterozygosity (Table 3.1a-c) were then compared to the three indices of species diversity: species richness, Shannon and functional diversities (Appendix K, Table K4). Table

4.1 represents the summary of all pairwise mixed linear models and Figure 4.2 shows plots of these correlations.

Amphinemura sulcicollis was the only species where a significant positive SGDC was found. This relationship exists between both genetic diversity estimates of *A. sulcicollis* and species richness and Shannon diversity (Figure 4.2, a-b, d-e; Figure 4.4). There was, however, no significant correlation between either genetic diversity estimates or functional diversity for this species (Figure 4.2, c and f).

Though there was a slight positive correlation within *I. grammatica* between both estimates of genetic diversity and species richness, and especially Shannon diversity (Figure 4.2, g-h, j-k), this was not significant (Table 4.1). There was only one site that showed significantly lower genetic diversity than all others (Table 3.4b), and although this site was also one of the lowest in species diversity (second or third lowest depending on the index, Appendix K, Figure K1), this site is shown as an outlier in the correlation (Figure 4.2, g-h, j-k). Again, functional diversity showed no correlation with genetic diversity.

Within *B. rhodani* no correlation was found between genetic diversity and species diversity (Table 4.1, Figure 4.2, m-r).

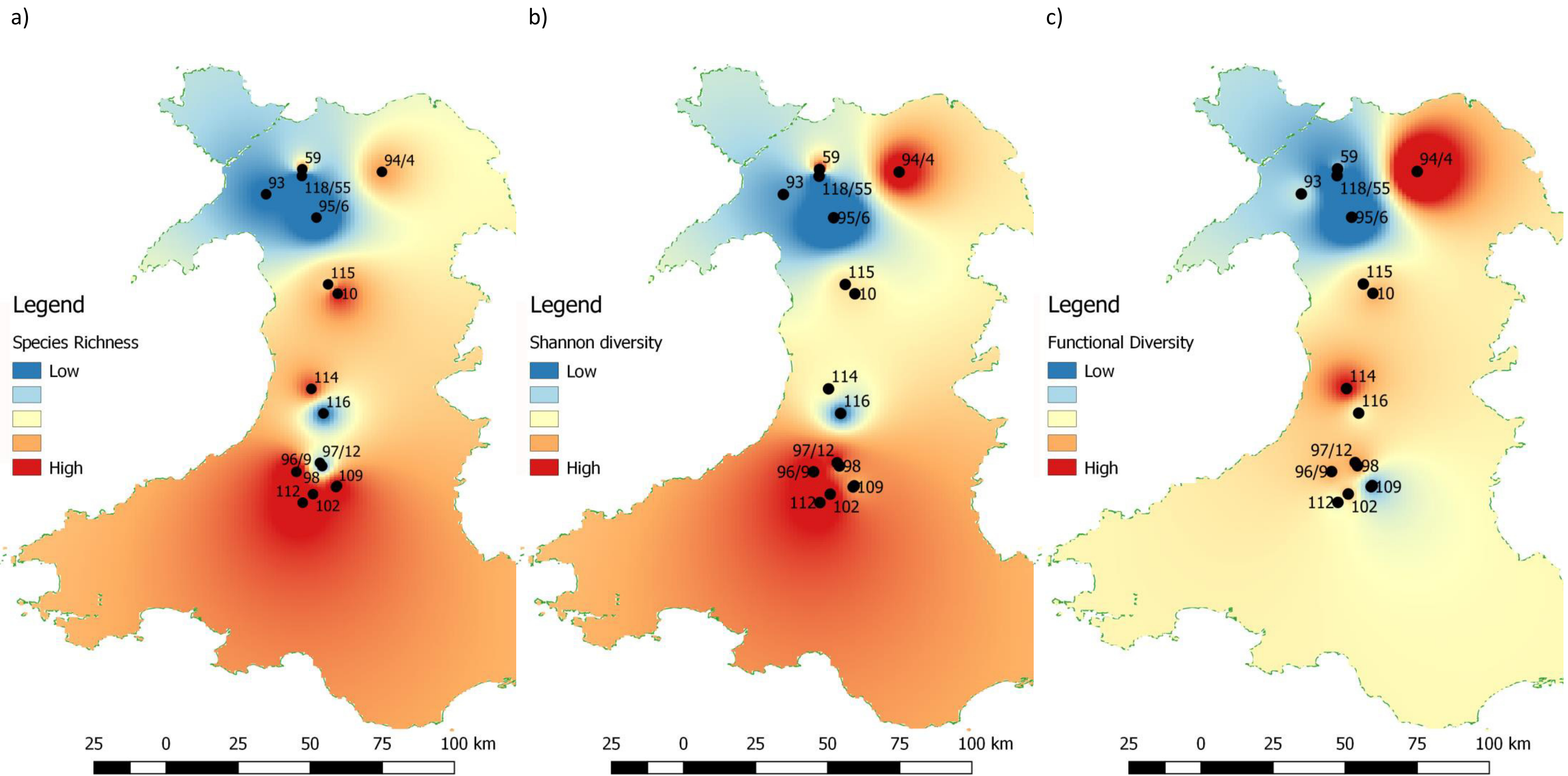
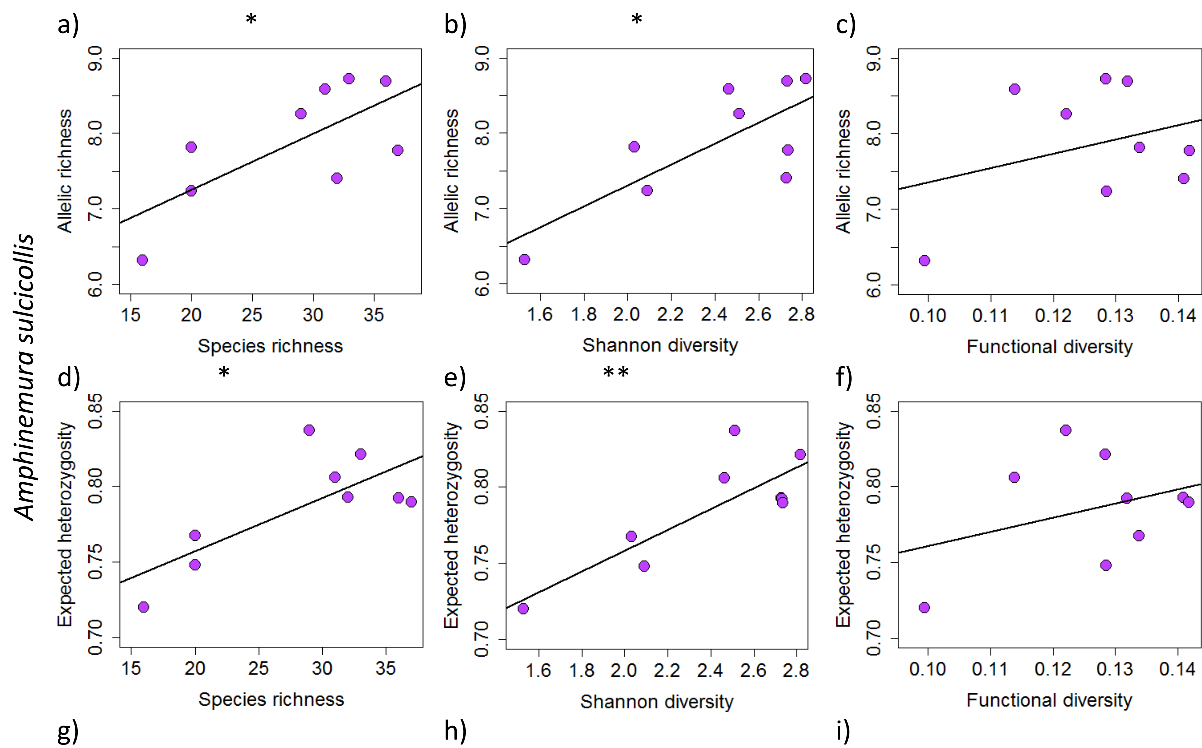


Figure 4.1. Interpolation maps created using QGIS across sites sampled in Wales, UK, showing a) species richness, b) Shannon diversity and c) functional diversity. Values for all indices are available in Appendix K Table K4. If data were available at a site for more than one year, a mean value was used.

Table 4.1. Summary of all mixed linear models preformed between the two indices of genetic diversity; (always on the y-axis) H_e = expected heterozygosity; AR = mean allelic richness and three indices of species diversity (on the x-axis); species richness, Shannon diversity and functional diversity. Showing adjusted R-squared (R^2), p-value (p) and number of sites (n) for each regression. Significant values in bold: * = p-value <0.05, ** = p-value <0.01.

			Species richness	Shannon diversity	Functional diversity
A. sulcicollis	H_e	R^2	0.48	0.63	0.00
		p	**0.022***	**0.007****	0.360
		n	9	9	9
	AR	R^2	0.44	0.51	-0.03
		p	**0.031***	**0.018***	0.404
		n	9	9	9
I. grammatica	H_e	R^2	-0.03	0.02	-0.12
		p	0.400	0.315	0.706
		n	9	9	9
	AR	R^2	0.06	0.09	-0.11
		p	0.256	0.224	0.680
		n	9	9	9
B. rhodani	H_e	R^2	-0.20	-0.10	-0.15
		p	0.896	0.531	0.654
		n	7	7	7
	AR	R^2	-0.16	-0.20	-0.20
		p	0.711	0.895	0.951
		n	7	7	7



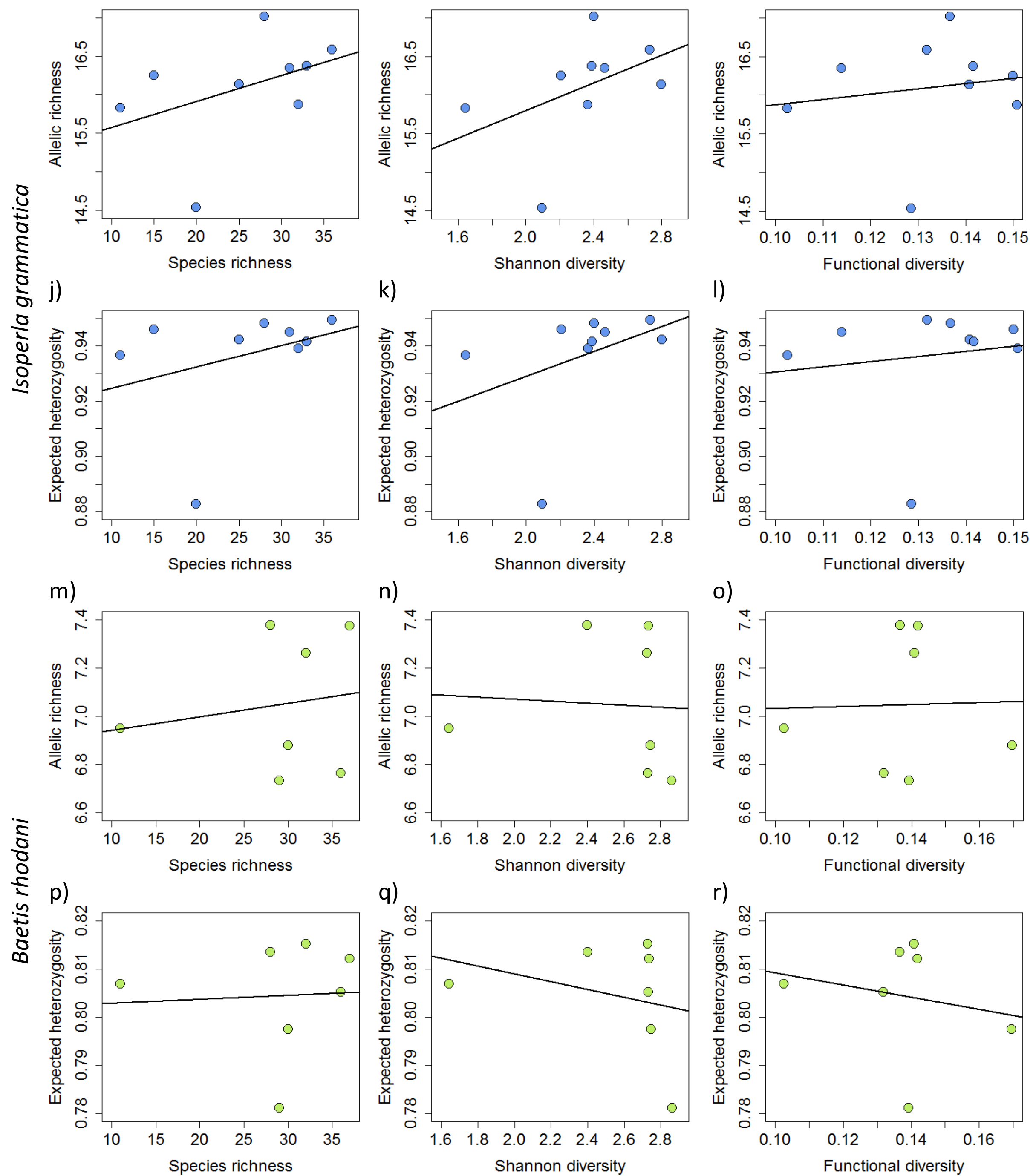


Figure 4.2. Correlations between genetic diversity indices (expected heterozygosity and allelic richness) and species diversity indices (species richness, Shannon diversity and functional diversity) for: a-f) *Amphinemura sulcicollis* (purple datum points); g-l) *Isoperla grammatica* (blue datum points) and m-r) *Baetis rhodani* (green datum points). Line of best fit created from a linear mixed model in RStudio. Stars above graphs denote significances: * = p-value <0.05, ** = p-value <0.01.

4.4.2 Species diversity vs environmental stressors

Each species diversity index (Appendix K, Table K4) was then compared to four environmental stressors (Appendix H, Table H1). Table 4.2 represents a summary of all pairwise mixed linear models and Figure 4.3 show example plots of these correlations. The most significant correlations were between pH and both species richness and Shannon diversity (Figure 4.3a-b), with low pH correlating with lower species diversity. Within Chapter 3, the genetic diversity of *A. sulcicollis* was shown to correlate with pH, to illustrate the patterns between genetic diversity, species diversity and pH, Figure 4.4 shows interpolation maps created in QGIS of each of these side by side. Figure 4.4 shows that where pH is low, particularly at sites 93, 95 and 116 it correlates with mean allelic richness of *A. sulcicollis* and in species richness. There was also a positive correlation between pH and functional diversity, though not significant (Figure 4.3c).

A significant positive correlation was also found between total oxidised nitrogen (TON) and Shannon diversity (Figure 4.3e). Although a slight positive correlation was found between TON and species richness, this was not significant (Figure 4.3d), and no correlation was found between TON and functional diversity (Figure 4.3f). No correlation was found between both metal stressors (Aluminium and Cadmium) and any species diversity index (Table 4.2).

Table 4.2. Summary of all mixed linear models preformed between the three indices of species diversity (always on the y-axis); species richness, Shannon diversity and functional diversity, and four environmental stressors (on the x-axis); aluminium (Al, µg/l), cadmium (Cd, µg/l), pH (pH-units) and total oxidised nitrogen (TON, mg/l). Showing adjusted R-squared (R^2), p-value (p) and number of sites (n) for each regression. Significant values in bold: * = p-value <0.05, ** = p-value <0.01, *** = p-value <0.001.

		Species richness	Shannon diversity	Functional diversity
Al	R^2	-0.04	-0.01	-0.06
	p	0.505	0.355	0.592
	n	14	14	14
Cd	R^2	-0.09	-0.08	0.16
	p	0.851	0.805	0.098
	n	13	13	13
pH	R^2	0.38	0.5877	0.17
	p	**0.009****	**0.001*****	0.069
	n	15	15	15
TON	R^2	0.11	0.26	0.01
	p	0.116	**0.025***	0.299
	n	16	16	16

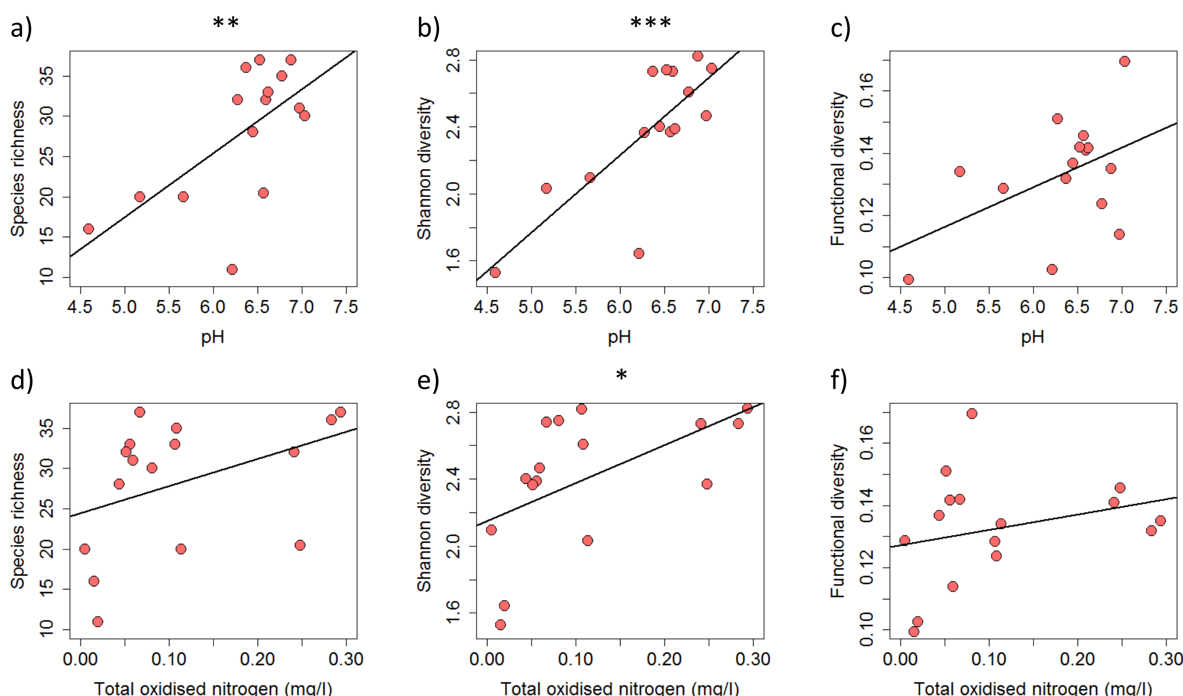


Figure 4.3. Correlations between species diversity indices (species richness, Shannon diversity and functional diversity) and environmental stressors a-c) pH, d-f) total oxidised nitrogen. Line of best fit created from a linear mixed model in RStudio. Stars above graphs denote significances: * = p-value <0.05, ** = p-value <0.01, *** = p-value <0.001.

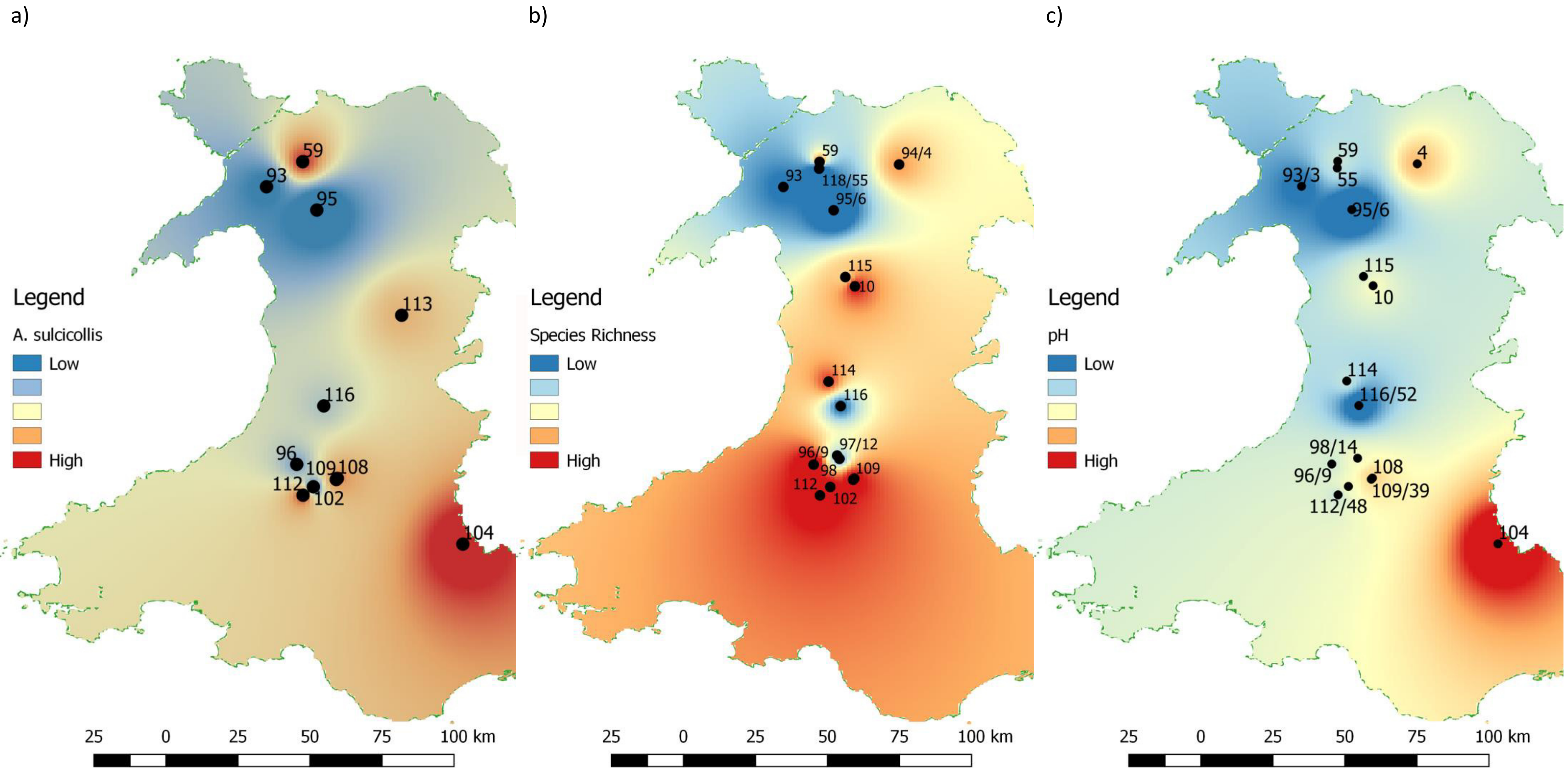


Figure 4.4. Side by side interpolation maps of a) the mean allelic richness of *Amphinemura sulcicollis*, b) species richness and c) pH. These three factors have all been shown to correlate with one another.

4.5 Discussion

Investigating the relationship between species and genetic diversity in three freshwater macroinvertebrates identified different patterns. It was hypothesised that *Amphinemura sulcicollis* was the most likely to show a SGDC as this species had the showed the most genetic structure in the previous chapter. As expected *A. sulcicollis* revealed a significant positive correlation between genetic diversity (allelic richness and expected heterozygosity) and species diversity (species richness and Shannon diversity). *Isoperla grammatica* showed a slightly positive correlation between the same factors but this was not significant, and *B. rhodani* showed no correlation.

The underlying driver of the significant positive correlation in *A. sulcicollis*, appears to be acidity (Figure 4.4). There was a significant positive correlation between pH and both species diversity (species richness and Shannon diversity; Figure 4.3a-b) and genetic diversity of *A. sulcicollis* (Figure 3.5a-b). Though a correlation does not necessarily mean causation, a reduction in species diversity at acidic streams is commonly seen, as the acid sensitive species are displaced by a few tolerant species (Ledger and Hildrew 2005; Svitok *et al.* 2014; Shaw *et al.* 2015). The effect of acidity on genetic diversity, on the other hand, is species-specific, even within this study. The genetic diversity of *B. rhodani* showed no correlation with acidity because it is acid sensitive and was not present at sites with low pH (Kowalik and Ormerod 2006; Murphy *et al.* 2013; Murphy *et al.* 2014). *Isoperla grammatica* showed a non-significant positive correlation between acidity and expected heterozygosity (Table 3.6) driven by just one site (site 93) having significantly lower genetic diversity, likely to be a result of low pH.

Within *A. sulcicollis* the genetic correlation with low pH was not universal throughout the sites, for example site 116 also had a low pH and a reduction in species diversity (Figure 4.4) but genetic diversity was not significantly lower (Table 3.4a; Appendix K, Figure K1). We suggest this is because this site shows no genetic isolation (during STRUCTURE analysis), and instead has high gene-flow with the surrounding sites (as shown in all three species studied in southern Wales) therefore low pH at the site does not exert the same influence on genetic diversity in the same way as species diversity. In Chapter 3, *A. sulcicollis* showed the

most structure (Figure 3.1a-c) of the three species, showing genetic isolation particularly at sites 95 and 93, - these sites have low pH which may explain the positive correlation between species and genetic diversity. *Isoperla grammatica* was only isolated at one site (Figure 3.2a-c) and *B. rhodani* had very weak genetic structure (Figure 3.3a-c) therefore a significant SGDC was unlikely to be found.

Although it has been observed that studies with discrete sampling units produce stronger positive SGDCs (Vellend 2014), previous studies have tended not to describe explicitly the genetic structure of the species they are investigating (e.g. Fray *et al.* (2016), Xu *et al*. (2016), Baselga *et al.* (2013) and Blum *et al.* (2012)). Instead, previous SGDC research is focused on the general pattern of genetic diversity, therefore most studies do not identify whether different sites are genetically differentiated. By first investigating the gene flow and genetic isolation of each species, this study was able to specifically investigate the role of isolation as well as possible drivers in SGDCs.

Baselga *et al.* (2013) is the only other study to investigate SGDCs in freshwater invertebrates that can disperse terrestrially, and though genetic isolation is not explicitly mentioned, they do conclude that their empirical results mirrored dispersal constrained models (i.e. the pattern of SGDC is due to limited dispersal which could drive isolation). Similarly, Lamy *et al.* (2013) concluded that habitat connectivity was the driver of SGDC for two freshwater gastropods. Sites that are more connected receive more alleles and species through immigration, resulting in greater species and genetic diversity. Vellend (2003) also attributed positive correlations as being a consequence of parallel effects of area and isolation among different islands studied.

A positive SGDC in the presence of genetic isolation is not always found, however, Albert *et al.* (2011) found that geographical isolation promotes species richness, whereas isolation would usually decrease the genetic diversity of a population, through mechanisms such as genetic drift and demographic bottlenecks, which may occur in the absence of an influence of the locality, or environmental heterogeneity driving a parallel response.

As hypothesised a combination of circumstances may be needed to drive a correlation between species and genetic diversity; 1) isolation, meaning that genetic diversity is being affected at the same spatial scale as species diversity; 2) a driver, where heterogeneity of the different localities produce patterns in both species and genetic diversity. Many studies conclude that it is particular drivers at a locality that influence both species and genetic diversity. For example, varying levels of anthropogenic disturbance, in the form of the amount of urbanised area (Frey *et al*. 2016) and stream conditions relating to agricultural land-use (Blum *et al.* 2012) both drove a positive correlation between species and genetic diversity, whereas soil pH and phosphorus availability led to a negative correlation (Xu *et al*, 2016). Within these studies the level of genetic isolation is not described so it is unknown whether isolation also had an effect. A large number of the empirical studies into SGDC have used plant species, and although all species have effective dispersal mechanisms, they rely on passive dispersal, possibly leading to greater structure and isolation between localities (Rundle *et al.* 2002; Tero *et al.* 2003). So although these studies do not mention isolation, it may still be an underlying cause. Derry *et al.* (2009) comparing boreal lakes, the first study on SGDC in the freshwater environment, found that species composition varied with acidity but haplotypes of zooplankton did not. This could be a species-specific response, a result of a lack of isolation or perhaps the CO1 mtDNA marker used did not reveal subtle genotypic differences.

Csergo *et al.* (2014) investigated SGDC within the habitat specialist plant, *Saponaria bellidifolia* and like in this current study (with Chapter 3), they first conducted a solely genetic investigation into the target species. They found that isolation as well as drivers within the locality (in their case outcrop area / heterogeneity), influenced genetic diversity (Csergö *et al.* 2009). They later compared the genetic diversity results with species diversity and as they expected found parallel patterns because of the isolation in the study area (an island-like system of outcrops) as well as heterogeneity of the sites acting as a drivers of the parallel patterns.

Though correlation was found with species richness and Shannon diversity, no correlation was found comparing the relationship between genetic diversity and functional diversity. Functional diversity represents particular life history traits of the species present at the sites (Appendix K, Table K3). As functional diversity does not vary a great deal between sites with

environmental heterogeneity, and is not correlated with any environmental stressors investigated, it could be that the traits used to compile the functional diversity are not the traits affecting an individual’s sensitivity to acidity or metals. This suggests that the explanatory factor is an as yet unexplored trait, for example, a physiological ability to remove metals (Buchwalter *et al.* 2008) or possibly, the genotype of the species.

In conclusion, though species diversity and genetic diversity can correlate, as shown here with *A. sulcicollis* and numerous other empirical studies (Table 1.3), correlation depends upon multiple factors coinciding and therefore should not be assumed. Studying three freshwater invertebrates simultaneously, that vary in their sensitivity to certain stressors (drivers), and in studying their genetic structure, has helped highlight possible causes. In this case it is believed that a positive SGDC was found within *A. sulcicollis* due to isolation between sites and acidity driving a parallel response in both species and genetic diversity.

Chapter 5 - A preliminary investigation of demographic resilience in three invertebrate species

Highlights

- Certain sites within *Amphinemura sulcicollis* and *Isoperla grammatica* had significantly reduced genetic diversity, compromising their resilience to environmental change. The aim of this chapter was to analyse the underlying demographic cause of this loss of diversity.
- Signatures of genetic bottlenecks were tested using three different approaches using the software BOTTLENECK, MSVAR and MPVal.
- BOTTLENECK did not detect signatures of recent demographic bottlenecks, probably reflecting the application of a model dependent on excess heterozygosity to three species characterised by heterozygosity deficiency.
- MSVAR, a full Bayesian likelihood coalescent approach, revealed bottlenecks for all clusters, sites and species tested. This points to a common ancient, possibly post-glacial, bottleneck from which all three species have expanded, but does not explain differences among sites within the three species.
- MPval did detect bottleneck signatures in the sites that were identified as having lower genetic diversity in Chapters 3 and 4, suggesting that differences in genetic diversity at these sites could be due to recent bottlenecks.
- Acidification, and acid episodes, are a likely previous cause of bottlenecks at the sites studied.
- Further analysis is required to pinpoint the magnitude and duration of the recent bottlenecks detected, for example using Approximate Bayesian Computational methods.

5.0 Abstract

Variation in genetic diversity among species and its determinants is a central issue in evolutionary research, particularly in understanding ecosystem and species resilience. Previously, genetic diversity patterns were identified in three freshwater invertebrates *Amphinemura sulcicollis*, *Isoperla grammatica* and *Baetis rhodani*. This chapter aims to investigate possible causes for these patterns, hypothesising that low genetic diversity could be explained by recent, anthropogenically mediated demographic bottlenecks. Signatures of population decline were investigated using three commonly used methods; the heterozygosity excess test in BOTTLENECK, full Bayesian coalescent simulation in MSVAR and the allele sized based M-ratio in MPVal. The different methods used gave different results; heterozygosity excess tests revealed no signature of a bottleneck, though this may be because the species being assessed are characterised by high polymorphism and heterozygosity deficiency. The full Bayesian coalescent approach revealed a signature consistent with a historic bottleneck 1 - 10 thousand years ago that affected all three species, though this event is unlikely to explain the current patterns across sites. The M-ratio method (assuming pre-bottleneck theta values of 50, 500 and 60 for *A. sulcicollis*, *I. grammatica* and *B. rhodani* respectively) identified differences between sites and clusters within each species. Results from *A. sulcicollis* and *I. grammatica* supported the hypothesis of recent bottlenecks at the sites and Clusters identified as having significantly lower genetic diversity in Chapters 3 and 4. *Baetis rhodani* which varied less in genetic diversity, showed evidence of bottlenecks at more sites. It is proposed that post-industrial population declines were caused by region-wide acidification and site-specific episodes of acidity combined, causing genetic bottlenecks in all three species studied. These patterns are consistent with genetic response to past disturbances, although subsequent population recovery illustrates a degree of resilience to these events, though further research is necessary to fully explore this theme.

Key words: genetic diversity, population decline, BOTTLENECK, MSVAR, M-ratio, acidification, resilience

5.1 Introduction

Resilience and biodiversity are intrinsically linked, it is generally believed that the more diversity a system features, the more resilient it is (Oliver *et al.* 2015). Resilience, defined in ecology as the capacity of an ecosystem to respond to a disturbance by resisting damage and recovering quickly (Hodgson *et al.* 2015) is one of the policy buzzwords of the moment (Newton 2016). It has been used in the Convention on Biological Diversity's 2020 goal to "enhance ecosystem resilience" (Convention on Biological Diversity 2012) and in legislation from the Welsh Government which calls to "support social, economic and ecological resilience" as well as reversing the decline in biodiversity in Wales (Welsh Government 2016). The ability of a system to resist and recover from a disturbance is a central issue in conservation, but also affects a system's ability to produce ecosystem services (Nimmo *et al.* 2015), therefore it is logical that the study of resilience is so popular and why policy makers have picked up on the term. Research on ecological resilience is usually focused at the ecosystem level and studied in reference to tipping points (Newton 2011; Petraitis 2013; Newton and Cantarello 2015), however the genetic diversity of individual species within an ecosystem has been shown to effect the overall ecosystem resilience (Hughes and Stachowicz 2004; Reusch *et al.* 2005). Therefore, as widely distributed species, *Amphinemura sulcicollis*, *Isoperla grammatica* and *Baetis rhodani* could be used as model organisms to investigate genetic resilience within a freshwater ecosystem.

Resilience involves both resistance and recovery (Hodgson *et al.* 2015), therefore can be studied through investigation of past demographic history. Investigating fluctuations in population size is vital for understanding the impact past disturbances have had on the current population genetics of a species (Girod *et al.* 2011). Whether these disturbances are caused by climatic changes or human disturbances, they can have a key role in shaping biodiversity at the genetic, species and community levels (Banks *et al.* 2013). Potential causes of deviation in genetic diversity include natural selection and genetic drift, but it could also be a product of a recent disturbance, i.e. an event that causes a severe reduction in population size, which leads to reduction in genetic diversity, termed a bottleneck (Leblois *et al.* 2014). Populations suffering from a bottleneck tend to lose genetic variation and potentially suffer a loss of evolutionary potential and an increased risk of extinction

(Frankham *et al.* 1999). As demographic bottlenecks reduce the population size, they in turn increase rates of inbreeding and the risk of deleterious alleles becoming fixed, which are concerning factors from a conservation perspective (Keller and Waller 2002; Leblois *et al*. 2014).

The determinants of genetic diversity have been a central issue in modern evolutionary theory ever since the synthesis of natural selection with genetics in the 1930s (Leffler *et al.* 2012; Huang 2016). The study of understanding the nature of genetic variation has therefore had a long history but still remains fundamental to evolutionary research today (Huang 2016). Direct measurement of the impact of disturbances requires long-term studies that may be impossible for many species (Girod *et al.* 2011). However using the distribution of present-day genetic variation to infer past demographic changes is becoming increasingly popular (Beaumont *et al.* 2001; Girod *et al.* 2011; Hu *et al.* 2011).

During Chapter 3 and 4 the genetic diversity of three freshwater invertebrates, *A. sulcicollis, I. grammatica* and *B. rhodani*, were analysed. It was found that for *A. sulcicollis* and *I. grammatica* there were certain sites that showed significantly lower genetic diversity when compared to others. Here the aim was to investigate possible causes of the genetic diversity variation seen in previous chapters. Specifically, three different methods: BOTTLENECK; MSVAR using the full Bayesian method and MPval using the M- ratio method, were implemented to assess whether signatures of demographic change can be found in any of the three species. It is hypothesised that the reduced genetic diversity in certain sites within *A. sulcicollis*, and *I. grammatica* (as identified with Chapter 3 and 4) were caused by a recent bottleneck. It was also an aim of the study to compare the results of the three different methods, and it is hypothesised that due to the differences in assumptions of the different approaches they may not reveal the same signal. A broader aim of this research is to infer the genetic resilience of these species and how they have responded to past disturbance.

5.2 Methods

Genotype data were collected and scored as described in Chapter 2 (see Appendix F for full dataset), and related individuals were removed as described in Chapter 3. Demographic change was then investigated using three different but complementary methods. As in

Chapter 3, each species was investigated 'per site' (for all sites that contain more than 10 individuals) and per cluster (all individuals based on population differentiation results laid out in Chapter 3). To prevent mixtures of genetic signal individuals for each cluster were chosen based on q – values from STRUCTURE output data using the threshold 0.8 (specifically how the data were split into clusters (including sample sizes) is summarised in Appendix L, Figure L1-3, Table L1). Due to the large computational requirements by MSVAR, large clusters were constrained to no more than 50 randomly sampled individuals (as in Orozco-terWengel *et al.* (2013)), and the same datasets were used for all three methods so they could be compared.

5.3 Statistical analysis

5.3.1 BOTTLENECK

The first approach used the program BOTTLENECK v1.2.02 (Cornuet and Luikart 1996; Piry *et al.* 1999). This program using the heterozygosity excess method where identification of significant heterozygosity excess suggests the presence of a bottleneck. This is because when a population undergoes a bottleneck, heterozygosity is not always reduced immediately but the rarer alleles are expected to be lost, so the test compares heterozygosity with allelic diversity, if the heterozygosity is higher relative to the allelic diversity, then a recent bottleneck could explain this (Luikart and Cornuet 1998; Luikart *et al.* 1998). This program carries out a sign test (Luikart and Cornuet 1998) and a Wilcoxon's signed rank test (Luikart et al. 1998). Both tests can be performed under three different models: the infinite allele model (IAM), stepwise mutation model (SMM) and the two-phase model (TPM).

The IAM predicts that mutation can occur across any number of repeats and assumes new alleles are produced with every mutation event thus allowing infinite allelic diversity. However, this model has been known to produce false positive results (O'Connell and Wright 1997). The SMM assumes mutations involve the loss or gain of one repeat unit at a time and may generate the same allele any number of times. The TPM however combines

single step and multiple step mutation models and is the most appropriate model for microsatellites (Luikart and Cornuet 1998).

The Sign test calculates the expected number of loci with heterozygosity excess and the observed number of loci with heterozygosity excess. Similarly the Wilcoxon's signed rank test calculates the probability of heterozygosity deficiency; probability of heterozygosity excess, and a two-tailed probability of heterozygosity deficiency or excess.

These tests were performed per site and per cluster (as in Chapter 3), using 1000 replications and assuming 95% single step mutations with a variance of 12, following recommendations by Piry *et al.* (1999). The mode-shift indicator test was also used to look at the allele frequency distribution; if a population is under mutation–drift equilibrium it is expected to be L-shaped, however if the population had undergone a bottleneck, rarer alleles would be removed and distribution would be a shifted mode (Luikart and Cornuet 1998).

5.3.2 MSVAR

Signatures of a demographic change were also investigated by using a coalescent-based full likelihood, Bayesian approach implemented in MSVAR v1.3 (Beaumont 1999; Storz and Beaumont 2002). This program can be used to investigate the genealogical history of microsatellite data using coalescent theory, which attempts to trace mutations shared by all members of a population to a single ancestral copy, the 'most recent ancestor' (Nordborg 2004). MSVAR estimates demographic and genealogical parameters such as current effective population size (N_0), effective population size before an event (N_1) and time since event (Ta), using MCMC (Markov chain Monte Carlo) simulations. The event could be a bottleneck or an expansion.

As MSVAR simulations are computationally very intensive, not all sites and clusters were used, but instead representative sites from each cluster, for each species and some additional clusters. For *A. sulcicollis* four sites representing all four clusters were tested (sites 104, 96, 95 and 93), additionally Cluster 3 was run to compare to site 95, and Cluster 4 was run to compare to site 93 (see Appendix L, Figure L1 and Table L1 for further details).

For *I. grammatica* two sites from the two representative clusters were tested (sites 108 and 93), additionally Cluster 2 was analysed to compare with site 93 (see Appendix L, Figure L2 and Table L1). Lastly, to represent *Baetis rhodani*'s Cluster 1, two sites were tested (sites 102 and 94, see Appendix L, Figure L3 and Table L1).

A large variance was set for the prior parameter distributions so they would have less effect on the posterior distributions (like in Hu et al. 2011), but a variety of priors were still tested that represented different scenarios, for example, a bottleneck, no change (stable) and an expansion (Figure 5.1). Ten independent runs (run 1 - 10, Table 5.1) were performed for each dataset to test how the prior distribution would influence the posterior distribution. Generation time (g) was assumed to be 1 year for all three species as they are mostly univoltine (see section 1.2). No mutation rate estimates are available for these species therefore the average microsatellite mutation rate (μ) of $1x10^{-4}$ was used (Brohede *et al.* 2002; Bulut *et al.* 2009; Orozco-terWengel et al. 2013). A line was recorded every 100'000 simulations until a total of 100'000 lines were recorded ($1\ x10^{10}$ simulations). The first 10% of each independent chain was discarded to avoid influence of the starting values in parameter estimation.

Output from MSVAR was interpreted using the package boa v1.1.7 (Smith 2007) as a plugin for R v3.0.2 (R Core Team 2013) to calculate mean values of N_0, N_1 and Ta for each independent run. Convergence between the different runs was also assessed using the Brooks, Gelman and Rubin Convergence Diagnostic test (Gelman and Rubin 1992; Brooks and Gelman 1998).

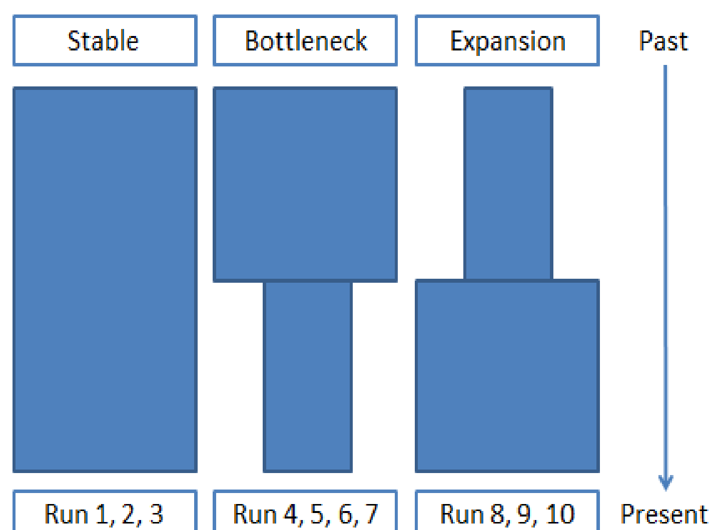


Figure 5.1. Depicting the three scenarios of effective population size that were tested using MSVAR runs 1-10: stable, bottleneck and expansion.

Table 5.1. Range of priors and variances tested in ten independent runs and what scenario they represent, using MSVAR. N_0 = current effective population time; *var* = variance; N_1 = Ancestral effective population size before event; μ = mutation rate; *Ta* = time interval between N_1 and N_0.

Run	Priors								Represents
	N_0	var	N_1	var	μ	var	Ta	var	
1	4	5	4	5	-4	5	4	5	Stable
2	4	10	4	10	-4	10	4	10	Stable, greater var
3	6	8	6	8	-4	8	4	8	Stable, higher N_o and N_1, medium var
4	3	5	5	5	-4	5	4	5	Bottleneck
5	3	10	5	10	-4	10	4	10	Bottleneck, greater variance
6	3	10	7	10	-4	10	6	10	Bottleneck, higher Nt_1 and Ta
7	5	8	7	8	-4	8	6	8	Bottleneck, higher N_o, N_1, medium var
8	5	5	3	5	-4	5	4	5	Expansion
9	5	10	3	10	-4	10	4	10	Expansion, greater variation
10	5	10	2	10	-4	10	6	10	Expansion, lower N_1 and higher Ta

5.3.3 MPVal

Thirdly, presence of a bottleneck signature was assessed using the M-ratio method used by the software MPVal and CriticalM (Garza and Williamson 2001). The M-ratio method

attempts to detect a recent bottleneck by, as with BOTTLENECK, detecting the loss of rare alleles. Instead of comparing the number of alleles to heterozygosity, however, MPVal uses a comparison between the number of different alleles per locus and the range in size of these alleles. Working on the assumption that during a bottleneck the number of different alleles will reduce but the size range will only reduce if the largest or smallest allele is lost, and since allele frequency distributions are rarely bell-shaped, this comparison (which is termed M), can be used to detect bottlenecks (Garza and Williamson 2001).

It is suggested that this method may be better suited than BOTTLENECK, to assess demographic change in large population sizes (Garza & Williamson 2001). The target species investigated here are characterised by very high polymorphism and Hardy Weinberg disequilibrium, but have correspondingly high size ranges (Appendix F), and fitting the assumptions of the M-ratio method, loci in all species rarely have bell-shaped allele frequency distribution, some loci have a right skewed distribution though many have very irregular patterns of loci size distribution.

The test was run using parameter values $\Delta g = 3.5$ (average size of mutation that is not one-step) and $p_s = 0.1$ (the percentage of time mutations are larger, i.e. not one step), these values are suggested by Garza and Williamson (2001) to be conservative.

The value for pre-bottleneck theta ($\theta = 4N_e\mu$) is unknown, therefore θ was estimated for all sites and clusters for each species using the program MSA v4.05 (Dieringer and Schlötterer 2003). *A. sulcicollis* ranged from a minimum of $\theta = 10$ (as the mutation rate is fixed by default in MPVal at 5×10^{-4} mutations per locus per generation (μ), this would equate to a N_e of 5000) to a maximum of $\theta = 50$ (Ne = 25,000), therefore the M- ratio method was tested at these two extremes. MSA calculated *I. grammatica*'s θ to be much higher at a minimum of $\theta = 100$ (N_e = 50,000) to a maximum of $\theta = 500$ (N_e = 250,000), corresponding with this species' polymorphism. *Baetis rhodani* showed a narrower range with a minimum of $\theta = 40$ (N_e = 20,000) to a maximum of $\theta = 60$ (N_e = 30,000). The average value of M of each site and cluster was then compared against a critical value (M_c) at two extremes of pre-bottleneck theta for each species to infer whether the population experienced a reduction in size.

5.4 Results

5.4.1 BOTTLENECK

As the Wilcoxon test has greater statistical power than the Sign test and is favoured in other studies (e.g. Costa *et al.* 2013; Minhós *et al.* 2016), the result of the Wilcoxon test under TPM only and the mode shift test are presented in Table 5.2 (other results can be found in Appendix L, Table L2a-c). Using this method, no significant heterozygosity excess was observed at any site or cluster in any of the three species and the mode-shift test mirrored this and revealed a normal L-shaped distribution for all species and sites tested (Table 5.2a-c). Heterozygosity deficiency on the other hand was observed, particularly within *B. rhodani*. However, no significant signature of a bottleneck was detected using this approach.

The Wilcoxon test under SMM supports this result, showing mostly heterozygosity deficiency rather than heterozygosity excess. Though the IAM model did show some significant heterozygosity, as this model is known to produce false positives (Appendix L, Table L2a-c). The sign test under all three models generally show more significant heterozygosity excess compared to the Wilcoxon test, however none that were consistent when comparing models and tests (Appendix L, Table L2a-c).

Table 5.2a-c. Showing results from the Wilcoxon test (two-phase (T.P.M.) model) and the mode shift test using BOTTLENECK for a) *Amphinemura sulcicollis*; b) *Isoperla grammatica* and c) *Baetis rhodani*. *P* = probability and *H* = heterozygosity. **Bold** = significance (p = <0.05).

Table 5.2a. *Amphinemura sulcicollis*.

Site	P (one tail for H deficiency)	P (one tail for H excess)	P (two tails for H excess or deficiency)	Mode shift
104	0.246	0.784	0.492	normal L-shaped distribution
109	0.278	0.754	0.557	normal L-shaped distribution
108	**0.007**	0.995	**0.014**	normal L-shaped distribution
102	0.500	0.539	1.000	normal L-shaped distribution
112	**0.016**	0.988	**0.032**	normal L-shaped distribution
96	0.097	0.920	0.193	normal L-shaped distribution
9	**0.016**	0.988	**0.032**	normal L-shaped distribution
116	0.138	0.884	0.275	normal L-shaped distribution
113	0.097	0.920	0.193	normal L-shaped distribution
95	0.161	0.862	0.322	normal L-shaped distribution
6	**0.042**	0.984	0.084	normal L-shaped distribution
59	**0.005**	0.997	**0.010**	normal L-shaped distribution
93	0.053	0.958	0.105	normal L-shaped distribution
Cluster 1	0.278	0.754	0.557	normal L-shaped distribution
Cluster 2	0.005	0.997	0.010	normal L-shaped distribution
Cluster 3	0.005	0.997	0.010	normal L-shaped distribution
Cluster 4	0.053	0.958	0.105	normal L-shaped distribution

Table 5.2b. *Isoperla grammatica*.

Site	P (one tail for H deficiency)	P (one tail for H excess)	P (two tails for H excess or deficiency)	Mode shift
105	0.577	0.461	0.922	normal L-shaped distribution
106	0.216	0.813	0.432	normal L-shaped distribution
108	0.903	0.116	0.232	normal L-shaped distribution
12	0.278	0.754	0.557	normal L-shaped distribution
97	0.652	0.385	0.770	normal L-shaped distribution
98	0.461	0.577	0.922	normal L-shaped distribution
112	0.539	0.500	1.000	normal L-shaped distribution
114	0.188	0.839	0.375	normal L-shaped distribution
115	0.313	0.722	0.625	normal L-shaped distribution
10	0.116	0.903	0.232	normal L-shaped distribution
118	0.313	0.722	0.625	normal L-shaped distribution
93	**0.001**	0.999	**0.003**	normal L-shaped distribution
Cluster 1	**0.042**	0.984	**0.084**	normal L-shaped distribution
Cluster 2	**0.012**	0.991	**0.024**	normal L-shaped distribution

Table 5.2c. *Baetis rhodani*.

Site	P (one tail for H deficiency)	P (one tail for H excess)	P (two tails for H excess or deficiency)	Mode shift
112	0.153	0.863	0.305	normal L-shaped distribution
102	**0.011**	0.996	**0.021**	normal L-shaped distribution
97	**0.011**	0.996	**0.021**	normal L-shaped distribution
96	**0.004**	0.997	**0.009**	normal L-shaped distribution
9	**0.040**	0.966	0.080	normal L-shaped distribution
106	**0.003**	0.997	**0.007**	normal L-shaped distribution
113	**0.002**	0.998	**0.004**	normal L-shaped distribution
115	**0.002**	0.999	**0.003**	normal L-shaped distribution
94	**0.040**	0.966	0.080	normal L-shaped distribution
118	0.064	0.945	0.127	normal L-shaped distribution
Cluster 1	**0.000**	1.000	**0.000**	normal L-shaped distribution

5.4.2 MSVAR

MSVAR was run for six different datasets for *A. sulcicollis* (60 independent runs); three different datasets for *I. grammatica* (30 independent runs) and two different dataset for *Baetis rhodani* (12 independent runs). As a result of the high polymorphism of each species, each of the 102 runs took over 30 days to complete; some *I. grammatica* runs took considerably longer, limiting what could be achieved in the time available (for this reason *B. rhodani* sites could not be tested at all ten independent runs and *I. grammatica* Cluster 2 and Site 93 had to be stopped before completion of $1x10^{10}$ simulations).

Testing for convergence between independent runs assures that the priors tested have not influenced the posterior and increases confidence in the result, however for these data the ten independent runs did not always converge satisfactorily (meaning that the multivariate potential scale reduction factor (MPSRF) was not always <1.2, which it ideally should be). For *A. sulcicollis* when comparing all ten independent runs only site 96 (1.08), site 93 (1.04) and cluster 4 (1.03) showed convergence. The MPSRF of sites 95, 104 and Cluster 3 however were higher than 1.2. The number of independent runs used in this study is higher than in the literature (ten compared to four as in Costa *et al.* 2013; Minhós *et al.* 2016), therefore convergence tests were re-run with just three independent runs (representing 1 stable, 1 bottleneck and 1 expansion); convergence was then found within site 95 (1.02 comparing run 2, 5 and 8) and site 104 (1.16 comparing run 1, 4 and 8) however Cluster 3 still did not

show convergence (1.69 comparing 1 4 and 8). Usually if convergence is not achieved it is recommended that simulations are run for longer, however these simulations were run for longer than other studies ($1x10^{10}$ compared with $5x10^{9}$ for polecats (Costa *et al.* 2013); $2x10^{9}$ for tomato frogs (Orozco-terWengel *et al*. 2013) and $9x10^{9}$ for colobus monkeys (Minhós *et al.* 2016)), and it was not possible to extend them.

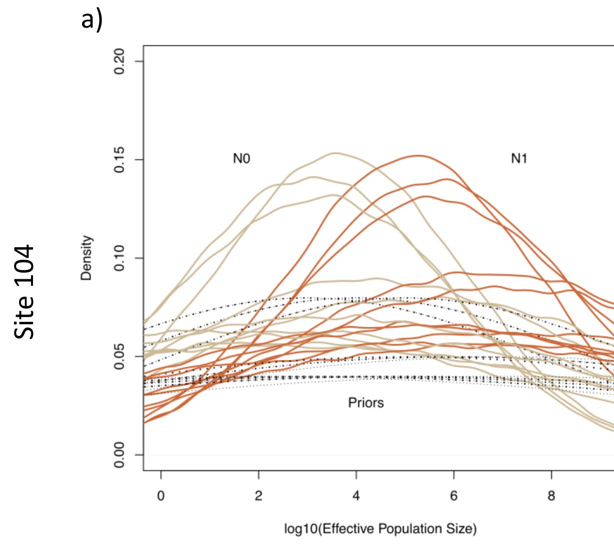
Similarly for *I. grammatica* none of the datasets converged using all ten runs and only Cluster 2 converged using just three runs (1.17 using run 1 4 and 8), whereas sites 108 and 93 did not converge. For *B. rhodani*, site 94 showed convergence with all 9 runs (1.09), however site 102 did not (3 runs 2, 5 and 6 had a MPSRF of 2.7).

Though the convergence tests must be taken into consideration when viewing the results, useful information can still be taken from all of the simulations. The posterior probabilities of each independent run were plotted for each species, comparing N_0 and N_1 and showing Ta separately. First, MSVAR showed a consistent signature of a population decline in all species, in almost all different MSVAR runs (Figure 5.2 – 5.4). Though there is variability in the values of N_1, N_0 and Ta for each species, regardless of cluster, the mean value for N_1 (ancestral effective population size, darker red line) was almost always a higher value compared to N_0 (the current population size, lighter red line) and were independent from the priors. Despite the convergence results, this was seen in almost all runs whether the priors represented stability, expansion or a bottleneck. Secondly, the time of this bottleneck was also very similar between species. Although we cannot be confident in the exact value of Ta (as it is quite variable), when removing obvious outliers in the data (see Appendix L, Table L3a-c) the time of the bottleneck ranged between $1\ x10^{3}$ and $1x10^{7}$ for *A. sulcicollis*, $1x10^{1}$ and $1x10^{5}$ for *I. grammatica* and $1x10^{2}$ and $1x10^{5}$ for *B. rhodani* and had a mean of between 1 and 10 thousand years ago for all three species (Figure 5.2 – 5.4).

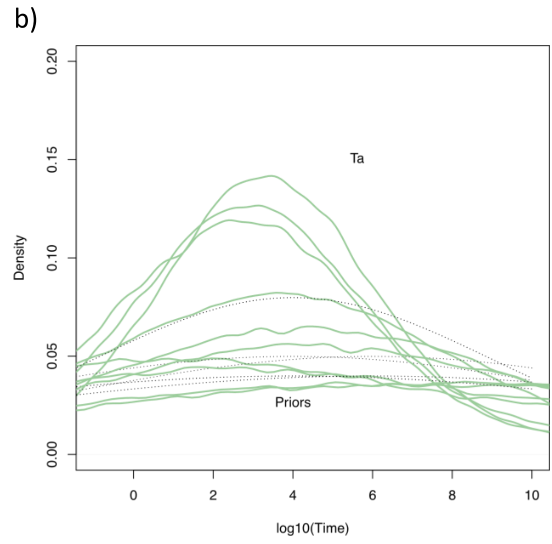
Summary statistics for N_0, N_1 and Ta (including Min, 1stQu, Median, Mean, 3rdQu and Max) for all MSVAR simulations are available in Appendix L (Figure L4 – L6, Table L3a-c).

Site 104

a)

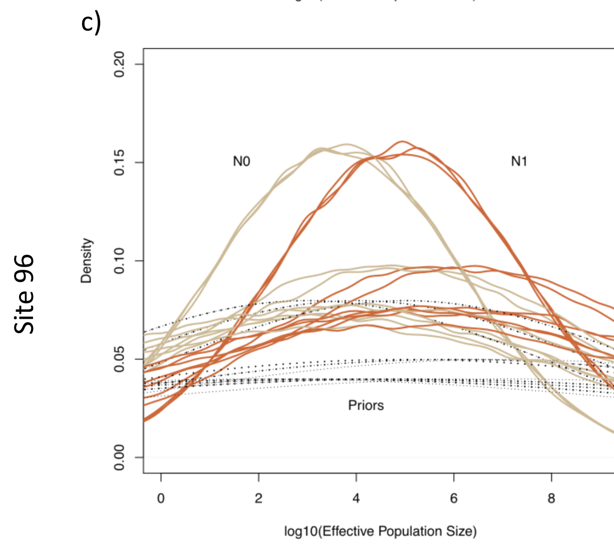


b)

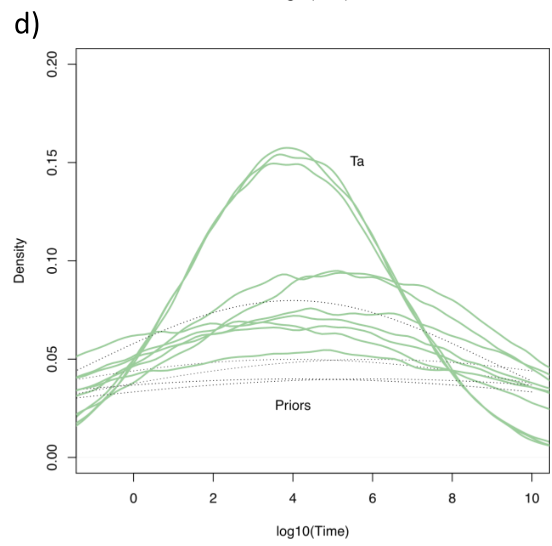


Site 96

c)

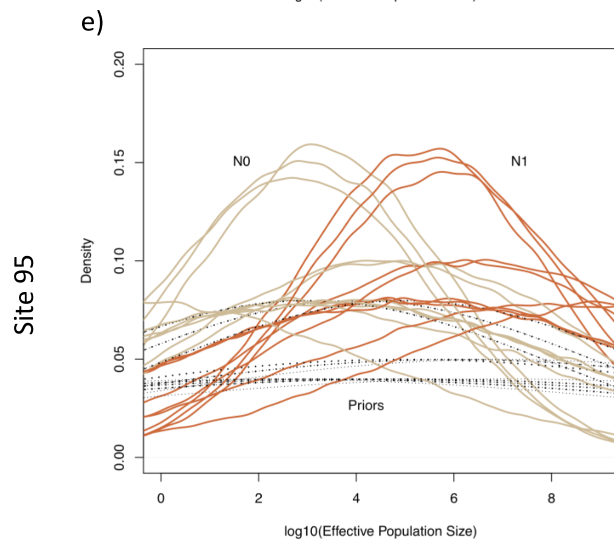


d)

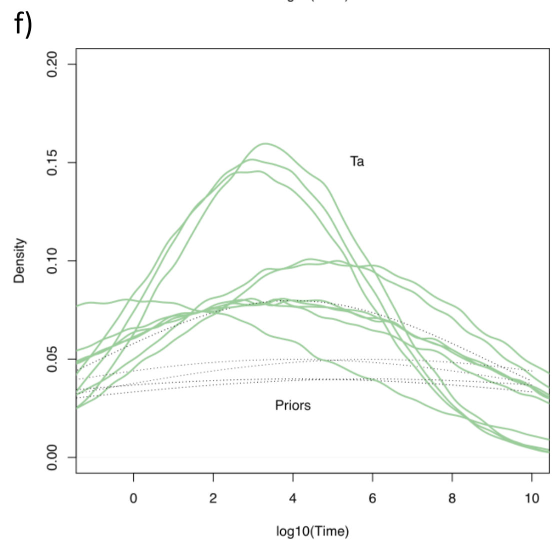


Site 95

e)



f)



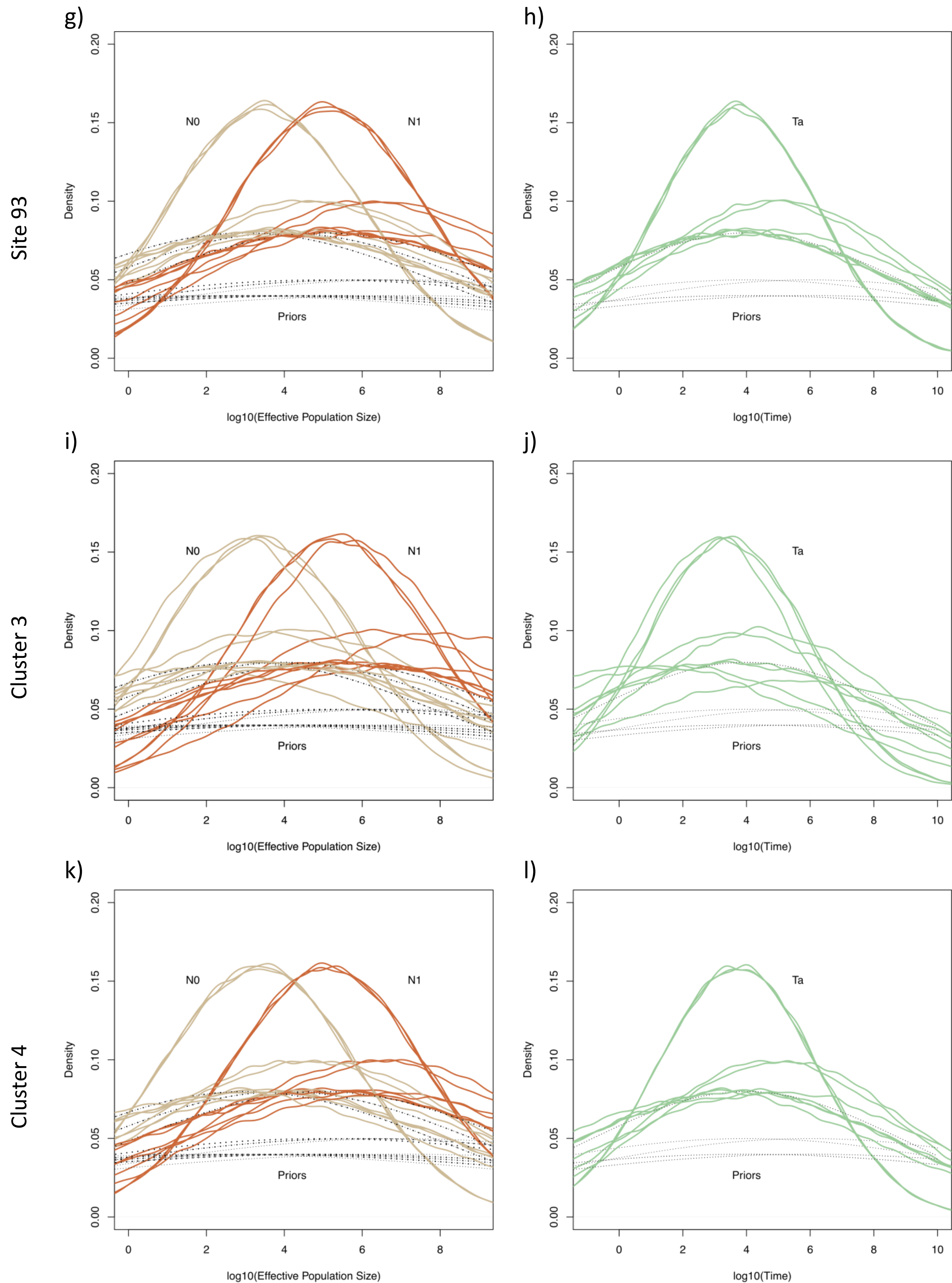


Figure 5.2. Showing the posterior distributions calculated using MSVAR comparing N_1 (effective population size at time Ta) = darker red line to N_0 (current effective population size) = lighter red line and posterior distributions of Ta (Time between N_0 and N_1) = green line, along with priors (dotted lines) for representative sites and clusters for Amphinemura *sulcicollis*. Each site/cluster contains ten

independent runs: a and b) Site 104; c and d) Site 96; e and f) Site 95; g and h) Site 93; I and j) Cluster 3; k and l) Cluster 4.

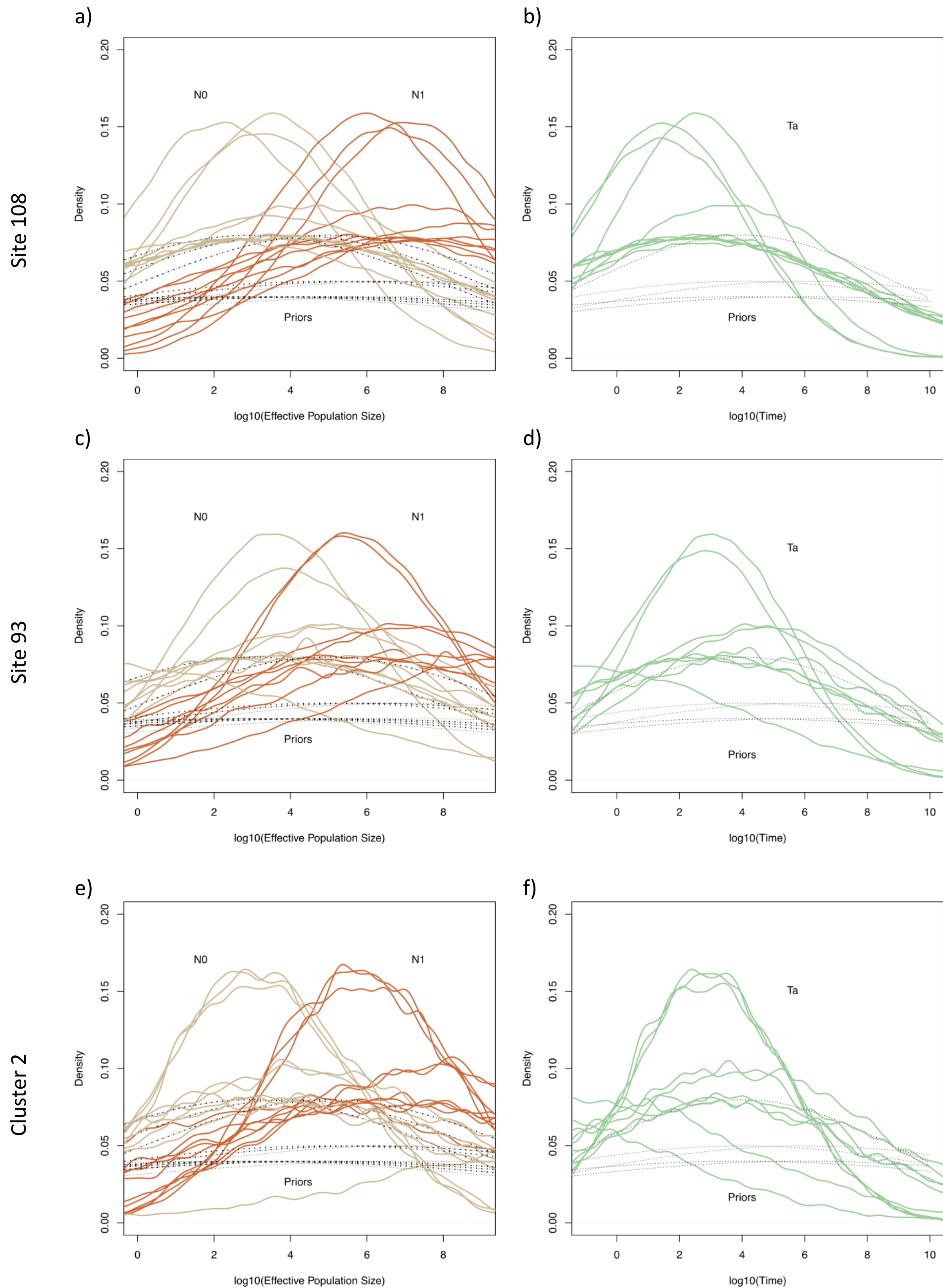


Figure 5.3. Showing the posterior distributions calculated using MSVAR comparing N_1 (effective population size at time Ta) = darker red line to N_0 (current effective population size) = lighter red line and posterior distributions of Ta (Time between N_0 and N_1) = green line, along with priors (dotted

lines) for representative sites and clusters for *Isoperla grammatica*. Each site/cluster contains ten independent runs: a and b) Site 108; c and d) Site 93; e and f) Cluster 2.

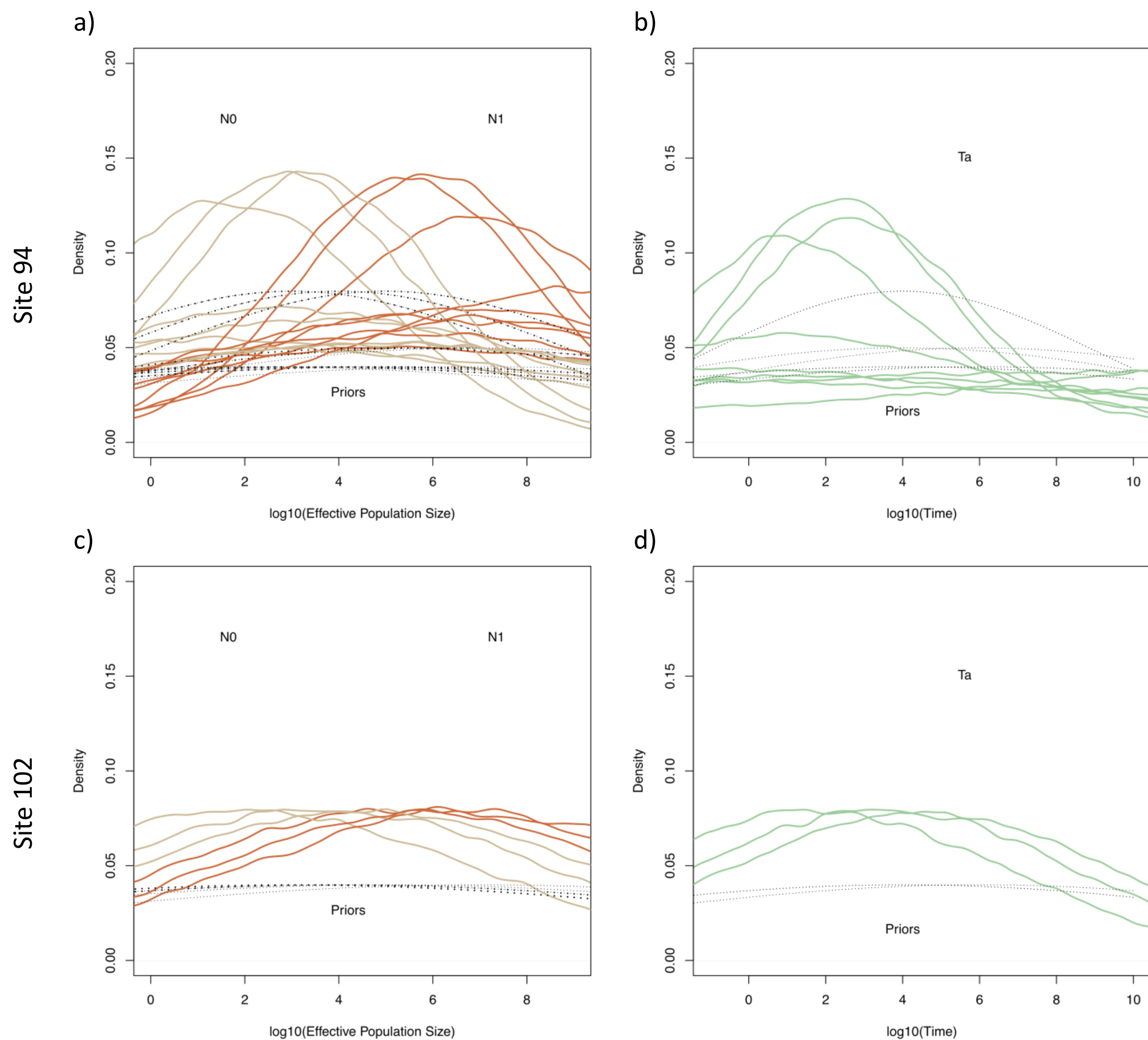


Figure 5.4. Showing the posterior distributions calculated using MSVAR comparing N_1 (effective population size at time Ta) = darker red line to N_0 (current effective population size) = lighter red line and posterior distributions of Ta (Time between N_0 and N_1) = green line, along with priors (dotted lines) for representative sites and clusters for *Baetis rhodani*. a and b) Site 94; c and d) Site 102; having 9 and 3 independent runs, respectively.

5.4.3 MPVal

The M-ratio method compares the average M of a site/cluster to a critical M (M_c) value assuming a certain pre-bottleneck theta size (θ), the higher the assumed θ the more

conservative the test. Particularly within *A. sulcicollis* and *I. grammatica* (which had the bigger ranges of θ according to results from MSA) the value of θ assumed had a significant impact on the results. When the Average M of all sites and clusters for *A. sulcicollis* was compared to the M_c assuming the lower θ of 10, all but site 104 show signatures of a bottleneck, however when the more conservative M_c was assumed (θ = 50), only sites 95, 6 (inter-annual sites at the same location) and 93, and Clusters 2 and 3 showed a signature of a bottleneck (Table 5.3a).

Similarly within *I. grammatica* when the Average M of all sites and clusters was compared to the M_c assuming the lower θ of 100, all sites and clusters apart from Cluster 1 showed evidence of a bottleneck, however when θ = 500 is assumed, no signature of a bottleneck was found in any site or cluster, apart from Cluster 2 (Table 5.3b). *B. rhodani* however, shows similar patterns between the two θ tested. At the lower range of θ all sites and clusters apart from 9 and 94 show a signature of a bottleneck, and when compared to the higher range of θ, this mostly remains the same apart from site 97 no longer shows evidence of a significant bottleneck (Table 5.3c).

Table 5.3a-c. Results from software MPVal and Critical M showing; *n* = sample size (2 x number of diploid individuals); *Site* = Site (cluster represented); *Average M* = as calculated using MPVal; M_c = Critical M value calculated assuming one of two extremes of θ (pre-bottleneck *theta*) and *d* = the different between the *Average M* and the M_c (*Average M* - M_c), where positive (blue) values indicate no signature of bottleneck, and negative (red) values indicate presence of a bottleneck. Calculated for a) *Amphinemura sulcicollis* b) *Isoperla grammatica* and c) *Baetis rhodani*.

Table5.3a. *Amphinemura sulcicollis*.

Site (cluster)	Average M	θ = 10		θ = 50	
		M_c	d	M_c	d
104 (1)	0.649	0.636	0.013	0.525	0.124
109 (2)	0.632	0.633	-0.001	0.514	0.118
108 (2)	0.631	0.633	-0.002	0.514	0.117
102 (2)	0.568	0.641	-0.073	0.525	0.043
112 (2)	0.673	0.641	0.032	0.525	0.148
96 (2)	0.603	0.628	-0.024	0.503	0.100
9 (2)	0.603	0.645	-0.042	0.532	0.070
116 (2)	0.591	0.646	-0.055	0.541	0.050
113 (2)	0.658	0.633	0.025	0.514	0.143
95 (3)	0.497	0.635	-0.138	0.514	-0.017
6 (3)	0.505	0.641	-0.136	0.525	-0.020
59 (2)	0.556	0.645	-0.088	0.532	0.024
93 (4)	0.490	0.633	-0.142	0.514	-0.024
Cluster 1	0.587	0.624	-0.037	0.494	0.093
Cluster 2	0.639	0.702	-0.063	0.663	-0.025
Cluster 3	0.594	0.682	-0.088	0.616	-0.022
Cluster 4	0.523	0.618	-0.095	0.479	0.044

Table 5.3b. *Isoperla grammatica*.

Site (cluster)	Average M	θ = 100		θ = 500	
		Mc	d	Mc	d
105 (1)	0.300	0.435	-0.134	0.264	0.036
106 (1)	0.363	0.487	-0.124	0.314	0.049
108 (1)	0.356	0.487	-0.130	0.314	0.042
12 (1)	0.356	0.487	-0.131	0.314	0.041
97 (1)	0.314	0.458	-0.144	0.286	0.028
98 (1)	0.384	0.487	-0.102	0.314	0.070
112 (1)	0.366	0.487	-0.121	0.314	0.052
114 (1)	0.344	0.476	-0.132	0.305	0.039
115 (1)	0.404	0.487	-0.083	0.314	0.089
10 (1)	0.329	0.487	-0.158	0.314	0.015
118 (1)	0.348	0.487	-0.139	0.314	0.034
93 (2)	0.327	0.476	-0.149	0.305	0.022
Cluster 1	0.632	0.626	0.007	0.493	0.139
Cluster 2	0.182	0.343	-0.161	0.190	-0.008

Table 5.3c. *Baetis rhodani*.

Site	Average M	θ = 40		θ = 60	
		Mc	d	Mc	d
112 (1)	0.465	0.544	-0.079	0.508	-0.043
102 (1)	0.514	0.578	-0.064	0.543	-0.030
97 (1)	0.539	0.544	-0.004	0.508	0.032
96 (1)	0.514	0.552	-0.038	0.517	-0.003
9 (1)	0.534	0.486	0.048	0.442	0.092
106 (1)	0.520	0.561	-0.041	0.527	-0.007
113 (1)	0.470	0.544	-0.074	0.508	-0.038
115 (1)	0.503	0.544	-0.040	0.508	-0.004
94 (1)	0.587	0.544	0.044	0.508	0.080
118 (1)	0.533	0.578	-0.045	0.543	-0.010
Cluster 1	0.528	0.682	-0.154	0.665	-0.137

To aid comparison within the Discussion, tables Table 5.4a-c summarise the results from BOTTLENECK, MSVAR and MPVal, as well as results from Chapter 3. These include observed and expected heterozygosity (H_o and uH_e) calculated in GenAIEx; deviations from Hardy–Weinberg equilibrium (HWE) calculated in Genepop; Inbreeding co-efficient (F_{IS}) and probability (p) calculated in Arlequin and mean allelic richness (adjusted for sample size: 'per site' analysis had a minimum sample size of 11, 14, and 8, whereas 'per cluster' analysis had a minimum sample size of 13, 8 and 47, for *A. sulcicollis*, *I.* grammatica and *B. rhodani*, respectively) calculated in FSTAT (see Chapter 3 for methods; 'per site' results are identical to those in Chapter 3, however 'per cluster' results have been re-calculated as this chapter uses different sample sizes, i.e. a maximum of 50, see Appendix L for 'how genotyping data were split into clusters').

Table 5.4a-c. Summary table showing: *Site (Cluster)* = site code with which cluster it represents in brackets; *n* = number of individuals; H_o = mean (across loci) observed heterozygosity; uH_e = mean (across loci) expected heterozygosity; *HWE* = p-value of Hardy-Weinberg Equilibrium test; F_{IS} = the inbreeding co-efficient; *p* = the F_{IS} p value; *AR* = mean (across loci) allelic richness adjusted for sample size. **Bold** represents significance. *BOTTLENECK*, *MSVAR* and *MPVal* represents a summary of these three methods, 'Yes' = this method found evidence of a bottleneck, 'No' = no evidence of bottleneck. For MPVal, this was based on the maximum θ tested for each species but * denotes when the result was consistent across both θ's tested.

Table 5.4a. *Amphinemura sulcicollis*.

Site (Cluster)	n	H_o	uH_e	HWE	F_{IS}	p	AR	BOTTLENECK	MSVAR	MPVal
104 (1)	17	0.676	0.842	**0.000**	0.222	**0.000**	9.00	No	Yes	No*
109 (2)	16	0.677	0.837	**0.000**	0.141	**0.000**	8.30	No	/	No
108 (2)	16	0.696	0.806	**0.000**	0.123	**0.000**	8.60	No	/	No
102 (2)	17	0.612	0.793	**0.000**	0.208	**0.000**	7.40	No	/	No
112 (2)	17	0.647	0.792	**0.000**	0.158	**0.000**	8.70	No	/	No
96 (2)	15	0.616	0.786	**0.000**	0.163	**0.000**	7.40	No	Yes	No
9 (2)	18	0.630	0.790	**0.000**	0.173	**0.000**	7.80	No	/	No
116 (2)	19	0.631	0.768	**0.000**	0.123	**0.002**	7.80	No	/	No
113 (2)	16	0.666	0.823	**0.000**	0.174	**0.000**	8.30	No	/	No
95 (3)	16	0.660	0.796	**0.000**	0.166	**0.001**	7.30	No	Yes	Yes*
6 (3)	17	0.573	0.720	**0.000**	0.174	**0.000**	6.30	No	/	Yes*
59 (2)	18	0.650	0.822	**0.000**	0.193	**0.000**	8.70	No	/	No
93 (4)	16	0.600	0.748	**0.000**	0.197	**0.000**	7.20	No	Yes	Yes*
Cluster 1	14	0.650	0.839	**0.000**	0.264	**0.000**	9.44	No	/	No
Cluster 2	50	0.636	0.808	**0.000**	0.187	**0.000**	8.48	No	/	Yes*
Cluster 3	33	0.600	0.755	**0.000**	0.194	**0.000**	6.98	No	Yes	Yes*
Cluster 4	13	0.631	0.747	**0.000**	0.153	**0.000**	7.70	No	Yes	No

Table 5.4b. *Isoperla grammatica*.

Site (Cluster)	n	H_o	uH_e	HWE	F_{IS}	p	AR	BOTTLENECK	MSVAR	MPVal
105 (1)	15	0.663	0.943	**0.000**	0.308	**0.000**	15.76	No	/	No
106 (1)	20	0.640	0.939	**0.000**	0.329	**0.000**	16.13	No	/	No
108 (1)	20	0.692	0.945	**0.000**	0.291	**0.000**	16.35	No	Yes	No
12 (1)	20	0.628	0.936	**0.000**	0.338	**0.000**	15.92	No	/	No
97 (1)	17	0.585	0.948	**0.000**	0.393	**0.000**	16.36	No	/	No
98 (1)	20	0.688	0.946	**0.000**	0.291	**0.000**	16.25	No	/	No
112 (1)	20	0.666	0.949	**0.000**	0.305	**0.000**	16.58	No	/	No
114 (1)	19	0.671	0.939	**0.000**	0.306	**0.000**	15.87	No	/	No
115 (1)	20	0.700	0.948	**0.000**	0.278	**0.000**	17.01	No	/	No
10 (1)	20	0.705	0.942	**0.000**	0.268	**0.000**	16.37	No	/	No
118 (1)	20	0.641	0.937	**0.000**	0.328	**0.000**	15.83	No	/	No
93 (2)	19	0.710	0.883	**0.000**	0.213	**0.000**	14.54	No	Yes	No
Cluster 1	50	0.662	0.953	**0.000**	0.277	**0.000**	10.89	No	/	No*
Cluster 2	9	0.679	0.859	**0.000**	0.171	**0.000**	8.72	No	Yes	Yes*

Table 5.4c. *Baetis rhodani*.

Site (Cluster)	n	H_o	uH_e	HWE	F_{IS}	p	AR	BOTTLENECK	MSVAR	MPVal
112	16	0.572	0.805	**0.000**	0.296	**0.000**	6.76	No	/	Yes*
102	20	0.616	0.815	**0.000**	0.266	**0.000**	7.26	No	Yes	Yes*
97	16	0.612	0.781	**0.000**	0.229	**0.000**	6.73	No	/	No
96	17	0.618	0.818	**0.000**	0.255	**0.000**	7.45	No	/	Yes*
9	11	0.640	0.812	**0.000**	0.253	**0.000**	7.37	No	/	No*
106	18	0.576	0.786	**0.000**	0.281	**0.000**	6.52	No	/	Yes*
113	16	0.640	0.778	**0.000**	0.186	**0.000**	6.87	No	/	Yes*
115	16	0.640	0.813	**0.000**	0.235	**0.000**	7.38	No	/	Yes*
94	16	0.607	0.798	**0.000**	0.249	**0.000**	6.88	No	Yes	No*
118	20	0.640	0.807	**0.000**	0.219	**0.000**	6.95	No	/	Yes*
Cluster 1	50	0.629	0.798	**0.000**	0.162	**0.000**	15.22	No	/	Yes*

5.5 Discussion

Demographic changes, such as bottlenecks, leave signals in the genome that can be identified by simulating the genealogy they are expected to create using microsatellite loci (Cornuet and Luikart 1996; Beaumont 1999; Garza and Williamson 2001; Minhós et al. 2016). The aim was to better understand the variation of genetic diversity in the three species studied; *A. sulcicollis*, *I. grammatica* and *B. rhodani* and to infer their genetic resilience in the face of disturbance. Previously it was found that at site 93, both *A. sulcicollis* (Cluster 4), and *I. grammatica* (Cluster 2) were shown to have significantly lower genetic diversity, additionally for *A. sulcicollis* inter-annual sites 95 and 6 (Cluster 3) also showed significantly lower genetic diversity than the rest of the sites studied (Figure 3.4a-b). *Baetis rhodani* showed less variation in genetic diversity across sites, though site 106 was shown to have a lower genetic diversity. It was hypothesised that reductions in genetic diversity at these sites could be due to a recent bottleneck.

5.5.1 Statistical inference

At first inspection the results from the three methods used to assess past demographic change seem to contradict one another. However, this was not unexpected, as they each have their limitations and assumptions (see Section 5.3) and do not capture the same signals from the data (Minhós *et al.* 2016). BOTTLENECK is a well-established method used in many studies (e.g. Bernard *et al.* 2016; Cornetti *et al.* 2016; Graignic *et al.* 2016) and follows the assumption that a sudden population decline will reduce the frequency of rare alleles in a population at a faster rate than heterozygosity will decline and searches for heterozygosity excess in comparison to the number of alleles. The program BOTTLENECK detected no signals of demographic change within any of the target species, however, as shown by Tables 5.4a-c, in every species, whether analysed by site or cluster, uH_e is always higher than H_o which leads to significant Hardy-Weinberg disequilibrium because of heterozygosity deficiency. Additionally while heterozygosity is found to be low, the allelic richness is very high in all species, therefore, it is not surprising that this method cannot detect heterozygosity excess, hence a signal of demographic change, in these species.

MSVAR, using a full Bayesian method has been shown to be more powerful and reveal bottlenecks from further back in time when compared to methods which utilise summary statistics like BOTTLENECK (Girod *et al.* 2011; Peery *et al.* 2012; Leblois *et al.* 2014). Using coalescent theory, MSVAR showed a consistent pattern of population decline in all sites and clusters tested in all three species. However, Chikhi *et al.* (2010) warn that genetic differentiation/gene flow, genetic diversity, and the sampling scheme, can all potentially create false bottleneck signals within MSVAR, especially in populations with large effective sizes. Here, this was counteracted by testing both example sites and separating the data into clusters based on their genetic structure and removing those individuals that may confuse the genetic signal (Appendix L). In this case considering that the same signal is seen in both sites and clusters for all species, it appears to be a true signal of an ancient bottleneck which affected all three species. Because of convergence issues, the exact timing of the bottleneck cannot be certain, however in most independent runs, in all species it was more than 10 thousand years ago, possibly suggesting that all three species went through a bottleneck during the last glacial maximum (21 thousand years ago, (Theissinger *et al.* 2013), with deglacial events between 40 and 12 thousand years ago (Bowen *et al.* 2002)).

Genetic studies have found signatures of ancient lineages in other freshwater invertebrate species. Endo *et al.* (2015) identified areas of high genetic diversity in a collection of alpine species, and concluded that diversification events could have dated back to the early to mid-Pleistocene, and that diversity hotspots were possible glacial refugial communities. Theissinger *et al.* (2013) investigated the Pleistocene and Holocene history of the stonefly *Arcynopteryx dichroa*, and concluded that the genetic diversification seen was linked to glacial expansion and contraction events in the Pliocene and Pleistocene periods. However, neither of these studies specifically investigated bottlenecks in their species; however they inferred them from genetic diversity estimates. Since the aim here was to investigate recent bottlenecks to try and explain the differences in genetic diversity and resilience within and between the three species, and as others have noted, MSVAR is not likely to detect recent declines (Girod *et al.* 2011; Salmona *et al.* 2012; Dussex *et al.* 2015).

Lastly, MPVal was used to assess each species using the M-ratio method. This method is the only one that identified differences within and between species, especially when the results are compared to the most conservative M_c (calculated with the highest value of θ). The

range of θ for each species was calculated from the data, where the sites and clusters with the lowest genetic diversity gave the value used for the minimum θ, and those with the highest genetic diversity gave the value for the maximum θ. Considering the hypothesis being tested is that sites with low genetic diversity may have undergone a bottleneck, it follows that the results should be compared to a M_c which assumes a pre bottleneck θ of the highest found in that respective species, as it is assumed that the sites with high genetic diversity (therefore high value of θ) have not gone through a bottleneck. The higher value of θ for each species is also better supported by values of ancient effective population size (N_1) calculated in MSVAR. Therefore, using the most conservative M_c, within *A. sulcicollis* M-ratio finds that sites 93, 95 and 6 show signals of recent bottlenecks, and site 104 is the only site that does not show a bottleneck when compared to either M_c. This mirrors the genetic diversity data (Table 5.4a) and supports the hypothesis that a bottleneck may be responsible for this reduction in genetic diversity. When analysing the data 'per cluster' however, Cluster 2 shows a signal of bottleneck which was not expected. It is possible that, because this dataset consists of a randomly selected 50 individuals from a large group, this has affected the ratio between the allele size range and allele frequency distribution (Appendix L, Table. L1.). Cluster 3, which represents individuals from sites 95 and 6 (as well as partial sites), mirrors the 'per site' result and shows a bottleneck; Cluster 4 however does not, despite being made up of individuals from site 93 (which did), though perhaps this is due to the sample size (13 rather than 16 individuals) which has again affected the M-ratio in an unforeseen way. To find a signature of a bottleneck at these specific sites, especially when compared to the other statistics in Table 5.4a that show consistent Hardy-Weinberg disequilibrium and significant Inbreeding co-efficient at all sites, highlights that it is not just this general pattern that the program is detecting, and therefore is likely to be a true signature of a population decline.

Within *I. grammatica*, it is only when analysing the data 'per cluster' that differences are found, Cluster 1 (which is the ad hoc cluster, Appendix L, Figure L2, Table L1) showed no evidence of a bottleneck when compared to either M_c, whereas Cluster 2 did (again, when compared to either M_c). This fits with the hypothesis, as Cluster 2 represents individuals from site 93 shown to have significantly lower genetic diversity, though (opposite to *A. sulcicollis* with Cluster 4) a bottleneck is detected in the Cluster with a lower sample size but

not in the site with a slightly higher sample size (Table 5.4b). *Baetis rhodani* surprisingly showed evidence for more bottlenecks than any other species, though they do not tally with genetic diversity (AR; Table 5.4c), only site 97, 9 and 94 did not show evidence of a bottleneck.

It is not uncommon for these different methods to yield different results, Costa *et al.* (2013) and Dussex *et al.* (2015) also compared BOTTLENECK, MPVal and MSVAR to assess demographic history. Both found no signature of a bottleneck with the heterozygosity excess method; a signal of bottleneck in some clusters using the M-ratio model (for Dussex *et al*. only), but consistent signatures of a bottleneck when using MSVAR. Minhós *et al.* (2016) compared BOTTLENECK, MSVAR and a different third method, - the Extended Bayesian Skyline Plot. Using these they found that the heterozygosity excess tests was only able to detect bottlenecks in some of the datasets, whereas again a consistent signal of population collapse for both species was found using the two Bayesian methods.

5.5.2 Conclusions

Comparing the three methods has given insight into the demographic history of *A. sulcicollis*, *I. grammatica* and *B. rhodani*. BOTTLENECK found no signature of population decline but perhaps is not suited to study the demographic history of species characterised by heterozygosity deficiency. The full Bayesian method, however, revealed a possible historical bottleneck which affected all three species, regardless of genetic differentiation and clusters, but therefore does not explain the differences in genetic diversity seen within and between species. MPVal, however, identified signature of a recent bottleneck in the sites previously identified in Chapters 3 and 4 as being unique in their chemistry and having significantly lower genetic diversity, therefore validating our hypothesis. Within *A. sulcicollis* and *I. grammatica* it is possible that bottlenecks have occurred at these sites because of the acidity and metal concentrations discussed in Chapter 3 and 4. This suggests that though these species have undergone a bottleneck, they have persisted at these sites, therefore while not being resistant (remaining unchanged) they have shown resilience, and bounced back from a once low population size. Further ecological research, however, is needed to fully investigate this theme of resilience, as we do not currently know whether these

populations are recovering. Past disturbances may also explain why *B. rhodani* shows evidence of bottlenecks at more sites than both other species. This was unpredicted due to there being less variation in genetic diversity over the sites studied, however, this result makes biological sense as this is an acid sensitive species and has been known to be affected by acid episodes (short periods (hours to weeks) of reduced pH generated by rainstorms or snow melt) (Kowalik and Ormerod 2006; Kowalik *et al.* 2007). These bottlenecks could be due to past acidification events that did not occur at all sites but have been a major post-industrial feature of the region (Ormerod *et al.* 1989; Ormerod and Durance 2009). These signatures of bottlenecks could be responding to acidification at two geographical scales: 1) across the whole Welsh region as 50% of the total stream length was once affected (Ormerod *et al.* 1989); 2) temporal acid episodes. Unlike *A. sulcicollis* and *I. grammatica*, *B. rhodani* was not present at any sites below a pH of 6, however past acidification or acid episodes that have been recorded within the sites studied could have affected the populations. Previous studies have used this phenomenon to explain why biological recovery has not necessarily followed chemical recovery from acidification (Kowalik *et al.* 2007; Pye *et al.* 2012). In terms of abundance stream assemblages appear to recover from acidification (Ormerod and Durance 2009), however, these results may have revealed the genetic response to these past disturbances.

5.6 Further work

Using these methods to investigate past bottlenecks has given us new and interesting information into the demographic of *A. sulcicollis*, *I. grammatica* and *B. rhodani*, however, to fully investigate the concepts of resistance and resilience within these species further research is needed. The next step would ideally be to compliment the methods used here with an Approximate Bayesian Computational (ABC) framework that allows the incorporation of prior knowledge to reduce computation constraints without a significant drop in accuracy (Beaumont *et al.* 2002). This would allow us to verify the ancient and more recent bottlenecks already found and also to estimate values for key parameters of interest such as pre- and post- bottleneck effective population sizes and time and duration of the

bottleneck. Using ABC enables more complex scenarios to be investigated. Currently, we can only speculate as to the genetic recovery of these species however, it is not only bottlenecks and expansions that can be identified but also signatures of recovery, allowing focus to be on the resilience of the species.

While a preliminary attempt was made to use programs that include DIYABC (Cornuet *et al.* 2008) and ABCtoolbox (Wegmann *et al.* 2010), unfortunately a lack of computational time prevented this from being included in this chapter.

The theme of resistance and resilience is part of an over-arching aim within the DURESS project, it is a big question which may only be answered through collaboration with other researchers that are studying other trophic levels in the freshwater environment from biofilm, to fish and birds, as part of achieving this aim.

Chapter 6 - General discussion

6.1 Overview

The main theme running through this study is biodiversity. This is essential for the long-term resilience of an ecosystem, including its functions and provision of services (Oliver *et al.* 2015). The study explored how biodiversity varies across geographical distance, and how it responds to environmental stressors at the species and genetic level. The streams of upland Wales provided an ideal location for such a study. Not only does freshwater represent one of the most threatened types of ecosystem worldwide, with biodiversity declining at a greater rate than in any other ecosystem (Sala *et al.* 2000; Dudgeon *et al.* 2006; Sievert *et al.* 2016), but also as the region represents a wide range of environmental heterogeneity, differing, for example, in land-use, acidity and metal concentration.

In the context of conserving freshwater ecosystem biodiversity, this thesis attempted to fill certain knowledge gaps in the literature. Generally, that few risk assessments have taken advantage of the use of molecular markers that can provide more cost-effective and reliable indicators of demographic change than some traditional approaches (Schwartz *et al.* 2007). The study highlighted the fact that availability of genetic resources, in terms of microsatellite markers, are very limited for freshwater invertebrates, particularly within two wide and diverse orders of Insecta; Plecoptera and Ephemeroptera (Section 1.5). Freshwater invertebrate populations make excellent models for basic ecological studies because of their abundance, diversity, ease of sampling and functional importance (Strayer 2006).

The study focused on three invertebrate species: *Amphinemura sulcicollis*, *Isoperla grammatica* and *Baetis rhodani*. These species were chosen because of their abundance and widespread distribution. Not only was this useful for studying large geographical areas but it also ensured that the resources developed would be useful to other researchers in the future.

6.2 Synthesis of outcomes and results

Within Chapter 2, whether cryptic diversity was present in the three species was investigated. While *B. rhodani* had previously been shown to have cryptic diversity (Williams *et al.* 2006), *A. sulcicollis* and *I. grammatica* had never been studied for this purpose. As expected, cryptic *B. rhodani* species was confirmed within upland Wales. At site 14 (Llyn Brianne, Tywi Catchment, Figure 2.6), in particular, a cryptic species was recorded with haplotypes more closely related to *Baetis liebenauae* (a species that does not occur in Britain) than to other *B. rhodani* individuals. This could represent a previously undetected taxon. Neither *A. sulcicollis* nor *I. grammatica* showed any evidence of cryptic diversity within the study area; a diagnostic tool was, therefore, only created for *B. rhodani* to ensure that cryptic diversity did not confound results.

Through Chapter 2 resources have been made available not only for use in this study but more generally to increase the resources available within freshwater invertebrate genetics. To ensure that this happened accessibility of all raw data were paramount. To this end, all mtDNA CO1 sequences (Appendix A, Table A2), raw Next Generation Sequencing (NGS) data and all assemblies created have been made available to download from Genbank (see Macdonald *et al.* (2016a) for Genbank accession numbers); furthermore, all necessary information on methods is available in a Genomic resource note posted to the preprint sever BioRxiv (since being posted on BioRxiv (29 March 2016 – 1 July 2016), this note has had 346 abstract viewings and 76 pdf downloads and 58 reads via Researchgate).

As well as creating genetic resources, software was also developed. The program *PrimerPipeline* provides a user-friendly program with a graphical interface, previously unavailable for microsatellite maker development. Since its creation, as well as being used to develop the microsatellite primers in this study, it has been used by numerous colleagues and the website has had 282 unique users from more than 22 countries (from 21 January 2016 – 1 July 2016). To promote this software more widely, a technical note for publication in GigaScience is currently under preparation.

The development of novel microsatellite markers (Macdonald *et al.* 2016b) means that *I. grammatica*, *A. sulcicollis* and *B. rhodani* now have between 18-21 microsatellites each (including primers by Williams *et al.* (2002)). This represents a large resource as there were

previously only three other species of Ephemeroptera (with between 2 – 10 loci available), and only three other species of Plecoptera (with between 4 – 13 loci available) for which microsatellites were available (Table 1.2).

The aim of Chapter 3, the first study to utilise microsatellites to investigate three species of freshwater invertebrate simultaneously, was to investigate the fine-scale genetic differentiation, population structure and genetic diversity of the species and their response to environmental heterogeneity and environmental stressors. By doing this, inferences could be made about their dispersal ability and, therefore, stream connectivity. All species were found to have high dispersal ability, owing to their adult terrestrial life-stage; all three species showed panmixa across catchments in southern and mid-Wales. *Amphinumera sulcicollis* and *I. grammatica* did, however, show genetic isolation and reduced genetic diversity in certain sites in North Wales. This was reflected by low pH levels and higher metal concentration at these sites. *Baetis rhodani* did not show this pattern of genetic diversity, their acid sensitivity meant that they were not present at these sites.

The reduced genetic diversity seen at these sites could be a genetic response to acidity not observed before in these species. *Amphinemura sulcicollis* and *I. grammatica* were previously considered to be acid tolerant because although it is well documented that acid sensitive species are removed from acidic sites, these or related species have been shown to persist (Eriksen and Pettersen 2016; Petrin *et al.* 2007). The implication of this result is a possible explanation as to why biological recovery does not always follow chemical recovery. As a consequence of past acidification in a demographically isolated location, decreased genetic diversity could hinder recovery due to reduced resilience in the face of other disturbances. Further research is necessary to determine if this genetic response is recorded in other locations.

In Chapter 4, the genetic diversity of each species was compared to the species diversity across whole macroinvertebrate assemblages. The aim was to determine whether a species-genetic diversity correlation (SGDC) existed. When using species richness and Shannon diversity, a significant positive SGDC was only found for *A. sulcicollis*. The underlying driver of this correlation was thought to be a combination of genetic isolation and acidity; both species diversity and genetic diversity for *A. sulcicollis*, had a significant positive correlation

with pH across the various locations. Studying the genetic structure of three freshwater invertebrates, that vary in their sensitivity to certain stressors (drivers), simultaneously, has helped highlight possible causes of SGDCs. The fact that differences were observed between species shows that a SGDC can never be assumed, even between related species.

As reported in Chapter 3, reduced genetic diversity was found at acidic sites. To investigate the cause of this in Chapter 5 the hypothesis that reduced genetic diversity at these sites was caused by a past bottleneck was tested. Signatures of genetic bottlenecks were tested using three different approaches, using the software BOTTLENECK, MSVAR and MPVal. Only the M-ratio method using MPVal detected differences between sites/clusters and species, although MSVAR found an ancient, possibly post-glacial, bottleneck common to all species and sites tested. Within *A. sulcicollis* and *I. grammatica*, evidence of bottlenecks was found within the sites/clusters previously identified as having reduced genetic diversity. This validated the hypothesis that reduced genetic diversity was caused by previous bottlenecks. *Baetis rhodani* showed evidence of previous bottlenecks in almost all sites, and when analysed as one cluster. These bottlenecks could be due to past acidification events that did not occur at all sites but have been a major post-industrial feature of the region (Ormerod *et al.* 1989; Ormerod and Durance 2009). Bottleneck signatures could be responding to acidification at two geographical scales: 1) across the whole Welsh region (50% of the total stream length having been affected (Ormerod *et al.* 1989)), or 2) episodic acidification that is still apparent at many sites (Kowalik and Ormerod 2006; Kowalik *et al.* 2007).

6.3 Limitations

6.3.1 Observation study

All the results in this study come from sampling in a natural environment. It was therefore not possible to control or record all possible influences that could have affected the results. The aim was to sample across a range of heterogeneous environments so that causes for patterns could be established. Major efforts were made to reduce confounding effects by collecting as many data on each site as possible. Altitude was not shown to be a factor;

although the sites studied range from 160 - 440 m; each species was collected from a similar range (190 – 440 m, 160 – 370 m and 210 – 400 m for *A. sulcicollis*, *I. grammatica* and *B. rhodani*, respectively). The sites also represented a combination of six different land-uses (coniferous woodland, deciduous woodland, improved pasture, rough pasture, moorland and heathland). This also did not seem to influence genetic isolation or genetic diversity. Sites for all three species represented all of these land-uses (Appendix A, Table A1). Weather data were also gathered at catchment level but again, no patterns were seen.

6.3.2 Number of different sites

When comparing results within species the number of sites used varied between 10 – 13 (9 – 11 when temporal sites were not included). This is a relatively low sample size per species, although comparable to other invertebrate studies (Table 1.2, for example Alp *et al.* (2012) used 11 and 14 sites to compare two freshwater invertebrates; and Asmyhr *et al.* (2014), Watts and Thompson (2012), and Huey *et al.* (2011) used 8, 9 and 10 sites respectively to study one species). The number of sites was reduced further when comparing genetic data to chemical and species data as this information was not available at every site. The number of sites used to investigate SGDC was comparable to that reported in the literature (Table 1.3; for example, Evanno *et al.* (2009) and Odat *et al.* (2004) compared SGDC over five and ten local sites). This current study does not, however, compare to larger scale studies such as that of Taberlet *et al.* (2012) which had 249 sites spanning several countries.

The number of sites studied was restricted logistically because of time availability and budget. This was further constrained due to the need for high levels of repetition when genotyping to ensure accuracy. The number of sites per species is, however, a limiting factor in this study. For example, within *I. grammatica*, there is only one site that shows demographic isolation and significantly reduced genetic diversity. There is evidence to suggest this is due to low acidity at the site. This is also illustrated in *A. sulcicollis*, although with only one site displaying such a result it is difficult to prove it is not an outlier. If more sites were to be added, sites based on a pH gradient would probably be most suited in order to validate the hypothesis of the general genetic response to acidification.

6.3.2 Hardy-Weinberg disequilibrium

All three species show Hardy-Weinberg disequilibrium and consequently, evidence of null alleles, at a relatively high proportion of loci. While this is an issue if the cause is artificial, this has been found to be a common factor within highly polymorphic invertebrate species. All efforts were taken to: 1) ensure high levels of accuracy at the scoring and binning stages, and 2) assess the influence of these loci on the results. An investigation into the effect of null alleles on STRUCTURE results for *A. sulcicollis* was carried out in Chapter 2, and, as reported in Chapter 3, a large proportion of the analysis was carried out with and without loci displaying this characteristic. Individual *A. sulcicollis* loci were found not to influence the STRUCTURE plots unduly, and for each species the same patterns were seen using both datasets in Chapter 3.

6.4 Future directions

The immediate next step for this research is, as mentioned in Chapter 5, to try an Approximate Bayesian Computational (ABC) framework to assess the demographic resilience of each species. It would also be advisable to collaborate with other researchers working within the extensive research program funded by the Natural Environment Research Council (NERC): DURESS (Diversity of Upland Rivers for Ecosystem Service Sustainability: http://nerc-duress.org). Colleagues have been preforming genetic studies on trout, dipper and biofilm within the same locations as the macroinvertebrates reported in this study. Through collaboration and comparison of studies across various trophic levels it is hoped that key questions on ecosystem resilience can be answered and the connection between biodiversity and ecosystem service resilience unravelled (http://nerc-duress.org/work/theme-3-resilience-and-thresholds).

While this study has advanced our understanding the findings also highlight certain knowledge gaps and areas where further research is required to validate conclusions. Within *A. sulcicollis* and *I. grammatica*, a genetic response to acidity was suggested to

explain the reduced genetic diversity at acidic sites. This conclusion is only based on a limited number of sites. It is recommended that further studies into the genetic response to environmental stressors should be tested at a greater number of locations, ideally controlling for a larger range of environmental stressors. This research could be extended to see if this pattern is found in other freshwater macroinvertebrates. Further research could focus on recovery from acidification to determine whether past bottlenecks and reduced genetic diversity hinder recovery in macroinvertebrate assemblages.

6.5 Conclusions

In its entirety, this study has shown the benefit of adding a genetic component to biodiversity investigations. Work reported in Chapter 3 showed that although species are present at low acidity sites it does not necessarily follow that they are unaffected. Lower genetic diversity at these locations could reduce their resilience by reducing their evolutionary potential. Some researchers have suggested that species diversity could potentially predict genetic diversity (e.g. He *et al.* 2008) and that species diversity could be used as a surrogate for genetic diversity and *vice versa*. The content of Chapter 4 highlights that this is not always the case, and in order to investigate the genetic health of a species the genetic diversity must be assessed, not inferred.

Investigating the genetic health of ecosystems may help achieve international policy goals to reduce the decline of biodiversity. The Convention on Biological Diversity (CBD)'s Aichi Targets for 2010 to the United Nations (Pereira *et al.* 2013), for example, seek to minimise genetic erosion and safeguard genetic diversity (Aichi Target 13), while one of the goals of CBD 2020 is to "enhance ecosystem resilience" (Convention on Biological Diversity 2012). More local legislation from the Welsh Government seeks to reduce biodiversity decline in Wales to "support social, economic and ecological resilience". Undoubtedly, further research is required to untangle the connection between biodiversity, both at the species and genetic level, and resilience in freshwater and other natural environments.

References

A

Aarssen, L. W. (1983). Ecological combining ability and competitive combining ability in plants: towards a general evolutionary theory of coexistence in systems of competition. *The American Naturalist* **122:**707-731.

Abbot, P. (2001). Individual and population variation in invertebrates revealed by Inter-simple Sequence Repeats (ISSRs). *Journal of Insect Science* **1:**8.

Abdul-Muneer, P. M. (2014). Application of microsatellite markers in conservation genetics and fisheries management: recent advances in population structure analysis and conservation strategies. *Genetics Research International* **2014:**11.

Ablett, G., Hill, H. and Henry, R. J. (2006). Sequence polymorphism discovery in wheat microsatellite flanking regions using pyrophosphate sequencing. *Molecular Breeding* **17:**281-289.

Adams, R. I. and Vellend, M. (2011). Species diversity of grasses promotes genotypic diversity of clover populations in simulated communities. *Oikos* **120:**1584-1594.

Addison, J. A., Einfeldt, A. L., Kang, N. N. and Walde, S. J. (2015). Small-scale genetic structure in a stream-dwelling caddisfly in eastern Canada. *Marine and Freshwater Research* **66:**459-468.

Addison, J. A. and Hart, M. W. (2005). Spawning, copulation and inbreeding coefficients in marine invertebrates. *Biology Letters* **1:**450-453.

References

Aguilar, A. and Jones, W. J. (2009). Nuclear and mitochondrial diversification in two native California minnows: insights into taxonomic identity and regional phylogeography. *Molecular Phylogenetics and Evolution* **51:**373-381.

Albert, J. S., Carvalho, T. P., Petry, P., Holder, M. A., Maxime, E. L., Espino, J., Corahua, I., Quispe, R., Rengifo, B., Ortega, H. and Reis, R. E. (2011). Aquatic biodiversity in the Amazon: Habitat specialization and geographic isolation promote species richness. *Animals* **1:**205-241.

Alp, M., Keller, I., Westram, A. M. and Robinson, C. T. (2012). How river structure and biological traits influence gene flow: a population genetic study of two stream invertebrates with differing dispersal abilities. *Freshwater Biology* **57:**969-981.

Amore, V., Belardinelli, M., Guerra, L., Buonocore, F., Fausto, A. M., Ubero-Pascal, N. and Fochetti, R. (2009). Do all stoneflies nymphs have respiratory proteins? Further data on the presence of hemocyanin in the larval stages of plecoptera species. *Insect Molecular Biology* **18:**203-211.

Amos, W., Hoffman, J. I., Frodsham, A., Zhang, L., Best, S. and Hill, A. V. S. (2007). Automated binning of microsatellite alleles: problems and solutions. *Molecular Ecology Notes* **7:**10-14.

Anas, M. U. M., Scott, K. A. and Wissel, B. (2015). Carbon budgets of boreal lakes: State of knowledge, challenges, and implications. *Environmental Reviews* **23:**275-287.

Andrén, C. M. and Wiklund, A. K. E. (2013). Response of *Gammarus pulex* and *Baetis rhodani* to springtime acid episodes in humic brooks. *Science of The Total Environment* **463–464:**690-699.

References

Antonovics, J. (1976). The input from population genetics: “the new ecological genetics.”. *Systematic Botany* **1:**233-245.

Antonovics, J. (1978). The population genetics of species mixtures. In: Wilson, J.R. ed. *Plant Relations in Patures*. Melbourne, Austraila: CSIRO. pp. 426-449.

Antonovics, J. (1992). Toward community genetics. In: Fitz, R.S. and Simms, E. L. (ed.) *Plant resistance to herbivores and pathogens: ecology, evolution, genetics*. Chicago: University of Chicago Press.

Asmyhr, M. G., Hose, G., Graham, P. and Stow, A. J. (2014). Fine-scale genetics of subterranean syncarids. *Freshwater Biology* **59:**1-11.

Avolio, M. L. and Smith, M. D. (2013). Correlations between genetic and species diversity: Effects of resource quantity and heterogeneity. *Journal of Vegetation Science* **24:**1185-1194.

B

Bady, P., Dolédec, S., Fesl, C., Gayraud, S., Bacchi, M. and Schöll, F. (2005). Use of invertebrate traits for the biomonitoring of European large rivers: the effects of sampling effort on genus richness and functional diversity. *Freshwater Biology* **50:**159-173.

Baekken, T. (1981). Growth patterns and food habits of *Baetis rhodani*, *Capnia pygmaea* and *Diura nanseni* in a West Norwegian river. *Ecography* **4:**139-144.

Baillie, K. B., Belda-Baillie, A. C., Silvestre, V., Sison, M., Gomez, V. A., Gomez, D. E. and Monje, V. (2000). Genetic variation in *Symbiodinium* isolates from giant clams based on random-amplified-polymorphic DNA (RAPD) patterns. *Marine Biology* **136:**829-836.

References

Bandelt, H. J., Forster, P. and Rohl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* **16:**37-48.

Banks, S. C., Cary, G. J., Smith, A. L., Davies, I. D., Driscoll, D. A., Gill, A. M., Lindenmayer, D. B. and Peakall, R. (2013). How does ecological disturbance influence genetic diversity? *Trends in Ecology and Evolution* **28:**670-679.

Baselga, A., Fujisawa, T., Crampton-Platt, A., Bergsten, J., Foster, P. G., Monaghan, M. T. and Vogler, A. P. (2013). Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels. *Nature Communications* **4**. Doi:10.1038/ncomms2881

Beaumont, M., Barratt, E. M., Gottelli, D., Kitchener, A. C., Daniels, M. J., Pritchard, J. K. and Bruford, M. W. (2001). Genetic diversity and introgression in the Scottish wildcat. *Molecular Ecology* **10:**319-336.

Beaumont, M. A. (1999). Detecting population expansion and decline using microsatellites. *Genetics* **153:**2013-2029.

Beaumont, M. A., Zhang, W. Y. and Balding, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics* **162**:2025–2035.

Bell, G., Lechowicz, M. J., Appenzeller, A., Chandler, M., DeBlois, E., Jackson, L., Mackenzie, B., Preziosi, R., Schallenberg, M. and Tinker, N. (1993). The spatial structure of the physical environment. *Oecologia* **96:**114-121.

Bensch, S., Hasselquist, D., Nielsen, B. and Hansson, B. (1998) Higher fitness for philopatric than for immigrant males in a semi-isolated population of great reed warblers. *Evolution* **52**, 877–883.

References

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Sayers, E. W. (2013). GenBank. *Nucleic Acids Research* **41:**D36-42.

Benson, D. A., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J. and Wheeler, D. L. (2005). GenBank. *Nucleic Acids Research* **33:**D34-D38.

Bernard, A. M., Feldheim, K. A., Nemeth, R., Kadison, E., Blondeau, J., Semmens, B. X. and Shivji, M. S. (2016). The ups and downs of coral reef fishes: the genetic characteristics of a formerly severely overfished but currently recovering Nassau grouper fish spawning aggregation. *Coral Reefs* **35:**273-284.

Bernhardt, J. R. and Leslie, H. M. (2013). Resilience to climate chnge in coastal marine ecosystems. *Annual Review of Marine Science* **5:**371-392.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K. L., Meier, R., Winker, K., Ingram, K. K. and Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution* **22:**148-155.

Bickham, J. W., Sandhu, S., Hebert, P. D. N., Chikhi, L. and Athwal, R. (2000) Effects of chemical contaminants on genetic diversity in natural populations: implications for biomonitoring and ecotoxicology. *Mutation Research/Reviews in Mutation Research*, **463**:33-51.

Bilton, D. T. (2014). Dispersal in Dytiscidae. In: Yee, A.D. (ed.) *Ecology, Systematics, and the Natural History of Predaceous Diving Beetles (Coleoptera: Dytiscidae)*. Doi:10.1007/978-94-017-9109-0_9**:** Dordrecht: Springer Netherlands, pp. 387-407.

References

Bisconti, R., Canestrelli, D., Tenchini, R., Belfiore, C., Buffagni, A. and Nascetti, G. (2016). Cryptic diversity and multiple origins of the widespread mayfly species group *Baetis rhodani* (Ephemeroptera: Baetidae) on northwestern Mediterranean islands. *bioRxiv* Doi:10.1101/056556.

Blankenship, S. M., May, B. and Hedgecock, D. (2002). Evolution of a perfect simple sequence repeat locus in the context of its flanking sequence. *Molecular Biology and Evolution* **19:**1943-1951.

Blum, M. J., Bagley, M. J., Walters, D. M., Jackson, S. A., Daniel, F. B., Chaloud, D. J. and Cade, B. S. (2012). Genetic diversity and species diversity of stream fishes covary across a land-use gradient. *Oecologia* **168:**83-95.

Booth, R. E. and Grime, J. P. (2003). Effects of genetic impoverishment on plant community diversity. *Journal of Ecology* **91:**721-730.

Boumans, L. and Baumann, R. W. (2012). Amphinemura palmeni is a valid Holarctic stonefly species (Plecoptera: Nemouridae). *Zootaxa* **3537:**59-75.

Boutin-Ganache, I., Raposo, M., Raymond, M. and Deschepper, C. F. (2001). M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. *Biotechniques* **31:**24-26, 28.

Bowen, D. Q., Phillips, F. M., McCabe, A. M., Knutz, P. C. and Sykes, G. A. (2002). New data for the Last Glacial Maximum in Great Britain and Ireland. *Quaternary Science Reviews* **21:**89-101.

References

Bradley, D. C. and Ormerod, S. J. (2002a). Evaluating the precision of kick-sampling in upland streams for assessments of long-term change: the effects of sampling effort, habitat and rarity. *Archiv Fur Hydrobiologie* **155:**199-221.

Bradley, D. C. and Ormerod, S. J. (2002b). Long-term effects of catchment liming on invertebrates in upland streams. *Freshwater Biology* **47:**161-171.

Bradshaw, W. E. and Holzapfel, C. M. (2006). Evolutionary Response to Rapid Climate Change. *Science* **312:**1477-1478.

Briers, R. A., Cariss, H. M. and Gee, J. H. R. (2002). Dispersal of adult stoneflies (Plecoptera) from upland streams draining catchments with contrasting land-use. *Archiv Fur Hydrobiologie* **155:**627-644.

Brohede, J., Primmer, C. R., Møller, A. and Ellegren, H. (2002). Heterogeneity in the rate and pattern of germline mutation at individual microsatellite loci. *Nucleic Acids Research* **30:**1997-2003.

Brookfield, J. F. (1996). A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology* **5:**453-455.

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* **7:**434-455.

Brooks, T. M., Mittermeier, R. A., da Fonseca, G. A. B., Gerlach, J., Hoffmann, M., Lamoreux, J. F., Mittermeier, C. G., Pilgrim, J. D. and Rodrigues, A. S. L. (2006). Global biodiversity conservation priorities. *Science* **313:**58-61.

References

Brownlow, R. J., Dawson, D. A., Horsburgh, G. J., Bell, J. J. and Fish, J. D. (2008). A method for genotype validation and primer assessment in heterozygote-deficient species, as demonstrated in the prosobranch mollusc *Hydrobia ulvae*. *BMC Genetics* **9**:55.

Buchwalter, D. B., Cain, D. J., Martin, C. A., Xie, L., Luoma, S. N. and Garland, T. (2008). Aquatic insect ecophysiological traits reveal phylogenetically based differences in dissolved cadmium susceptibility. *Proceedings of the National Academy of Sciences* **105:**8321-8326.

Buchwalter, D. B. and Luoma, S. N. (2005). Differences in Dissolved Cadmium and Zinc Uptake among Stream Insects: Mechanistic Explanations. *Environmental Science & Technology* **39:**498-504.

Bulut, Z., McCormick, C. R., Gopurenko, D., Williams, R. N., Bos, D. H. and DeWoody, J. A. (2009). Microsatellite mutation rates in the eastern tiger salamander (*Ambystoma tigrinum tigrinum*) differ 10-fold across loci. *Genetica* **136:**501-504.

Burdon, J. J. (1987). Dieases and plant population biology. Cambridge, UK: Cambridge University Press.

Buss, D. F., Carlisle, D. M., Chon, T. S., Culp, J., Harding, J. S., Keizer-Vlek, H. E., Robinson, W. A., Strachan, S., Thirion, C. and Hughes, R. M. (2015). Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-scale programs. *Environmental monitoring and assessment* **187:**4132.

C

Camargo, J. A., Alonso, A. and Salamanca, A. (2005). Nitrate toxicity to aquatic animals: a review with new data for freshwater invertebrates. *Chemosphere* **58:**1255-1267.

References

Cao, L., Fu, X. and Wu, K. (2015). Development of 10 microsatellite markers from Pantala flavescens and their applicability in studying genetics diversity. *Molecular Biology Reports* **42:**1275-1279.

Cardoni, S., Tenchini, R., Ficulle, I., Piredda, R., Simeone, M. C. and Belfiore, C. (2015). DNA barcode assessment of Mediterranean mayflies (Ephemeroptera), benchmark data for a regional reference library for rapid biomonitoring of freshwaters. *Biochemical Systematics and Ecology* **62:**36-50.

Carew, M., Gagliardi, B. and Hoffmann, A. A. (2013). Mitochondrial DNA suggests a single maternal origin for the widespread triploid parthenogenetic pest species, *Paratanytarsus grimmii*, but microsatellite variation shows local endemism. *Insect Science* **20:**345-357.

Carew, M. E., Pettigrove, V. and Hoffmann, A. A. (2003). Identifying chironomids (Diptera: Chironomidae) for biological monitoring with PCR–RFLP. *Bulletin of Entomological Research* **93:**483-490.

Carreño-Rocabado, G., Peña-Claros, M., Bongers, F., Alarcón, A., Licona, J. C. and Poorter, L. (2012). Effects of disturbance intensity on species and functional diversity in a tropical forest. *Journal of Ecology* **100:**1453-1463.

Cervini, M., Henrique-Silva, F., Mortari, N. and Matheucci Jr, E. (2006). Genetic variability of 10 microsatellite markers in the characterization of Brazilian Nellore cattle (*Bos indicus*). *Genetics and Molecular Biology* **29:**486-490.

Chakraborty, R., De Andrade, M., Daiger, S. P. and Budowle, B. (1992). Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications. *Annals of Human Genetics* **56:**45-57.

References

Chapuis, M.-P. and Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24:**621-631.

Chapuis, M. P., Loiseau, A., Michalakis, Y., Lecoq, M. and Estoup, A. (2005). Characterization and PCR multiplexing of polymorphic microsatellite loci for the locust Locusta migratoria. *Molecular Ecology Notes* **5:**554-557.

Chavarriaga-Aguirre, P., Maya, M. M., Tohme, J., Duque, M. C., Iglesias, C., Bonierbale, M. W., Kresovich, S. and Kochert, G. (1999). Using microsatellites, isozymes and AFLPs to evaluate genetic diversity and redundancy in the cassava core collection and to assess the usefulness of DNA-based markers to maintain germplasm collections. *Molecular Breeding* **5:**263-273.

Chessman, B. C. (2015). Relationships between lotic macroinvertebrate traits and responses to extreme drought. *Freshwater Biology* **60:**50-63.

Chester, E. T., Miller, A. D., Valenzuela, I., Wickson, S. J. and Robson, B. J. (2015). Drought survival strategies, dispersal potential and persistence of invertebrate species in an intermittent stream landscape. *Freshwater Biology* **60:**2066-2083.

Chevene, F., DolÉAdec, S. and Chessel, D. (1994). A fuzzy coding approach for the analysis of long-term ecological data. *Freshwater Biology* **31:**295-309.

Chikhi, L., Sousa, V. C., Luisi, P., Goossens, B. and Beaumont, M. A. (2010). The confounding effects of population structure, genetic diversity and the sampling scheme on the detection and quantification of population size changes. *Genetics* **186:**983-995.

References

Cleary, D. F. R., Fauvelot, C., Genner, M. J., Menken, S. B. J. and Mooers, A. Ø. (2006). Parallel responses of species and genetic diversity to El Niño Southern Oscillation-induced environmental destruction. *Ecology Letters* **9:**304-310.

Clements, W. H., Carlisle, D. M., Lazorchak, J. M. and Johnson, P. C. (2000). Heavy metals structure benthic communities in Colorado mountian streams. *Ecological Applications* **10:**626-638.

Collier, K. J. and Smith, B. J. (1995). Sticky trapping of adult mayflies, stoneflies and caddisflies along three contrasting streams near Hamilton, New Zealand. *New Zealand Natural Sciences* **22:**1-9.

Compin, A. and Céréghino, R. (2003). Sensitivity of aquatic insect species richness to disturbance in the Adour–Garonne stream system (France). *Ecological Indicators* **3:**135-142.

Consortium, T. I. H. (2007). A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449:**851-861.

Convention on Biological Diversity. (2012). Aichi Biodiversity Targets [Online]. Available at: https://www.cbd.int/sp/targets/ [Accessed 28.06.16].

Cornetti, L., Benazzo, A., Hoban, S., Vernesi, C. and Bertorelle, G. (2016). Ancient, but not recent, population declines have had a genetic impact on alpine yellow-bellied toad populations, suggesting potential for complete recovery. *Conservation Genetics* **17:**727-743.

Cornuet, J.-M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J.-M., Balding, D. J., Guillemaud, T. and Estoup, A. (2008). Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24:**2713-2719.

References

Cornuet, J. M. and Luikart, G. (1996). Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data. *Genetics* **144**:2001-14.

Corrêa, T. C., Del Lama, S. N., De Souza, J. R. and Miño, C. I. (2015). Genetic structuring among populations of the great egret, *Ardea alba egretta*, in major Brazilian wetlands. *Aquatic Conservation: Marine and Freshwater Ecosystems* **26***:* 333–349.

Costa, M., Fernandes, C., Birks, J. D., Kitchener, A. C., Santos-Reis, M. and Bruford, M. W. (2013). The genetic legacy of the 19th-century decline of the British polecat: evidence for extensive introgression from feral ferrets. *Molecular Ecology* **22:**5130-5147.

Costanza, R., de Groot, R., Sutton, P., van der Ploeg, S., Anderson, S. J., Kubiszewski, I., Farber, S. and Turner, R. K. (2014). Changes in the global value of ecosystem services. *Global Environmental Change* **26:**152-158.

Covich, A. P., Palmer, M. A. and Crowl, T. A. (1999). The role of benthic invertebrate species in freshwater ecosystems: zoobenthic species influence energy flows and nutrient cycling. *BioScience* **49:**119-127.

Crawford, K. M. and Rudgers, J. A. (2012). Plant species diversity and genetic diversity within a dominant species interactively affect plant community biomass. *Journal of Ecology* **100:**1512-1521.

Crispo, E. and Chapman, L. J. (2008) Population genetic structure across dissolved oxygen regimes in an African cichlid fish. *Molecular Ecology*, **17**: 2134–2148.

Crnokrak, P. and Barrett, S. C. H. (2002). Perspective: purging the genetic load: a review of the experimental evidence. *Evolution* **56:**2347-2358.

References

Crook, D. A., Lowe, W. H., Allendorf, F. W., Erős, T., Finn, D. S., Gillanders, B. M., Hadwen, W. L., Harrod, C., Hermoso, V., Jennings, S., Kilada, R. W., Nagelkerken, I., Hansen, M. M., Page, T. J., Riginos, C., Fry, B. and Hughes, J. M. (2015). Human effects on ecological connectivity in aquatic ecosystems: integrating scientific approaches to support management and mitigation. *Science of The Total Environment* **534:**52-64.

Csergö, A.-M., Schönswetter, P., Mara, G., Deák, T., Boşcaiu, N. and Höhn, M. (2009). Genetic structure of peripheral, island-like populations: a case study of Saponaria bellidifolia Sm. (Caryophyllaceae) from the Southeastern Carpathians. *Plant Systematics and Evolution* **278:**33-41.

Csergo, A. M., Hufnagel, L. and Höhn, M. (2014). Positive relationship between genetic- and species diversity on limestone outcrops in the Carpathian Mountains. *Ecological Complexity* **20**:233–239.

D

da Silva-Méndez, G., Lorenzo-Carballa, M. O., Cordero-Rivera, A. and Watts, P. C. (2013). Microsatellite loci for two threatened dragonfly (Odonata: Anisoptera) species: *Oxygastra curtisii* (Dale, 1834) and *Macromia splendens* (Pictet, 1843). *Conservation Genetics Resources* **5:**1171-1174.

Dakin, E. E. and Avise, J. C. (2004). Microsatellite null alleles in parentage analysis. *Heredity* **93:**504-509.

Daniels, S. R., Stewart, B. A. and Cook, P. A. (2002). Congruent patterns of genetic variation in a burrowing freshwater crab revealed by allozymes and mt DNA sequence analysis. *Hydrobiologia* **468:**171-179.

References

Dawson, D. A. and Wilcock, H. R. (2002). Isolation of polymorphic microsatellite loci in the net-spinning caddisfly, *Polycentropus flavomaculatus* (Polycentropodidae). *Molecular Ecology Notes* **2:**514-517.

De Castro-Català, N., Muñoz, I., Armendáriz, L., Campos, B., Barceló, D., López-Doval, J., Pérez, S., Petrovic, M., Picó, Y. and Riera, J. L. (2015). Invertebrate community responses to emerging water pollutants in Iberian river basins. *Science of The Total Environment* **503–504:**142-150.

de Figueroa, J. M. T., Bo, T., López-Rodríguez, M. J. and Fenoglio, S. (2009). Life cycle of three stonefly species (Plecoptera) from an Apenninic stream (Italy) with the description of the nymph of Nemoura hesperiae. *Annales de la Société entomologique de France (N.S.)* **45:**339-343.

de Groot, R., Brander, L., van der Ploeg, S., Costanza, R., Bernard, F., Braat, L., Christie, M., Crossman, N., Ghermandi, A., Hein, L., Hussain, S., Kumar, P., McVittie, A., Portela, R., Rodriguez, L. C., ten Brink, P. and van Beukering, P. (2012). Global estimates of the value of ecosystems and their services in monetary units. *Ecosystem Services* **1:**50-61.

De Laender, F., Melian, C. J., Bindler, R., Van den Brink, P. J., Daam, M., Roussel, H., Juselius, J., Verschuren, D. and Janssen, C. R. (2014). The contribution of intra- and interspecific tolerance variability to biodiversity changes along toxicity gradients. *Ecology Letters* **17:**72-81.

de Mendoza, G., Ventura, M. and Catalan, J. (2015). Environmental factors prevail over dispersal constraints in determining the distribution and assembly of Trichoptera species in mountain lakes. *Ecology and Evolution* **5:**2518-2532.

Delettre, Y. R. and Morvan, N. (2000). Dispersal of adult aquatic Chironomidae (Diptera) in agricultural landscapes. *Freshwater Biology* **44:**399-411.

References

Derry, A. M., Arnott, S. E., Shead, J. A., Hebert, P. D. N. and Boag, P. T. (2009). Ecological linkages between community and genetic diversity in zooplankton among boreal shield lakes. *Ecology* **90:**2275-2286.

Dewoody, J., Nason, J. D. and Hipkins, V. D. (2006). Mitigating scoring errors in microsatellite data from wild populations. *Molecular Ecology Notes* **6:**951-957.

Dharmarajan, G., Beatty, W. S. and Rhodes, O. E. (2013). Heterozygote deficiencies caused by a Wahlund effect: Dispelling unfounded expectations. *The Journal of Wildlife Management* **77:**226-234.

Dieringer, D. and Schlötterer, C. (2003). microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* **3:**167-169.

Dohet, A., Hlúbiková, D., Wetzel, C. E., L'Hoste, L., Iffly, J. F., Hoffmann, L. and Ector, L. (2015). Influence of thermal regime and land use on benthic invertebrate communities inhabiting headwater streams exposed to contrasted shading. *Science of The Total Environment* **505:**1112-1126.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A.-H., Soto, D., Stiassny, M. L. J. and Sullivan, C. A. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews* **81:**163-182.

Durance, I. and Ormerod, S. J. (2007). Climate change effects on upland stream macroinvertebrates over a 25-year period. *Global Change Biology* **13:**942-957.

Dussex, N., Rawlence, N. J. and Robertson, B. C. (2015). Ancient and contemporary DNA reveal a pre-human decline but no population bottleneck associated with recent human persecution in the Kea (*Nestor notabilis*). *PLoS ONE* **10:**e0118522.

E

Earl, D. and vonHoldt, B. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4:**359-361.

Elbrecht, V., Feld, C. K., Gies, M., Hering, D., Sondermann, M., Tollrian, R. and Leese, F. (2014). Genetic diversity and dispersal potential of the stonefly *Dinocras cephalotes* in a central European low mountain range. *Freshwater Science* **33:**181-192.

Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics* **5:**435-445.

Ellingson, R. A. and Krug, P. J. (2016). Reduced genetic diversity and increased reproductive isolation follow population-level loss of larval dispersal in a marine gastropod. *Evolution* **70:**18-37.

Elliott, J. M. (1991). The effect of temperature on egg hatching for three populations of Isoperla grammatica and one population of *Isogenus nubecula* (Plecoptera: Perlodidae). *Entomologist's Gazette* **42:**61-65.

Elliott, J. M. (2002). Time spent in the drift by downstream-dispersing invertebrates in a Lake District stream. *Freshwater Biology* **47:**97-106.

References

Elliott, J. M. (2003a). A comparative study of the functional response of four species of carnivorous stoneflies. *Freshwater Biology* **48:**191-202.

Elliott, J. M. (2003b). Interspecific interference and the functional response of four species of carnivorous stoneflies. *Freshwater Biology* **48:**1527-1539.

Elliott, J. M. and Humpesch, U. H. (1983). A key to the adults of the British Ephemeroptera with notes on their ecology. Cumbria: Scientific Publications of the Freshwater Biological Association No. 47.

Elliott, J. M., Humpesch, U. H. and Macan, T. T. (1988). Larvae of the British Ephemeroptera: a key with ecological notes. Cumbria, England: Freshwater Biological Association Scientific Publication No.49.

Elton, C., S. (1958). The Ecology of Invasion by plants and animals. Chicago, IL: University of Chicago Press.

Endo, Y., Nash, M., Hoffmann, A. A., Slatyer, R. and Miller, A. D. (2015). Comparative phylogeography of alpine invertebrates indicates deep lineage diversification and historical refugia in the Australian Alps. *Journal of Biogeography* **42:**89-102.

Eriksen, T. E. and Pettersen, E. (2016). Fixed dose aluminum sulfate treatment to eradicate the salmon parasite *Gyrodactylus salaris*: effects on benthic macroinvertebrates. *Journal of Freshwater Ecology* Doi:10.1080/02705060.2015.1136849**:**1-13.

Etienne, R. S., Alonso, D. and McKane, A. J. (2007). The zero-sum assumption in neutral biodiversity theory. *Journal of Theoretical Biology* **248:**522-536.

European Commission. (2000). Directive 2000/60/EC of the European parliament and of the council of 23 october 2000 establishing a framework for community action in the field of water policy. *Official Journal of the European Union* **327:**1-72.

Éva Kisdi. (2002). Dispersal: risk spreading versus local adaptation. *The American Naturalist* **159:**579-596.

Evanno, G., Castella, E., Antoine, C., Paillat, G. and Goudet, J. (2009). Parallel changes in genetic diversity and species diversity following a natural disturbance. *Molecular Ecology* **18:**1137-1144.

Evanno, G., Regnaut, S. and Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology* **14:**2611-2620.

Excoffier, L. and Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resoures* **10:**564-567.

Excoffier, L., Smouse, P. E. and Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131:**479-491.

F

Fady, B. and Conord, C. (2010). Macroecological patterns of species and genetic diversity in vascular plants of the Mediterranean basin. *Diversity and Distributions* **16:**53-64.

References

Falush, D., Stephens, M. and Pritchard, J. K. (2003). Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* **164:**1567–1587.

Feeley, H. B. and Kelly-Quinn, M. (2015). The nymphal diet of the stonefly *Protonemura meyeri* (Pictet) (Plecoptera: Nemouridae) in four episodically acidic headwater streams in Ireland. *Irish Naturalists' Journal* **34 (2):**104-109.

Feeley, H. B., Kerrigan, C., Fanning, P., Hannigan, E. and Kelly-Quinn, M. (2011). Longitudinal extent of acidification effects of plantation forest on benthic macroinvertebrate communities in soft water streams: evidence for localised impact and temporal ecological recovery. *Hydrobiologia* **671:**217-226.

Feld, C. K., Birk, S., Eme, D., Gerisch, M., Hering, D., Kernan, M., Maileht, K., Mischke, U., Ott, I., Pletterbauer, F., Poikane, S., Salgado, J., Sayer, C. D., Van Wichelen, J. and Malard, F. (2016). Disentangling the effects of land use and geo-climatic factors on diversity in European freshwater ecosystems. *Ecological Indicators* **60:**71-83.

Fernandez-Silva, I., Whitney, J., Wainwright, B., Andrews, K. R., Ylitalo-Ward, H., Bowen, B. W., Toonen, R. J., Goetze, E. and Karl, S. A. (2013). Microsatellites for Next-Generation Ecologists: A Post-Sequencing Bioinformatics Pipeline. *PLoS ONE* **8:**e55990.

Feulner, P. G. D., Bielfeldt, W., Zachos, F. E., Bradvarovic, J., Eckert, I. and Hartl, G. B. (2004). Mitochondrial DNA and microsatellite analyses of the genetic status of the presumed subspecies *Cervus elaphus montanus* (Carpathian red deer). *Heredity* **93:**299-306.

Fialkowski, W., Klonowska-Olejnik, M., Smith, B. D. and Rainbow, P. S. (2003). Mayfly larvae (*Baetis rhodani* and *B. vernus*) as biomonitors of trace metal pollution in streams of a catchment draining a zinc and lead mining area of Upper Silesia, Poland. *Environmental Pollution* **121:**253-267.

Fjellheim, A. and Raddum, G. G. (2008). Growth and voltinism in the aquatic insects of a regulated river subject to groundwater inflows. *River Research and Applications* **24:**710-719.

Flook, P. K. and Rowell, C. H. (1997). The effectiveness of mitochondrial rRNA gene sequences for the reconstruction of the phylogeny of an insect order (Orthoptera). *Molecular Phylogenetics and Evolution* **8:**177-192.

Flores-Rentería, L. and Krohn, A. (2013). Scoring Microsatellite Loci. In: Kantartzi, K.S. (ed.) *Microsatellites: Methods and Protocols*. Doi:10.1007/978-1-62703-389-3_21**:** Totowa, NJ: Humana Press, pp. 319-336.

Fochetti, R. and Tierno de Figueroa, J. M. (2008). Plecoptera. Fauna d'Italia. Edizioni Calderini de Il Sole 24 ore, Milano.

Folmer, O., Black, M., Hoeh, W., Lutz, R. and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology* **3:**294-299.

Frankham, R., Ballou, J. and Briscoe, D. (2009). An introduction to conservation genetics. 2nd edition ed. Cambridge, UK: Cambridge University Press, p. 618.

Frankham, R., Ballou, J. D. and Briscoe, D. A. (2002). Introduction to conservation genetics. Cambridge: Cambridge University Press.

Frankham, R., Lees, K., Montgomery, M. E., England, P. R., Lowe, E. H. and Briscoe, D. A. (1999). Do population size bottlenecks reduce evolutionary potential? *Animal Conservation* **2:**255-260.

References

Freeland, J. R., Noble, L. R. and Okamura, B. (2000a). Genetic consequences of the metapopulation biology of a facultatively sexual freshwater invertebrate. *Journal of Evolutionary Biology* **13:**383-395.

Freeland, J. R., Noble, L. R. and Okamura, B. (2000b). Genetic diversity of North American populations of *Cristatella mucedo*, inferred from microsatellite and mitochondrial DNA. *Molecular Ecology* **9:**1375-1389.

Freeland, J. R., Rimmer, V. K. and Okamura, B. (2001). Genetic changes within freshwater bryozoan populations suggest temporal gene flow from statoblast banks. *Limnology and Oceanography* **46:**1121-1129.

Frey, D., Arrigo, N., Granereau, G., Sarr, A., Felber, F. and Kozlowski, G. (2016). Parallel declines in species and genetic diversity driven by anthropogenic disturbance: a multispecies approach in a French Atlantic dune system. *Evolutionary Applications* **9:**479-488.

Fridley, J. D. and Grime, J. P. (2009). Community and ecosystem effects of intraspecific genetic diversity in grassland microcosms of varying species diversity. *Ecology* **91:**2272-2283.

Froufe, E., Sobral, C., Teixeira, A., Lopes, A., Sousa, R., Varandas, S. and Lopes-Lima, M. (2013). Development and multiplexing of microsatellite loci for the near threatened freshwater mussel Potomida littoralis (Cuvier, 1798) using 454 sequencing. *Aquatic Conservation: Marine and Freshwater Ecosystems* **23:**619-623.

Funk, W. C., Murphy, M. A., Hoke, K. L., Muths, E., Amburgey, S. M., Lemmon, E. M. and Lemmon, A. R. (2016). Elevational speciation in action? Restricted gene flow associated with adaptive divergence across an altitudinal gradient. *Journal of Evolutionary Biology* **29:**241-252.

G

Garant, D., Forde, S. E. and Hendry, A. P. (2007), The multifarious effects of dispersal and gene flow on contemporary adaptation. *Functional Ecology*, **21**: 434–443.

Garbuz, D. G., Zatsepina, O. G., Przhiboro, A. A., Yushenova, I., Guzhova, I. V. and Evgen'Ev, M. B. (2008). Larvae of related Diptera species from thermally contrasting habitats exhibit continuous up-regulation of heat shock proteins and high thermotolerance. *Molecular Ecology* **17:**4763-4777.

Garcia, A. A. F., Benchimol, L. L., Barbosa, A. M. M., Geraldi, I. O., Souza, C. L. J. and de Souza, A. P. (2004). Comparison of RAPD, RFLP, AFLP and SSR markers for diversity studies in tropical maize inbred lines. *Genetics and Molecular Biology* **27:**579-588.

Garza, J. C. and Williamson, E. G. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular Ecology* **10:**305-318.

Geismar, J., Haase, P., Nowak, C., Sauer, J. and Pauls, S. U. (2015). Local population genetic structure of the montane caddisfly Drusus discolor is driven by overland dispersal and spatial scaling. *Freshwater Biology* **60:**209-221.

Geismar, J. and Nowak, C. (2013). Isolation and characterisation of new microsatellite markers for the stonefly Brachyptera braueri comparing a traditional approach with high throughput 454 sequencing. *Conservation Genetics Resources* **5:**413-416.

Gelman, A. and Rubin, D. B. (1992). Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* **7**:457-472. Doi:10.1214/ss/1177011136**:**457-472.

References

Gibbs, H. L., Gibbs, K. E., Siebenmann, M. and and Collins, L. (1998). Genetic differentiation among populations of the rare mayfly *Siphlonisca aerodromia* Needham. *Journal of the North American Benthological Society* **17:**464-474.

Girod, C., Vitalis, R., Leblois, R. and Fréville, H. (2011). Inferring population decline and expansion from microsatellite data: a simulation-based evaluation of the Msvar method. *Genetics* **188:**165–179.

Goffová, K., Čiamporová-Zat'ovičová, Z. and Čiampor Jr, F. (2015). Twenty-one new microsatellite markers for the ecologically important midge *Heterotrissocladius marcidus*, and their use in studies of alpine lakes. *Biochemical Systematics and Ecology* **61:**329-335.

Gomez-Uchida, D., Weetman, D., Hauser, L., Galleguillos, R. and Retamal, M. (2003). Allozyme and AFLP analyses of genetic population structure in the hairy Edible Crab from the Chilean coast. *Journal of Crustacean Biology* **23:**486-494.

Gotelli, N. J. and Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* **4:**379-391.

Graf, W., Grasser, U. and Weinzierl, A. (2002). Plecoptera.- Part III, 17pp., in Moog, O. (Ed.) (2002). *Fauna Aquatica Austriaca*. Austrian Federal Ministry of Agriculture Forestry Environment and Water Management, Wasserwirtschaftskataster Vienna.

Graignic, N., Tremblay, F. and Bergeron, Y. (2016). Genetic consequences of selection cutting on sugar maple (*Acer saccharum* Marshall). *Evolutionary Applications* **9:**777-790.

Griffith, M. B., Barrows, E. M. and Perry, S. A. (1998). Lateral dispersal of adult aquatic insects (Plecoptera, Trichoptera) following emergence from headwater streams in forested Appalachian catchments. *Annals of the Entomological Society of America* **91:**195-201.

Guarnizo, C. E., Paz, A., Muñoz-Ortiz, A., Flechas, S. V., Méndez-Narváez, J. and Crawford, A. J. (2015). DNA barcoding survey of anurans across the Eastern Cordillera of Colombia and the impact of the andes on cryptic diversity. *PLoS ONE* **10**:e0127312.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., LÉGer, P., Lepais, O., Lepoittevin, C., Malausa, T., Revardel, E., Salin, F. and Petit, R. J. (2011). Current trends in microsatellite genotyping. *Molecular Ecology Resources* **11:**591-611.

Gutiérrez, D. and Menéndez, R. (1997), Patterns in the distribution, abundance and body size of carabid beetles (Coleoptera: Caraboidea) in relation to dispersal ability. Journal of Biogeography, 24: 903–914.

Gutiérrez-Cánovas, C., Millán, A., Velasco, J., Vaughan, I. P. and Ormerod, S. J. (2013). Contrasting effects of natural and anthropogenic stressors on beta diversity in river organisms. *Global Ecology and Biogeography* **22:**796-805.

Gutierrez-Canovas, C., Sanchez-Fernandez, D., Velasco, J., Millan, A. and Bonada, N. (2015). Similarity in the difference: changes in community functional features along natural and anthropogenic stress gradients. *Ecology* **96:**2458-2466.

H

Haasl, R. J. and Payseur, B. A. (2011). Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity (Edinb)* **106:**158-171.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41:**95-98.

Harper, G. L., King, R. A., Dodd, C. S., Harwood, J. D., Glen, D. M., Bruford, M. W. and Symondson, W. O. C. (2005). Rapid screening of invertebrate predators for multiple prey DNA targets. *Molecular Ecology* **14:**819-827.

Harper, J. L. (1977). The population biology of plants. New York, NY: Academic Press.

He, T. and Lamont, B. B. (2010). Species versus genotypic diversity of a nitrogen-fixing plant functional group in a metacommunity. *Population Ecology* **52:**337-345.

He, T., Lamont, B. B., Krauss, S. L., Enright, N. J. and Miller, B. P. (2008). Covariation between intraspecific genetic diversity and species diversity within a plant functional group. *Journal of Ecology* **96:**956-961.

Hebert, P. D. N., Ratnasingham, S. and deWaard, J. R. (2003). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B* **270:**S96–S99.

Hedrick, P. W. (1999). Perspective: Highly Variable Loci and Their Interpretation in Evolution and Conservation. *Evolution* **53:**313-318.

Heino, J., Virkkala, R. and Toivonen, H. (2009). Climate change and freshwater biodiversity: detected patterns, future trends and adaptations in northern regions. *Biological reviews of the Cambridge Philosophical Society* **84:**39-54.

Hendry, A.P. (2004) Selection against migrants contributes to the rapid evolution of ecologically dependent reproductive isolation. *Evolutionary Ecology Research* **6**:1219–1236.

References

Higgins, J. K., MacLean, H. J., Buckley, L. B. and Kingsolver, J. G. (2014). Geographic differences and microevolutionary changes in thermal sensitivity of butterfly larvae in response to climate. *Functional Ecology* **28:**982-989.

Hoban, S., Arntzen, J. A., Bruford, M. W., Godoy, J. A., Rus Hoelzel, A., Segelbacher, G., Vilà, C. and Bertorelle, G. (2014). Comparative evaluation of potential indicators and temporal sampling protocols for monitoring genetic erosion. *Evolutionary Applications* **7:**984-998.

Hoban, S., Hauffe, H., Pérez-Espona, S., Arntzen, J., Bertorelle, G., Bryja, J., Frith, K., Gaggiotti, O., Galbusera, P., Godoy, J., Hoelzel, A. R., Nichols, R., Primmer, C., Russo, I.-R., Segelbacher, G., Siegismund, H., Sihvonen, M., Vernesi, C., Vilà, C. and Bruford, M. (2013). Bringing genetic diversity to the forefront of conservation policy and management. *Conservation Genetics Resources* **5**:593-598.

Hodgson, D., McDonald, J. L. and Hosken, D. J. (2015). What do you mean,'resilient'? *Trends in Ecology & Evolution* **30:**503-506.

Hoffmann, A. A. and Willi, Y. (2008).Detecting genetic responses to environmental change. *Nature Reviews Genetics* **9**:421-432.)

Holland, M. M. and Parson, W. (2011). GeneMarker® HID: A reliable software tool for the analysis of forensic STR data. *Journal of Forensic Sciences* **56:**29-35.

Hooper, D. U., Chapin, F. S., Ewel, J. J., Hector, A., Inchausti, P., Lavorel, S., Lawton, J. H., Lodge, D. M., Loreau, M., Naeem, S., Schmid, B., Setälä, H., Symstad, A. J., Vandermeer, J. and Wardle, D. A. (2005). Effects of biodiversity on ecosystem functioning: a consenus of current knowledge. *Ecological Monographs* **75:**3-35.

References

Hosius, B., Leinemann, L., Bergmann, F., Maurer, W., Tabel, U. and Eder, W. (2001). Species diversity and genetic diversity - Is there any relationship? *Allgemeine Forst- und Jagdzeitung* **172:**87-91.

Hu, Y., Guo, Y. U., Qi, D., Zhan, X., Wu, H. U. A., Bruford, M. W. and Wei, F. (2011). Genetic structuring and recent demographic history of red pandas (*Ailurus fulgens*) inferred from microsatellite and mitochondrial DNA. *Molecular Ecology* **20:**2662-2675.

Huang, S. (2016). New thoughts on an old riddle: What determines genetic diversity within and between species? *Genomics* **108:**3-10.

Hubbell, S. P. (2001). The Unified Neutral Theory of Biodiversity and Biogeography. Princeton, NJ: Princeton University Press.

Huey, J. A., Schmidt, D. J., Balcombe, S. R., Marshall, J. C. and Hughes, J. M. (2011). High gene flow and metapopulation dynamics detected for three species in a dryland river system. *Freshwater Biology* **56:**2378-2390.

Hughes, A. R. and Stachowicz, J. J. (2004). Genetic diversity enhances the resistance of a seagrass ecosystem to disturbance. *Proceedings of the National Academy of Sciences of the United States of America* **101:**8998-9002.

Hutchinson, G. (1993). A Treatise on Limnology. Vol. 4. The Zoobenthos. New York: John Wiley & Sons.

Hynes, H. B. N. (1941). The taxonomy and ecology of the nymphs of Bristish Plecoptera with notes on the adults and eggs. *Transactions of the Royal Entomological Society of London* **91:**459-557.

Hynes, H. B. N. (1961). The invertebrate fauna of a Welsh mountain stream. *Archive für Hydrobiologie* **57:**344-388.

Hynes, H. B. N. (1977). A key to the adults and nymphs of British soneflies (Plecoptera) with notes on their ecology and distribution. Third edition. Scientific Publication No. 17. ed. Far Sawrey, Ambleside: Freshwater Biological Association.

I

Imholt, C., Gibbins, C. N., Malcolm, I. A., Langan, S. and Soulsby, C. (2010). Influence of riparian cover on stream temperatures and the growth of the mayfly Baetis rhodani in an upland stream. *Aquatic Ecology* **44:**669-678.

J

Jackson, J. K., McElravy, E. R. I. C. P. and Resh, V. I. n. H. (1999). Long-term movements of self-marked caddisfly larvae (Trichoptera: Sericostomatidae) in a California coastal mountain stream. *Freshwater Biology* **42:**525-536.

Jackson, J. K. and Resh, V. H. (1989). Activities and ecological role of adult aquatic insects in the riparian zone of streams. Berkeley, CA, USA: USDA Forest Service.

Jackson, M. C., Loewen, C. J., Vinebrooke, R. D. and Chimimba, C. T. (2016). Net effects of multiple stressors in freshwater ecosystems: a meta-analysis. *Global Change Biology* **22**:180-189.

Jakobsson, M. and Rosenberg, N. A. (2007). CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* **23:**1801-1806.

Jaramillo, F. and Destouni, G. (2015). Local flow regulation and irrigation raise global human water consumption and footprint. *Science* **350:**1248-1251.

Jardine, T. D., Kidd, K. A. and Rasmussen, J. B. (2012). Aquatic and terrestrial organic matter in the diet of stream consumers: Implications for mercury bioaccumulation. *Ecological Applications* **22:**843-855.

Johannesson, K. and AndrÉ, C. (2006). Invited reveiw: Life on the margin: genetic isolation and diversity loss in a peripheral marine ecosystem, the Baltic Sea. *Molecular Ecology* **15:**2013-2029.

Johnson, G. B. (1973). Relationship of enzyme polymorphism to species diversity. *Nature* **242:**193-194.

Johnson, M. S. and Black, R. (1984). The Wahlund effect and the geographical scale of variation in the intertidal limpet Siphonaria sp. *Marine Biology* **79:**295-302.

Jones, J. W., Neves, R. J. and Hallerman, E. M. (2015). Historical demography of freshwater mussels (Bivalvia: Unionidae): genetic evidence for population expansion and contraction during the late Pleistocene and Holocene. *Biological Journal of the Linnean Society* **114:**376-397.

Jones, O. R. and Wang, J. (2010). COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular Ecology Resources* **10:**551-555.

Junge, C., Museth, J., Hindar, K., Kraabøl, M. and Vøllestad, L. A. (2014). Assessing the consequences of habitat fragmentation for two migratory salmonid fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems* **24:**297-311.

K

Kadmon, R. and Allouche, O. (2007). Integrating the effects of area, isolation, and habitat heterogeneity on species diversity: A unification of island biogeography and niche theory. *American Naturalist* **170:**443-454.

Kalia, R. K., Rai, M. K., Kalia, S., Singh, R. and Dhawan, A. K. (2011). Microsatellite markers: an overview of the recent progress in plants. *Euphytica* **177:**309-334.

Kalinowski, S. T., Taper, M. L. and Marshall, T. C. (2007). Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16:**1099-1106.

Karandikar, R. L. (2006). On the Markov Chain Monte Carlo (MCMC) method. *Sadhana* **31:**81-104.

Karasawa, M. M. G., Vencovsky, R., Silva, C. M., Cardim, D. C., Bressan, E. A., Oliveira, G. C. X. and Veasey, E. A. (2012). Comparison of microsatellites and isozymes in genetic diversity studies of *Oryza glumaepatula* (Poaceae) populations. *Revista de Biologia Tropical* **60:**1463-1478.

Karlin, A. A., Guttman, S. I. and Rathbun, S. L. (1984). Spatial autocorrelation analysis of heterozygosity and geographic distribution in populations of *Desmognathus fuscus* (Amphibia, Plethodontidae). *Copeia* **2:**343-356.

Kefford, B. J., Papas, P. J. and Nugegoda, D. (2003). Relative salinity tolerance of macroinvertebrates from the Barwon River, Victoria, Australia. *Marine and Freshwater Research* **54:**755-765.

Keller, D. and Holderegger, R. (2013). Damselflies use different movement strategies for short- and long-distance dispersal. *Insect Conservation and Diversity* **6:**590-597.

Keller, D., Van Strien, M. J. and Holderegger, R. (2012). Do landscape barriers affect functional connectivity of populations of an endangered damselfly? *Freshwater Biology* **57:**1373-1384.

Keller, L. F. and Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution* **17:**230-241.

Knight, R. (2012). Biodiversity loss: How accurate are the numbers? BBC News. [Online]. Available at: http://www.bbc.co.uk/news/magazine-17826898 [Accessed: May 2013].

Knops, J. M. H., Tilman, D., Haddad, N. M., Naeem, S., Mitchell, C. E., Haarstad, J., Ritchie, M. E., Howe, K. M., Reich, P. B., Siemann, E. and Groth, J. (1999). Effects of plant species richness on invasion dynamics, disease outbreaks, insect abundances and diversity. *Ecology Letters* **2:**286-293.

Kowalik, R. A., Cooper, D. M., Evans, C. D. and Ormerod, S. J. (2007). Acidic episodes retard the biological recovery of upland British streams from chronic acidification. *Global Change Biology* **13:**2439-2452.

Kowalik, R. A. and Ormerod, S. J. (2006). Intensive sampling and transplantation experiments reveal continued effects of episodic acidification on sensitive stream invertebrates. *Freshwater Biology* **51:**180-191.

Kurita, T., Aoyama, H., Saitoh, S., Shinzato, N., Sawada, K., Kuriwada, T., Hironaka, K., Inomata, N., Yamahira, K. and Toda, M. (2014). Isolation and validation of eight

microsatellite loci in a damselfly Ischnura senegalensis (Odonata: Coenagrionidae) by pyrosequencing a bead-enriched library. *Applied Entomology and Zoology* **49:**623-626.

Kuusela, K. and Huusko, A. R. I. (1996). Post-emergence migration of stoneflies (Plecoptera) into the nearby forest. *Ecological Entomology* **21:**171-177.

L

Laikre, L., Nilsson, T., Primmer, C. R., Ryman, N. and Allendorf, F. W. (2009). Importance of genetics in the interpretation of favourable conservation status. *Conservation Biology* **23:**1378-1381.

Laliberté, E. and Legendre, P. (2010). A distance-based framework for measuring functional diversity from multiple traits. *Ecology* **91:**299-305.

Lamy, T., Jarne, P., Laroche, F., Pointier, J. P., Huth, G., Segard, A. and David, P. (2013). Variation in habitat connectivity generates positive correlations between species and genetic diversity in a metacommunity. *Molecular Ecology* **22:**4445-4456.

Lankau, R. A. and Strauss, S. Y. (2007). Mutual Feedbacks Maintain Both Genetic and Species Diversity in a Plant Community. *Science* **317:**1561-1563.

Laroche, F., Jarne, P., Lamy, T., David, P. and Massol, F. (2015). A neutral theory for interpreting correlations between species and genetic diversity in communities. *The American Naturalist* **185:**59-69.

Larrañaga, A., Layer, K., Basaguren, A., Pozo, J. and Woodward, G. (2010). Consumer body composition and community structure in a stream is altered by pH. *Freshwater Biology* **55:**670-680.

Lavandier, P. and Dumas, J. (1971). Cycles de développement de quelques invertébrés benthiques dans des ruisseaux des Pyrénées centrales. . *Annales de Limnologie* **7:**151-172.

Leakey, R. and Lewin, R. (1995). The Sixth Extinction: Biodiversity and its Survival. London: Phoenix.

Leblois, R., Pudlo, P., Néron, J., Bertaux, F., Reddy Beeravolu, C., Vitalis, R. and Rousset, F. (2014). Maximum-likelihood inference of population size contractions from microsatellite data. *Molecular Biology and Evolution* **31:**2805-2823.

Ledger, M. E. and Hildrew, A. G. (2005). The ecology of acidification and recovery: changes in herbivore-algal food web linkages across a stream pH gradient. *Environmental Pollution* **137:**103-118.

Lees, J. H. and Ward, R. D. (1987). Genetic variation and biochemical systematics of british nemouridae. *Biochemical Systematics and Ecology* **15:**117-125.

Leffler, E. M., Bullaughey, K., Matute, D. R., Meyer, W. K., Ségurel, L., Venkat, A., Andolfatto, P. and Przeworski, M. (2012). Revisiting an old riddle: what determines genetic diversity levels within species? *PLoS Biology* **10:**e1001388.

Lehmann, T., Besansky, N. J., Hawley, W. A., Fahey, T. G., Kamau, L. and Collins, F. H. (1997). Microgeographic structure of Anopheles gambiae in western Kenya based on mtDNA and microsatellite loci. *Molecular Ecology* **6:**243-253.

Leigh, J. W. and Bryant, D. (2015). Popart: full-feature software for haplotype network construction. *Methods in Ecology and Evolution* **6:**1110-1116.

References

LeRoy, C. J., Wooley, S. C. and Lindroth, R. L. (2012). Genotype and soil nutrient environment influence aspen litter chemistry and in-stream decomposition. *Freshwater Science* **31:**1244-1253.

Leys, M., Keller, I., Räsänen, K., Gattolliat, J.-L. and Robinson, C. T. (2016). Distribution and population genetic variation of cryptic species of the Alpine mayfly *Baetis alpinus* (Ephemeroptera: Baetidae) in the Central Alps. *BMC Evolutionary Biology* **16:**1-15.

Liu, C.-y., Yuan, C., Li, X.-t., Li, Z.-H., Zhao, P., Qian, Z.-Q., Li, J.-f. and Liu, Z.-L. (2015). Development of microsatellites from *Rehmannia glutinosa* transcriptome and evaluation of genetic diversity among *Rehmannia* species. *Biochemical Systematics and Ecology* **59:**177-182.

Lock, K. and Goethals, P. L. M. (2008). Distribution and ecology of the stoneflies (Plecoptera) of Flanders (Belgium). *Annales de Limnologie* **44:**203-213.

Lomolino, M. V. (2000). A call for a new paradigm of island biogeography. *Global Ecology and Biogeography* **9:**1-6.

Lopes-Lima, M., Sousa, R., Teixeira, A., Varandas, S., Riccardi, N., Aldridge, D. C. and Froufe, E. (2015). Newly developed microsatellite markers for the pan-European duck mussel, *Anodonta anatina*: revisiting the main mitochondrial lineages. *Aquatic Conservation: Marine and Freshwater Ecosystems* **26***:* 307–318*.*

López-rodríguez, M. J., Bo, T., Tierno De Figueroa, J. M. and Fenoglio, S. (2010). Nymphal trophic behaviour of two Nemouridae species (Insecta, Plecoptera) in the Curone creek (northern Apennines, Italy). *Entomological Science* **13:**288-292.

Luikart, G. and Cornuet, J.-M. (1998). Empirical evaluation of a test for identifying recently bottlenecked populations from allele frequency data. *Conservation Biology* **12:**228-237.

Luikart, G., Sherwin, W. B., Steele, B. M. and Allendorf, F. W. (1998). Usefulness of molecular markers for detecting population bottlenecks via monitoring genetic change. *Molecular Ecology* **7:**963-974.

M

Macarthur, R. and Levins, R. (1967). The limiting similarity, convergence, and divergence of coexisting species. *The American Naturalist* **101:**377-385.

MacArthur, R. H. and Wilson, E. O. (1967). The Theory of Island Biogeography. Princeton University Press, Princeton.

Macdonald, H. C., Cunha, L. and Bruford, M. W. (2016a). Development of genomic resources for four potential environmental bioindicator species: *Isoperla grammatica*, *Amphinemura sulcicollis*, *Oniscus asellus* and *Baetis rhodani*. *bioRxiv* Doi:10.1101/046227.

Macdonald, H. C., Ormerod, S. J. and Bruford, M. W. (2016b). Enhancing capacity for freshwater conservation at the genetic level: a demonstration using three stream macroinvertebrates. *Aquatic Conservation: Marine and Freshwater Ecosystems* Doi: 10.1002/aqc.2691**:** *in press*.

Macher, J. N., Salis, R. K., Blakemore, K. S., Tollrian, R., Matthaei, C. D. and Leese, F. (2016). Multiple-stressor effects on stream invertebrates: DNA barcoding reveals contrasting responses of cryptic mayfly species. *Ecological Indicators* **61:**159-169.

References

Macneil, C., Dick, J. T. A. and Elwood, R. W. (1997). The trophic ecology of freshwater Gammarus spp. (Crustacea: Amphipoda): problems and perspectives concerning the functional feeding group concept. *Biological Reviews* **72:**349-364.

Magurran, A. E. (2005). Ecology: linking species diversity and genetic diversity. *Current Biology* **15:**R597-R599.

Maitland, P. S. (1966). The distribution, life cycle, and predators of *Amphinemura sulcicollis* (Stephens) (Plecoptera) in the River Endrick, Scotland. *The Entomologist* **99:**72-81.

Malmgvist, B., Sjöström, P. and Frick, K. (1991). The diet of two species of Isoperla (Plecoptera: Perlodidae) in relation to season, site, and sympatry. *Hydrobiologia* **213:**191-203.

Malmqvist, B. and Sjöström, P. (1989). The life cycle and growth of *Isoperla grammatica* and *I. difformis* (Plecoptera) in southernmost Sweden: intra- and interspecific considerations. *Hydrobiologia* **175:**97-108.

Manfrin, A., Traversetti, L., Pilotto, F., Larsen, S. and Scalici, M. (2016). Effect of spatial scale on macroinvertebrate assemblages along a Mediterranean river. *Hydrobiologia* **765:**185-196.

Marchant, S., Moran, A. L. and Marko, P. B. (2015). Out-of-the tropics or trans-tropical dispersal? The origins of the disjunct distribution of the gooseneck barnacle *Pollicipes elegans*. *Frontiers in Zoology* **12:**1-16.

Marshall, J. L. and Camp, C. D. (2006). Environmental correlates of species and genetic richness in lungless salamanders (family plethodontidae). *Acta Oecologica* **29:**33-44.

References

Marske, K. A., Rahbek, C. and Nogués-Bravo, D. (2013). Phylogeography: spanning the ecology-evolution continuum. *Ecography* **36:**1169-1181.

Mason, N. W. H., Mouillot, D., Lee, W. G. and Wilson, J. B. (2005). Functional richness, functional evenness and functional divergence: the primary components of functional diversity. *Oikos* **111:**112-118.

Master, L., Flack, S. and Stein, B. (1998). Rivers of life: critical watersheds for protecting freshwater biodiversity. Nature Conservancy, Arlington, VA.

Masters, Z. (2002). Factors affecting stream macroinvertebrate recovery from acidification. Thesis PhD thesis, Cardiff University, Cardiff.

Masters, Z., Peteresen, I., Hildrew, A. G. and Ormerod, S. J. (2007). Insect dispersal does not limit the biological recovery of streams from acidification. *Aquatic Conservation: Marine and Freshwater Ecosystems* **17:**375-383.

Matschiner, M. and Salzburger, W. (2009). TANDEM: integrating automated allele binning into genetics and genomics workflows. *Bioinformatics* **25(15):**1982-1983.

McNeely, J. A., Miller, K. R., Reid, W. V., Mittermeier, R. A. and Werner, T. B. (1990). Conserving the world's biological diversity. IUCN, World Resources Institute, Conservation International, WWF-US and the World Bank, Washington, DC.

Meglecz, E., Petenian, F., Danchin, E., D'Acier, A. C., Rasplus, J.-Y. and Faure, E. (2004). High similarity between flanking regions of different microsatellites detected within each of two species of Lepidoptera: *Parnassius apollo* and *Euphydryas aurinia*. *Molecular Ecology* **13:**1693-1700.

References

Messmer, V., Jones, G. P., Munday, P. L. and Planes, S. (2012). Concordance between genetic and species diversity in coral reef fishes across the Pacific Ocean biodiversity gradient. *Evolution* **66:**3902-3917.

Meyer, C. P. and Paulay, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology* **3:**e422.

Mills, L. S. and Allendorf, F. W. (1996). The one-migrant-per-generation rule in conservation and management. *Conservation Biology* **10**:1509-1518.

Minhós, T., Chikhi, L., Sousa, C., Vicente, L. M., Ferreira da Silva, M., Heller, R., Casanova, C. and Bruford, M. W. (2016). Genetic consequences of human forest exploitation in two colobus monkeys in Guinea Bissau. *Biological Conservation* **194:**194-208.

Moreira, X., Mooney, K. A., Zas, R. and Sampedro, L. (2012). Bottom-up effects of host-plant species diversity and top-down effects of ants interactively increase plant performance. *Proceedings of the Royal Society B: Biological Sciences* **279:**4464-4472.

Moss, B. (1998). Ecology of Freshwaters: Man and Medium, Past to Future. Oxford, UK.

Mueller, U. G. and Wolfenbarger, L. L. (1999). AFLP genotyping and fingerprinting. *Trends in Ecology and Evolution* **14:**389-394.

Murphy, J. F., Davy-Bowker, J., McFarland, B. and Ormerod, S. J. (2013). A diagnostic biotic index for assessing acidity in sensitive streams in Britain. *Ecological Indicators* **24:**562-572.

Murphy, J. F., Winterbottom, J. H., Orton, S., Simpson, G. L., Shilland, E. M. and Hildrew, A. G. (2014). Evidence of recovery from acidification in the macroinvertebrate assemblages of UK fresh waters: A 20-year time series. *Ecological Indicators* **37:**330-340.

N

Navarro-Ortega, A., Acuña, V., Bellin, A., Burek,P., Cassiani, G., Choukr-Allah, R., Dolédec, S., Elosegi, A., Ferrari, F., Ginebreda, A., Grathwohl, P., Jones, C., Rault, P. K., Kok, K., Koundouri, P., Ludwig, R. P., Merz, R., Milacic, R., Muñoz, I., Nikulin, G., Paniconi, C., Paunović, M., Petrovic, M., Sabater, L., Sabater, S., Skoulikidis, N. T., Slob, A., Teutsch, G., Voulvoulis, N. and Barceló, D. (2015). Managing the effects of multiple stressors on aquatic ecosystems under water scarcity. The GLOBAQUA project, *Science of The Total Environment* **503–504**: 3-9.

Naveda-Rodríguez, A., Bildstein, K. L. and Vargas, F. H. (2016). Geographic patterns of species richness of diurnal raptors in Venezuela. *Biodiversity and Conservation* **25:**1037-1052.

Nei, M. (1987). Molecular Evolutionary Genetics. Colombia University Press, New York.

Nelson, M. F. and Anderson, N. O. (2013). How many marker loci are necessary? Analysis of dominant marker data sets using two popular population genetic algorithms. *Ecology and Evolution* **3:**3455-3470.

Newton, A. C. (2011). Social-ecological resilience and biodiversity conservation in a 900-yearold protected area. *Ecology and Society* **16**:13.

Newton, A. C. (2016). Biodiversity risks of adopting resilience as a policy goal. *Conservation Letters* Doi:10.1111/conl.12227.

Newton, A. C. and Cantarello, E. (2015). Restoration of forest resilience: an achievable goal? *New Forests* **46:**645-668.

Nimmo, D. G., Mac Nally, R., Cunningham, S. C., Haslem, A. and Bennett, A. F. (2015). Vive la resistance: reviving resistance for 21st century conservation. *Trends Ecol Evol* **30:**516-523.

Nordborg, M. (2004). Coalescent Theory. *Handbook of Statistical Genetics*. Doi:10.1002/0470022620.bbc21**:** John Wiley & Sons, Ltd.

Nosil, P., Vines, T.H. and Funk, D.J. (2005) Perspective: reproductive isolation caused by natural selection against immigrants from divergent habitats. *Evolution* **59**:705–719.

Nowak, C., Hankeln, T., Schmidt, E. R. and Schwenk, K. (2006). Development and localization of microsatellite markers for the sibling species *Chironomus riparius* and *Chironomus piger* (Diptera: Chironomidae). *Molecular Ecology Notes* **6:**915-917.

Nowak, C., Vogt, C., Oehlmann, J., Pfenninger, M., Schwenk, K., Streit, B. and Oetken, M. (2012). Impact of genetic diversity and inbreeding on the life-history of *Chironomus midges* over consecutive generations. *Chemosphere* **88:**988-993.

Nukazawa, K., Kazama, S. and Watanabe, K. (2015). A hydrothermal simulation approach to modelling spatial patterns of adaptive genetic variation in four stream insects. *Journal of Biogeography* **42:**103-113.

O

O'Connell, M. and Wright, J. M. (1997). Microsatellite DNA in fishes. *Reviews in Fish Biology and Fisheries* **7:**331-363.

Odat, N., Hellwig, F. H., Jetschke, G. and Fischer, M. (2010). On the relationship between plant species diversity and genetic diversity of *Plantago lanceolata* (Plantaginaceae) within and between grassland communities. *Journal of Plant Ecology* **3:**41-48.

Odat, N., Jetschke, G. and Hellwig, F. H. (2004). Genetic diversity of *Ranunculus acris* L. (Ranunculaceae) populations in relation to species diversity and habitat type in grassland communities. *Molecular Ecology* **13:**1251-1257.

Oliver, T. H., Heard, M. S., Isaac, N. J. B., Roy, D. B., Procter, D., Eigenbrod, F., Freckleton, R., Hector, A., Orme, C. D. L., Petchey, O. L., Proença, V., Raffaelli, D., Suttle, K. B., Mace, G. M., Martín-López, B., Woodcock, B. A. and Bullock, J. M. (2015). Biodiversity and resilience of ecosystem functions. *Trends in Ecology & Evolution* **30:**673-684.

Ormerod, S. J., Donald, A. P. and Brown, S. J. (1989). The influence of plantation forestry on the pH and aluminium concentration of upland welsh streams: a re-examination. *Environmental Pollution* **62:**47-62.

Ormerod, S. J. and Durance, I. (2009). Restoration and recovery from acidification in upland Welsh streams over 25 years. *Journal of Applied Ecology* **46:**164-174.

Ormerod, S. J. and Jenkins, A. (1994). The biological effects of acid episodes. In: Acidification of Freshwater Ecosystems: Implications for the Future. 272. Wiley, Chichester.

Orozco-terWengel, P., Andreone, F., Louis, E. and Vences, M. (2013). Mitochondrial introgressive hybridization following a demographic expansion in the tomato frogs of Madagascar, genus Dyscophus. *Molecular Ecology* **22:**6074-6090.

Östergren, J., Nilsson, J., Lundqvist, H., Dannewitz, J. and Palm, S. (2015). Genetic baseline for conservation and management of sea trout in the northern Baltic Sea. *Conservation Genetics* **17**:177-191.

References

P

Paetkau, D., Amstrup, S. C., Born, E. W., Calvert, W., Derocher, A. E., Garner, G. W., Messier, F., Stirling, I., Taylor, M. K., Wiig, Ø. and Strobeck, C. (1999). Genetic structure of the world's polar bear populations. *Molecular Ecology* **8:**1571-1584.

Palsbøll, P. J., Bérubé, M. and Allendorf, F. W. (2007). Identification of management units using population genetic data. *Trends in Ecology & Evolution* **22:**11-16.

Papadopoulou, A., Anastasiou, I., Spagopoulou, F., Stalimerou, M., Terzopoulou, S., Legakis, A. and Vogler, A. P. (2011). Testing the species--genetic diversity correlation in the Aegean archipelago: toward a haplotype-based macroecology? *The American Naturalist* **178:**241-255.

Pathirana, E., Takemon, Y., Takayanagi, A., Abbas, E. M., Tanida, K., Shimizu, N. and Kato, M. (2012). Isolation of repetitive sequence elements from the genome of a mayfly *Rhithrogena japonica* (Ephemeroptera: Heptageniidae). *Journal of Advanced Science* **24:**1-7.

Paz-Vinas, I., Loot, G., Stevens, V. M. and Blanchet, S. (2015). Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. *Molecular Ecology* **24:**4586-4604.

Peakall, R. and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28:**2537-2539.

Peakall, R. O. D. and Smouse, P. E. (2006). genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* **6:**288-295.

Pease, C. M., Lande, R. and Bull, J. J. (1989). A model of population growth, dispersal and evolution in a changing environment. *Ecology* **70:**1657-1664.

Peery, M. Z., Kirby, R., Reid, B. N., Stoelting, R., Doucet-Beer, E., Robinson, S., Vasquez-Carrillo, C., Pauli, J. N. and Palsboll, P. J. (2012). Reliability of genetic bottleneck tests for detecting recent population declines. *Molecular Ecology* **21:**3403-3418.

Pelosi, C., Pey, B., Caro, G., Cluzeau, D., Peigné, J., Bertrand, M. and Hedde, M. (2016). Dynamics of earthworm taxonomic and functional diversity in ploughed and no-tilled cropping systems. *Soil and Tillage Research* **156:**25-32.

Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., Bruford, M. W., Brummitt, N., Butchart, S. H. M., Cardoso, A. C., Coops, N. C., Dulloo, E., Faith, D. P., Freyhof, J., Gregory, R. D., Heip, C., Höft, R., Hurtt, G., Jetz, W., Karp, D. S., McGeoch, M. A., Obura, D., Onoda, Y., Pettorelli, N., Reyers, B., Sayre, R., Scharlemann, J. P. W., Stuart, S. N., Turak, E., Walpole, M. and Wegmann, M. (2013). Essential biodiversity variables. *Science* **339:**277-278.

Pérez-Portela, R., Noyer, C. and Becerro, M. A. (2015). Genetic structure and diversity of the endangered bath sponge Spongia lamella. *Aquatic Conservation: Marine and Freshwater Ecosystems* **25:**365-379.

Perna, C. and Burrows, D. (2005) Improved dissolved oxygen status following removal of exotic weed mats in important fish habitat lagoons of the tropical Burdekin River floodplain, Australia. *Marine Pollution Bulletin* **51**: 138-148.

Petersen, I., Masters, Z., Hildrew, A. G. and Ormerod, S. J. (2004). Dispersal of adult aquatic insects in catchments of differing land use. *Journal of Applied Ecology* **41:**934-950.

Petersen, I., Winterbottom, J. H., Orton, S., Friberg, N., Hildrew, A. G., Spiers, D. C. and Gurney, W. S. C. (1999). Emergence and lateral dispersal of adult Plecoptera and Trichoptera from Broadstone Stream, U.K. *Freshwater Biology* **42:**401-416.

Petraitis, P. (2013). Multiple Stable States in Natural Ecosystems. Oxford, UK: Oxford University Press.

Petrin, Z. (2011). Species traits predict assembly of mayfly and stonefly communities along pH gradients. *Oecologia* **167:**513-524.

Petrin, Z., Laudon, H. and Malmqvist, B. (2007). Does freshwater macroinvertebrate diversity along a pH-gradient reflect adaptation to low pH? *Freshwater Biology* **52:**2172-2183.

Pfenninger, M. and Schwenk, K. (2007). Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology* **7:**121.

Pfrender, M. E., Hawkins, C. P., Bagley, M., Courtney, G. W., Creutzburg, B. R., Epler, J. H., Fend, S., Schindel, D., Ferrington, L. C., Jr., Hartzell, P. L., Jackson, S., Larsen, D. P., Levesque, A., Morse, J. C., Petersen, M. J., Ruiter, D. and Whiting, M. (2010). Assessing macroinvertebrate biodiversity in freshwater ecosystems: advances and challenges in DNA-based approaches. *The Quarterly Review of Biology* **85:**319-340.

Phillipsen, I. C. and Lytle, D. A. (2013). Aquatic insects in a sea of desert: population genetic structure is shaped by limited dispersal in a naturally fragmented landscape. *Ecography* **36:**731-743.

Pimentel, D. (1968). Population regulation and genetic feedback. *Science* **159:**1432-1437.

Piry, S., Luikart, G. and Cornuet, J.-M. (1999). Computer note. BOTTLENECK: a computer program for detecting recent reductions in the effective size using allele frequency data. *Journal of Heredity* **90:**502-503.

References

Postaire, B., Aurelle, D., Bourmaud, C. A. F., Bruggemann, J. H. and Magalon, H. (2015). Isolation and characterisation of 26 microsatellite loci from a widespread tropical hydrozoan, *Macrorhynchia phoenicea* (Leptothecata, Aglaopheniidae), and cross-amplification in closely related species. *Biochemical Systematics and Ecology* **62:**137-141.

Postel, S. and Richter, B. (2003). Rivers for life: managing water for people and nature. Island Press, Washington.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S. and Rafalski, A. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding* **2:**225-238.

Pritchard, J. K., Stephens, M. and Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics* **155:**945–959.

Purser, A. F. (1966). Increase in heterozygote frequency with differential fertility. *Heredity (Edinb)* **21:**322-327.

Puşcaş, M., Taberlet, P. and Choler, P. (2008). No positive correlation between species and genetic diversity in European alpine grasslands dominated by Carex curvula. *Diversity and Distributions* **14:**852-861.

Putman, A. I. and Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution* **4:**4399-4428.

Pye, M. C., Vaughan, I. P. and Ormerod, S. J. (2012). Episodic acidification affects the breakdown and invertebrate colonisation of oak litter. *Freshwater Biology* **57:**2318-2329.

Q

QGIS Development Team. (2016). QGIS Geographic Information System. Open Source Geospatial Foundation. [Online]. Available at: http://qgis.osgeo.org [Accessed: May 2016].

R

Raeymaekers, J. A. M., Maes, G. E., Audenaert, E. and Volckaert, F. A. M. (2005). Detecting Holocene divergence in the anadromous–freshwater three-spined stickleback (*Gasterosteus aculeatus*) system. *Molecular Ecology* **14:**1001-1014.

Raymond, M. and Rousset, F. (1995). GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86:**248-249.

Rebora, M., Lucentini, L., Palomba, A., Panara, F. and Gaino, E. (2005). Genetic differentiation among populations of *Baetis rhodani* (Ephemeroptera, Baetidae) in three Italian streams. *Italian Journal of Zoology* **72:**121-126.

Resh, V. H. and Rosenberg, D. M. (2014). Economic Aspects of Freshwater Invertebrates.*Thorp and Covich's Freshwater Invertebrates: Ecology and General Biology: Fourth Edition*. Vol. 1. pp. 93-109.

Reusch, T. B. H., Ehlers, A., Hämmerli, A. and Worm, B. (2005). Ecosystem recovery after climatic extremes enhanced by genotypic diversity. *Proceedings of the National Academy of Sciences of the United States of America* **102:**2826-2831.

Rice, W. R. (1989). Analyzing tabled of statistical tests. *Evolution* **43:**223-225.

Rice, W. R. and Hostert, E. E. (1993). Laboratory experiments on speciation: what have we learned in 40 Years? *Evolution* **47:**1637-1653.

Robertson, H. L., Guzik, M. T. and Murphy, N. P. (2014). Persistence in the desert: Ephemeral waterways and small-scale gene flow in the desert spring amphipod, *Wangiannachiltonia guzikae*. *Freshwater Biology* **59:**653-665.

Robinson, J. D., Diaz-Ferguson, E., Poelchau, M. F., Pennings, S., Bishop, T. D. and Wares, J. (2010). Multiscale diversity in the marshes of the georgia coastal ecosystems LTER. *Estuaries and Coasts* **33:**865-877.

Rodrigues, A. S. L., Andelman, S. J., Bakarr, M. I., Boitani, L., Brooks, T. M., Cowling, R. M., Fishpool, L. D. C., da Fonseca, G. A. B., Gaston, K. J., Hoffmann, M., Long, J. S., Marquet, P. A., Pilgrim, J. D., Pressey, R. L., Schipper, J., Sechrest, W., Stuart, S. N., Underhill, L. G., Waller, R. W., Watts, M. E. J. and Yan, X. (2004). Effectiveness of the global protected area network in representing species diversity. *Nature* **428:**640-643.

Rodríguez, F., Pérez, T., Hammer, S. E., Albornoz, J. and Domínguez, A. (2010). Integrating phylogeographic patterns of microsatellite and mtDNA divergence to infer the evolutionary history of chamois (genus Rupicapra). *BMC Evolutionary Biology* **10**:222.

Rosenzweig, M. L. (1995). Species Diversity in Space and Time. Cambridge: Cambridge University Press.

Rousset, F. (2008). Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Molecular Ecology. Resources* **8:**103-106.

Rundle, S. D., Foggo, A., Choiseul, V. and Bilton, D. T. (2002). Are distribution patterns linked to dispersal mechanism? An investigation using pond invertebrate assemblages. *Freshwater Biology* **47:**1571-1581.

S

Saito, R. and Tojo, K. (2016). Comparing spatial patterns of population density, biomass, and genetic diversity patterns of the habitat generalist mayfly *Isonychia japonica* Ulmer (Ephemeroptera:Isonychiidae) in the Chikuma–Shinano River basin. *Freshwater Science* **35** doi:10.1086/686537.

Sala, O. E., Stuart Chapin , F., III, Armesto, J. J., Berlow, E., Bloomfield, J., Dirzo, R., Huber-Sanwald, E., Huenneke, L. F., Jackson, R. B., Kinzig, A., Leemans, R., Lodge, D. M., Mooney, H. A., Oesterheld, M. n., Poff, N. L., Sykes, M. T., Walker, B. H., Walker, M. and Wall, D. H. (2000). Global biodiversity scenarios for the year 2100. *Science* **287:**1770-1774.

Salmona, J., Salamolard, M., Fouillot, D., Ghestemme, T., Larose, J., Centon, J.-F., Sousa, V., Dawson, D. A., Thebaud, C. and Chikhi, L. (2012). Signature of a Pre-Human Population Decline in the Critically Endangered Reunion Island Endemic Forest Bird *Coracina newtoni*. *PLoS ONE* **7:**e43524.

Saura, S., Bodin, Ö. and Fortin, M.-J. (2014). Editor's choice: Stepping stones are crucial for species' long-distance dispersal and range expansion through habitat networks. *Journal of Applied Ecology* **51:**171-182.

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18:**233-234.

Schuldt, A. and Assmann, T. (2010). Invertebrate diversity and national responsibility for species conservation across Europe - a multi-taxon approach. *Biological Conservation* **143:**2747-2756.

References

Schultheis, A. S., Marchant, R. and Hughes, J. M. (2008). Contrasting patterns of genetic structure and disequilibrium in populations of a stone-cased caddisfly (Tasimiidae) from northern and southern Australia. *Marine and Freshwater Research* **59:**235-245.

Schwartz, M. K., Luikart, G. and Waples, R. S. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution* **22:**25-33.

Sei, M., Lang, B. K. and Berg, D. J. (2009). Genetic and community similarities are correlated in endemic-rich springs of the northern Chihuahuan Desert. *Global Ecology and Biogeography* **18:**192-201.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal* **27:**379-423.

Shaw, P. W., Pierce, G. J. and Boyle, P. R. (1999a). Subtle population structuring within a highly vagile marine invertebrate, the veined squid *Loligo forbesi*, demonstrated with microsatellite DNA markers. *Molecular Ecology* **8:**407-417.

Shaw, P. W., Turan, C., Wright, J. M., O'Connell, M. and Carvalho, G. R. (1999b). Microsatellite DNA analysis of population structure in Atlantic herring (*Clupea harengus*), with direct comparison to allozyme and mtDNA RFLP analyses. *Heredity* **83:**490-499.

Shaw, R. F., Johnson, P. J., Macdonald, D. W. and Feber, R. E. (2015). Enhancing the biodiversity of ditches in intensively managed UK farmland. *PLoS ONE* **10**:e0138306.

Sievert, N. A., Paukert, C. P., Tsang, Y.-P. and Infante, D. (2016). Development and assessment of indices to determine stream fish vulnerability to climate change and habitat alteration. *Ecological Indicators* **67:**403-416.

References

Silvertown, J., Biss, P. M. and Freeland, J. (2009). Community genetics: Resource addition has opposing effects on genetic and species diversity in a 150-year experiment. *Ecology Letters* **12:**165-170.

Smith, B. J. (2007). boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference *Journal of Statistical Software* **21(11):**1-37.

Smith, C., Good, M., Murphy, J. F., Giller, P. S. and O'Halloran, J. (2000). Life-history patterns and spatial and temporal overlap in an assemblage of lotic Plecoptera in the Araglin catchment study area, Ireland. *Archiv fur Hydrobiologie* **150:**117-132.

Smith, E. P. and van Belle, G. (1984). Nonparametric Estimation of Species Richness. *Biometrics* **40:**119-129.

Smith, P. J. (2005). Random amplified polymorphic DNA (RAPD). Stock Identification Methods: applications in fishery science. Cadrin, SX, KD Friedland and J. Waldman, eds. Elsevier Academic Press, Amsterdam, pp.371-388.

Smith, P. J. and Collier, K. J. (2001). Allozyme diversity and population genetic structure of the caddisfly Orthopsyche fimbriata and the mayfly *Acanthophlebia cruentata* in New Zealand streams. *Freshwater Biology* **46:**795-805.

Sode, A. and Wiberg-Larsen, P. (1993). Dispersal of adult Trichoptera at a Danish forest brook. *Freshwater Biology* **30:**439-446.

Soule, M. and Yang, S. Y. (1973). Genetic variation in side blotched lizards on islands in the Gulf of California. *Evolution* **27:**593-600.

References

Spellerberg, I. F. and Fedor, P. J. (2003). A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener' Index. *Global Ecology and Biogeography* **12:**177-179.

Statzner, B. and Bêche, L. A. (2010). Can biological invertebrate traits resolve effects of multiple stressors on running water ecosystems? *Freshwater Biology* **55:**80-119.

Storey, R. (2015). Macroinvertebrate community responses to duration, intensity and timing of annual dry events in intermittent forested and pasture streams. *Aquatic Sciences* **78**:395-414.

Stork, N. E. (2007). Biodiversity: world of insects. *Nature* **448:**657-658.

Storz, J. F. and Beaumont, M. A. (2002). Testing for genetic evidence of population expansion and contraction: an empirical analysis of microsatellite DNA variation using a hierarchical Bayesian model. *Evolution* **56:**154–166.

Strayer, D. L. (2006). Challenges for freshwater invertebrate conservation. *Journal of the North American Benthological Society* **25:**271-287.

Struebig, M. J., Kingston, T., Petit, E. J., Le Comber, S. C., Zubaid, A., Mohd-Adnan, A. and Rossiter, S. J. (2011). Parallel declines in species and genetic diversity in tropical forest fragments. *Ecology Letters* **14:**582-590.

Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology & Evolution* **15:**199-203.

Svensson, B. W. (1974). Population movement of adult Trichoptera at a south Swedish stream. *Oikos* **25:**157-175.

Svensson, E. (1977). Growth rates of the stonefly, *Amphinemura sulcicollis* Stephens, in the upper River Wear. *Freshwater Biology* **7:**455-460.

Svitok, M., Novikmec, M., Bitušík, P., Máša, B., Oboňa, J., Očadlík, M. and Michalková, E. (2014). Benthic communities of low-order streams affected by acid mine drainages: a case study from central Europe. *Water (Switzerland)* **6:**1312-1338.

T

Taberlet, P., Zimmermann, N. E., Englisch, T., Tribsch, A., Holderegger, R., Alvarez, N., Niklfeld, H., Coldea, G., Mirek, Z., Moilanen, A., Ahlmer, W., Marsan, P. A., Bona, E., Bovio, M., Choler, P., Cieślak, E., Colli, L., Cristea, V., Dalmas, J. P., Frajman, B., Garraud, L., Gaudeul, M., Gielly, L., Gutermann, W., Jogan, N., Kagalo, A. A., Korbecka, G., Küpfer, P., Lequette, B., Letz, D. R., Manel, S., Mansion, G., Marhold, K., Martini, F., Negrini, R., Niño, F., Paun, O., Pellecchia, M., Perico, G., Piekoś-Mirkowa, H., Prosser, F., Puşcaş, M., Ronikier, M., Scheuerer, M., Schneeweiss, G. M., Schönswetter, P., Schratt-Ehrendorfer, L., Schüpfer, F., Selvaggi, A., Steinmann, K., Thiel-Egenter, C., van Loo, M., Winkler, M., Wohlgemuth, T., Wraber, T. and Gugerli, F. (2012). Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. *Ecology Letters* **15:**1439-1448.

Tachet, H., Richoux, P., Bournaud, M. and Usseglio-Polatera, P. (2002). Invertébrés d'eau douce. Systematique, biologie, écologie (2nd corrected impression). CNRS éditions, Paris, France.

Tallmon, D. A., Luikart, G. and Waples, R. S. (2004). The alluring simplicity and complex reality of genetic rescue. *Trends in Ecology and Evolution* **19**:489–496.

References

Tamura, K. and Nei, M. (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**:512-26.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28:**2731-2739.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular Biology and Evolution* **30:**2725-2729.

Taper, M. L. and Case, T. J. (1992). Coevolution among competitors. *Journal of Evolutionary Biology* **8:**63-109.

Taubmann, J., Theissinger, K., Feldheim, K. A., Laube, I., Graf, W., Haase, P., Johannesen, J. and Pauls, S. U. (2011). Modelling range shifts and assessing genetic diversity distribution of the montane aquatic mayfly *Ameletus inopinatus* in Europe under climate change scenarios. *Conservation Genetics* **12:**503-515.

Tenesa, A., Navarro, P., Hayes, B. J., Duffy, D. L., Clarke, G. M., Goddard, M. E. and Visscher, P. M. (2007). Recent human effective population size estimated from linkage disequilibrium. *Genome Research* **17:**520-526.

Tero, N., Aspi, J., Siikamäki, P., Jäkäläniemi, A. and Tuomi, J. (2003). Genetic structure and gene flow in a metapopulation of an endangered plant species, *Silene tatarica*. *Molecular Ecology* **12:**2073-2085.

References

Teske, P. R., Sandoval-Castillo, J., Sasaki, M. and Beheregaray, L. B. (2015). Invasion success of a habitat-forming marine invertebrate is limited by lower-than-expected dispersal ability. *Marine Ecology Progress Series* **536:**221-227.

Theissinger, K., Bálint, M., Feldheim, K. A., Haase, P., Johannesen, J., Laube, I. and Pauls, S. U. (2013). Glacial survival and post-glacial recolonization of an arctic–alpine freshwater insect (*Arcynopteryx dichroa*, Plecoptera, Perlodidae) in Europe. *Journal of Biogeography* **40:**236-248.

Theissinger, K., Feldheim, K. A., Seitz, A. and Pauls, S. U. (2009). Isolation and characterization of 11 polymorphic trinucleotide microsatellite markers in the stonefly *Arcynopteryx compacta* (Plecoptera: Perlodidae). *Molecular Ecology Resources* **9:**357-359.

Theissinger, K., Feldheim, K. A., Taubmann, J., Seitz, A. and Pauls, S. U. (2008). Isolation and characterization of 10 highly polymorphic di- and trinucleotide microsatellite markers in the mayfly *Ameletus inopinatus* (Ephemeroptera: Siphlonuridae). *Molecular Ecology Resources* **8:**1285-1287.

Thiel-Egenter, C., Alvarez, N., Holderegger, R., Tribsch, A., Englisch, T., Wohlgemuth, T., Colli, L., Gaudeul, M., Gielly, L., Jogan, N., Linder, H. P., Negrini, R., Niklfeld, H., Pellecchia, M., Rioux, D., Schönswetter, P., Taberlet, P., van Loo, M., Winkler, M., IntraBioDiv, C. and Gugerli, F. (2011). Break zones in the distributions of alleles and species in alpine plants. *Journal of Biogeography* **38:**772-782.

Thomas, M. A., Walsh, K. A., Wolf, M. R., McPheron, B. A. and Marden, J. H. (2000). Molecular phylogenetic analysis of evolutionary trends in stonefly wing structure and locomotor behavior. *Proceedings of the National Academy of Sciences* **97:**13178-13183.

Tilman, D. (1982). Resource Competition and Community Structure. Princeton, NJ: Princeton University Press.

Turlure, C., Vandewoestijne, S. and Baguette, M. (2014). Conservation genetics of a threatened butterfly: comparison of allozymes, RAPDs and microsatellites. *BMC Genetics* **15:**1-12.

U

UN. (2015). Elaboration of data by United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: The 2015 Revision. (Medium-fertility variant). [Online]. Available at: Source: Worldometers (www.Worldometers.info) [Accessed: 02.06.2016].

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B., Remm, M. and Rozen, S. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research* **40**:e115.

V

Valencia-Cuevas, L., Piñero, D., Mussali-Galante, P., Valencia-Ávalos, S. and Tovar-Sánchez, E. (2014). Effect of a red oak species gradient on genetic structure and diversity of Quercus castanea (Fagaceae) in Mexico. *Tree Genetics & Genomes* **10**:641-652.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. and Shipley, P. (2004). Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4:**535-538.

Van Oosterhout, C., Weetman, D. and Hutchinson, W. F. (2006). Estimation and adjustment of microsatellite null alleles in nonequilibrium populations. *Molecular Ecology Notes* **6:**255-256.

References

Van Valen, L. (1965). Morphological variation and width of ecological niche. *The American Naturalist* **99:**377-390.

Vellend, M. (2003). Island biogeography of genes and species. *American Naturalist* **162:**358-365.

Vellend, M. (2004). Parallel effects of land-use history on species diversity and genetic diversity of forest herbs. *Ecology* **85:**3043-3055.

Vellend, M. (2005). Species diversity and genetic diversity: parallel processes and correlated patterns. *The American Naturalist* **166:**199-215.

Vellend, M. (2008). Effects of diversity on diversity: consequences of competition and facilitation. *Oikos* **117:**1075-1085.

Vellend, M. (2010). Conceptual synthesis in community ecology. *Quarterly Review of Biology* **85:**183-206.

Vellend, M. and Geber, M. A. (2005). Connections between species diversity and genetic diversity. *Ecology Letters* **8:**767-781.

Vellend, M., Lajoie, G., Bourret, A., Múrria, C., Kembel, S. W. and Garant, D. (2014). Drawing ecological inferences from coincident patterns of population- and community-level biodiversity. *Molecular Ecology* **23:**2890-2901.

Vignal, A., Milan, D., SanCristobal, M. and Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution* **34:**275-305.

Villalobos, F., Lira-Noriega, A., Soberón, J. and Arita, H. T. (2013). Range-diversity plots for conservation assessments: using richness and rarity in priority setting. *Biological Conservation* **158:**313-320.

Viñas, M. D., Blanco-Bercial, L., Bucklin, A., Verheye, H., Bersano, J. G. F. and Ceballos, S. (2015). Phylogeography of the copepod *Calanoides carinatus* s.l. (Krøyer) reveals cryptic species and delimits *C. carinatus* s.s. distribution in SW Atlantic Ocean. *Journal of Experimental Marine Biology and Ecology* **468:**97-104.

W

Walsh, P. S., Metzger, D. A. and Higuchi, R. (1991). Chelex-100 as a medium for simple extraction of DNA for PCR-basedtyping from forensic material. *BioTechniques* **10:**506-513.

Waples, R. S. (1990). Conservation Genetics of Pacific Salmon. II. Effective Population Size and the Rate of Loss of Genetic Variability. *Journal of Heredity* **81:**267-276.

Watts, P. C. and Thompson, D. J. (2012). Developmental plasticity as a cohesive evolutionary process between sympatric alternate-year insect cohorts. *Heredity* **108:**236-241.

Watts, P. C., Wu, J. H., Westgarth, C., Thompson, D. J. and Kemp, S. J. (2004). A panel of microsatellite loci for the Southern Damselfly, *Coenagrion mercuriale* (Odonata: Coenagrionidae). *Conservation Genetics* **5:**117-119.

Weatherley, N. S. and Ormerod, S. J. (1987). The impact of acidification on macroinvertebrate assemblages in welsh streams: towards an empirical model. *Environmental Pollution* **46:**223-240.

References

Wegmann, D., Leuenberger, C., Neuenschwander, S. and Excoffier, L. (2010). ABCtoolbox: A versatile toolkit for approximate Bayesian computations. *BMC Bioinformatics* **11**:116.

Wehenkel, C., Bergmann, F. and Gregorius, H.-R. (2006). Is there a trade-off between species diversity and genetic diversity in forest tree communities? *Plant Ecology* **185:**151-161.

Wei, X. and Jiang, M. (2012). Contrasting relationships between species diversity and genetic diversity in natural and disturbed forest tree communities. *New Phytologist* **193:**779-786.

Welsh Government. (2016). [Online]. Available at: http://thewaleswewant.co.uk/goals/resilient [Accessed: 4.07.2016].

Westemeier, R. L., Brawn, J. D., Simpson, S. A., Esker, T. L., Jansen, R. W., Walk, J. W., Kershner, E. L., Bouzat, J. L. and Paige, K. N. (1998). Tracking the long-term decline and recovery of an isolated population. *Science* **282:**1695-1698.

Whiles, M. R. (2013). Freshwater Invertebrate Ecology. Oxford Bibliographies in Ecology, Oxford University Press.

Whiteley, A. R., Spruell, P. and Allendorf, F. W. (2006). Can common species provide valuable information for conservation? *Molecular Ecology* **15:**2767-2786.

Whitham, T. G., Young, W. P., Martinsen, G. D., Gehring, C. A., Schweitzer, J. A., Shuster, S. M., Wimp, G. M., Fischer, D. G., Bailey, J. K., Lindroth, R. L., Woolbright, S. and Kuske, C. R. (2003). Community and ecosystem genetics: A consequence of the extended phenotype. *Ecology* **84:**559-573.

Whittaker, R. H. (1975). Communities and Ecosystems. New York: MacMillian Publishing Company.

Whittaker, R. J. (2000). Scale, succession and complexity in island biogeography: are we asking the right questions? *Global Ecology and Biogeography* **9:**75-85.

Wickson, S. J., Chester, E. T., Valenzuela, I., Halliday, B., Lester, R. E., Matthews, T. G. and Miller, A. D. (2014). Population genetic structure of the Australian caddisfly Lectrides varians Mosely (Trichoptera: Leptoceridae) and the identification of cryptic species in south-eastern Australia. *Journal of Insect Conservation* **18:**1037-1046.

Wilcock, H. R., Bruford, M. W., Hildrew, A. G. and Nichols, R. A. (2005). Recruitment, kin and the spatial genetic structure of a caddisfly *Plectrocnemia conspersa* in a southern English stream. *Freshwater Biology* **50:**1499-1514.

Wilcock, H. R., Bruford, M. W., Nichols, R. A. and Hildrew, A. G. (2007). Landscape, habitat characteristics and the genetic population structure of two caddisflies. *Freshwater Biology* **52:**1907-1929.

Wilcock, H. R., Hildrew, A. G. and Nichols, R. A. (2001a). Genetic differentiation of a European caddisfly: past and present gene flow among fragmented larval habitats. *Molecular Ecology* **10:**1821-1834.

Wilcock, H. R., Hildrew, A. G., Nichols, R. A. and Bruford, M. W. (2001b). Microsatellites for the net-spinning caddisfly Plectrocnemia conspersa (Polycentropodidae). *Molecular Ecology Notes* **1:**318-319.

Williams, G. C. (1975). Sex and Evolution. Princeton, NJ: Princeton University Press.

Williams, H. (2003). The genetic structure and dispersal of the mayfly *Baetis rhodani* in relation to acidication and inter-catchment distance. Thesis Ph.D., Cardiff University.

Williams, H. C., Ormerod, S. J. and Bruford, M. W. (2006). Molecular systematics and phylogeography of the cryptic species complex *Baetis rhodani* (Ephemeroptera, Baetidae). *Molecular Phylogenetics and Evolution* **40:**370-382.

Williams, H. C., Wilcock, H. R. and Bruford, M. W. (2002). Microsatellite loci for the mayfly *Baetis rhodani* (Baetidae, Ephemeroptera). *Molecular Ecology Notes* **2:**411-412.

Williamson, T. J., Cross, W. F., Benstead, J. P., Gíslason, G. M., Hood, J. M., Huryn, A. D., Johnson, P. W. and Welter, J. R. (2016). Warming alters coupled carbon and nutrient cycles in experimental streams. *Global Change Biology* **22:**2152-2164.

Wolf, K. L. and Flora, K. (2010). Mental Health and Function - A Literature Review. [Online]. In: Green Cities: Good Health. College of the Environment, University of Washington.: Available at: www.greenhealth.washington.edu [Accessed: 02.06.2016].

Woodward, G., Bonada, N., Brown, L. E., Death, R. G., Durance, I., Gray, C., Hladyz, S., Ledger, M. E., Milner, A. M., Ormerod, S. J., Thompson, R. M. and Pawar, S. (2016). The effects of climatic fluctuations and extreme events on running water ecosystems. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371** Doi: 10.1098/rstb.2015.0274.

X

Xenopoulos, M. A., Lodge, D. M., Alcamo, J., Märker, M., Schulze, K. and Van Vuuren, D. P. (2005). Scenarios of freshwater fish extinctions from climate change and water withdrawal. *Global Change Biology* **11:**1557-1564.

Xu, W., Liu, L., He, T., Cao, M., Sha, L., Hu, Y., Li, Q. and Li, J. (2016). Soil properties drive a negative correlation between species diversity and genetic diversity in a tropical seasonal rainforest. *Scientific Reports* **6:**20652.

Y

Yan, N. D., Leung, B., Keller, W., Arnott, S. E., Gunn, J. M. and Raddum, G. G. (2003). Developing conceptual frameworks for the recovery of aquatic biota from acidification. *AMBIO: A Journal of the Human Environment* **32:**165-169.

Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R., Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., Goddard, M. E. and Visscher, P. M. (2010). Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* **42:**565-569.

Yu, J.-N., Won, C., Jun, J., Lim, Y. and Kwak, M. (2011). Fast and Cost-Effective Mining of Microsatellite Markers Using NGS Technology: An Example of a Korean Water Deer *Hydropotes inermis argyropus*. *PLoS ONE* **6:**e26933.

Z

Zaccarelli, N., Petrosillo, I., Zurlini, G. and Riitters, K. H. (2008). Source/sink patterns of disturbance and cross-scale mismatches in a panarchy of social-ecological landscapes. *Ecology and Society* **13**:26.

Zhang, H., Wang, Z., Wang, S. and Li, H. (2012). Progress of genome wide association study in domestic animals. *Journal of Animal Science and Biotechnology* **3:**1-10.

References

Zheng, H. P., Li, L. and Zhang, G. F. (2012). Inbreeding depression for fitness-related traits and purging the genetic load in the hermaphroditic bay scallop *Argopecten irradians* irradians (Mollusca: Bivalvia). *Aquaculture* **366:**27-33.

Zhou, B., Yu, D., Ding, Z. and Xu, X. (2016). Comparison of genetic diversity in four Typha species (Poales, Typhaceae) from China. *Hydrobiologia* **770:**117-128.

Appendix A - Site Information and Sequencing Data

This appendix gives details of all freshwater sites sampled and all mtDNA CO1 sequences submitted to Genbank.

Table of tables

Table A1 257

Table A2 265

Table A1. Details of all sites where samples were collected in Wales, UK. Including *Site Code*[1]; *Site name*; *Year* = the year in which the samples were collected (both years the samples were collected in May); *E* and *N* = the Eastings and Northings (Coordinates are in EPSG:27700 - OSGB 1936 / British National Grid); *Alt* = Altitude; *Land-use*; *Catchment*; *Spp* = the species code (A = *Amphinemura sulcicollis*, B = *Baetis rhodani*, I = *Isoperla grammatica* and P = *Plectronemia conspersa*; *P. conspersa* was subsequently removed from ongoing analysis); a * beside the species code shows that this site was used in the final data set for that species; and *Total* = total number of individuals collected (identified to Genus level in the field). If the site was resampled the following year, it appears below in **bold** with a new site number.

Site Code	Site name	Year	E	N	Alt (m)	Land-use	Catchment	Spp	Total
1	Afon Bidno	2012	287650	282200	316	Moor/ heathland Rough Pasture	Wye	A B I P	22 / / 19
2	Afon Calettwr	2012	285760	349330	235	Moor/Heathland Rough Pasture	Conwy	A B I P	/ 80 4 4
3	Afon Colwyn	2012	257600	350900	190	Moor/ Heathland Rough Pasture	Glaslyn	A B I P	/ 26 65 4
93	Afon Colwyn	2013						A* B I* P	19 1 61 65
4	Afon Fechan	2012	297800	358700	400	Moorland/ Coniferous	Dee	A* B I P	33 40 6 3
94	Afon Fechan	2013						A* B* I P	50+ 40+ / 3
5	Afon Gain	2012	275500	333530	305	Moor/ Heathland	Mawddach	A B I P	2 30 39 2
6	Afon Pistyll	2012	275070	342810	440	Rough pasture	Dwyryd	A* B I P	45 / / 33
95	**Afon Pistyll**	**2013**						**A*** **B**	**50+** **/**

Appendix A - Site Information and Sequencing Data

Site Code	Site name	Year	E	N	Alt (m)	Land-use	Catchment	Spp	Total
								I **P**	**6** **33**
7	Aran	2012	314700	267200	290	Rough Pasture Tips/Waste (Disued tip)	Wye	A B I P	/ 48 2 1
8	Berwyn	2012	269400	259700	168	Rough Pasture	Teifi	A B I P	28 54 50 1
9	Brefi	2012	268100	254500	210	Moor/ Heathland Rough Pasture	Teifi	A* B* I P	26 68 70 34
96	**Brefi**	**2013**						**A*** **B*** **I** **P**	**19** **70** **51** **34**
10	Cerist (Afon)	2012	282450	316380	160	Moor/Heathland Rough Pasture	Dyfi	A B I* P	~3-5 54 117 9
11	Ceirw	2012	276830	328920	250	Rough Pasture Coniferous Woodland	Mawddach	A B I P	117 26 2 /
12	CI1	2012	276268	257632	370	Rough pasture/ moorland	Tywi	A B I* P	25 60 30 25
97	**CI1**	**2013**						**A** **B*** **I*** **P**	**3** **50** **17** **25**
13	CI2	2012	276379	257381	373	Rough Pasture	Tywi	A B I P	/ 6 53 18
14	CI4	2012	277065	256513	364	Rough Pasture	Tywi	A B I P	2 63 26 24
98	**CI4**	**2013**						**A** **B** **I*** **P**	**/** ***20*** **41** **24**
15	CI5	2012	277454	255697	350	Rough Pasture	Tywi	A B I	2 74 2

Site Code	Site name	Year	E	N	Alt (m)	Land-use	Catchment	Spp	Total
								P	20
99	**CI5**	**2013**						**A** **B** **I** **P**	**/** **40+** **2** **20**
16	Clywedog	2012	308300	265000	205	Rough Pasture	Wye	A B I P	/ 54 52 4
17	Cnyffiad	2012	290700	252300	269	Rough Pasture Deciduous Woodland	Wye	A B I P	1 41 3 /
18	Dulais	2012	269900	240000	225	Rough Pasture Coniferous Woodland	Tywi	A B I P	21 34 / /
19	Dulas North	2012	277860	310450	140	Rough Pasture Coniferous Woodland	Dyfi	A B I P	24 38 44 23
100	**Dulas North**	**2013**						**A** **B** **I** **P**	**4** **1** **13** **23**
20	Edw at Hundred house	2012	311400	254400	200	Improved pasture	Wye	A B I P	/ 113 26 14
101	**Edw at Hundred house**	**2013**						**A** **B** **I** **P**	**/** **88** **35** **14**
21	Gl1	2012	273879	246723	216	Deciduous woodland	Tywi	A B I P	52 7 7 2
102	**Gl1**	**2013**						**A*** **B*** **I** **P**	**27** **57** **24** **2**
22	Gl2	2012	272948	246591	245	Deciduous	Tywi	A B I P	50 44 84 13
103	**Gl2**	**2013**						**A** **B** **I** **P**	**9** **22** **43** **13**

Site Code	Site name	Year	E	N	Alt (m)	Land-use	Catchment	Spp	Total
23	Giar (Ceiliog)	2012	251650	240030	205	Rough Pasture	Teifi	A B I P	33 31 62 28
24	Groes - Teifi	2012	269400	259900	175	Rough Pasture	Teifi	A B I P	8 27 14 28
25	Grwyne Fawr	2012	325800	226800	374	Coniferous Woodland Deciduous Woodland	Usk	A B I P	62 40 22 /
104	**Grwyne Fawr**	**2013**						**A*** **B** **I** **P**	**28** **75** **19** **/**
26	Gwy headstream	2012	282391	285385	395	Moor/ Heathland Coniferous Woodland	Wye	A B I P	26 37 62 24
27[2]	Hafren headstream	2012	284287	287756	357	Coniferous woodland	Severn	A B I P	20 1 5 20
28	Hay Dulas	2012	324400	240600	217	Deciduous Woodland	Wye	A B I P	32 / / /
29	Hirnant	2012	299200	256900	203	Deciduous/ rough pasture	Wye	A B I P	/ 47 88 /
30	Honddu at Capel	2012	325600	231500	318	Rough Pasture	Wye	A B I P	22 52 44 /
105	**Honddu at Capel**	**2013**						**A** **B** **I*** **P**	**/** **45** **19** **/**
31	Irfon at Builth	2012	303300	251700	126	Deciduous	Wye	A B I P	/ 44 7 5
32	Ithon at Llandewi	2012	310400	268300	235	Rough Pasture	Wye	A B I P	/ 47 15 24
106	**Ithon at**	**2013**						**A**	**/**

Site Code	Site name	Year	E	N	Alt (m)	Land-use	Catchment	Spp	Total
	Llandewi							**B* I* P**	**70 23 24**
33	LI1	2012	280871	252985	303	Coniferious	Tywi	A B I P	41 / 43 20
34	LI2	2012	281060	251643	307	Coniferious	Tywi	A B I P	21 / / 38
35	LI3	2012	281467	250743	344	Coniferous Woodland	Tywi	A B I P	26 / 2 2
107	**LI3**	**2013**						**A B I P**	**55 / 9 2**
36	LI4	2012	281617	250043	335	Coniferious	Tywi	A B I P	30 3 22 29
37	LI5	2012	282015	249739	326	Rough pasture	Tywi	A B I P	80 50 17 17
38	LI6	2012	282215	249639	326	Rough Pasture	Tywi	A B I P	32 50 25 20
108	**LI6**	**2013**						**A* B I* P**	**39 70 42 20**
39	LI7	2012	281832	249233	330	Moor/ Heathland	Tywi	A B I P	125 42 28 10
109	**LI7**	**2013**						**A* B I P**	**80 70 / 10**
40	LI8	2012	280533	248866	322	Moor/ Heathland Coniferous Woodland	Tywi	A B I P	70 / / 23
110	**LI8**	**2013**						**A B**	**20 /**

Site Code	Site name	Year	E	N	Alt (m)	Land-use	Catchment	Spp	Total
								I P	**/ 23**
41	Lower Chwefru	2012	299700	253000	201	Rough Pasture Coniferous Woodland	Wye	A B I P	/ 55 22 /
42	Lower Hafren headstream	2012	284287	287756	357	Coniferous woodland	Severn	A B I P	/ 20 21 9
43	Lugg at Monaughty	2012	323700	268300	193	Rough Pasture Coniferous Woodland	Wye	A B I P	/ 54 60 /
44	Marteg	2012	295200	271500	236	Rough Pasture, Deciduous Woodland	Wye	A B I P	1 20 20 /
111	**Marteg**	**2013**						**A B I P**	**/ 48 4 /**
45	Meurig	2012	271800	267500	175	Rough Pasture	Teifi	A B I P	1 36 4 1
46	Nant Clawdd	2012	265200	246200	210	Improved pasture Deciduous Woodland	Cothi	A* B I* P	115 56 58 /
47	Nant Clwedog Uchaf	2012	264400	251600	230	Moor/ Heathland Deciduous Wooland	Teifi	A B I P	18 19 35 10
48	Nant Dar	2012	270300	243800	220	Coniferous Woodland	Cothi	A B I P	50 47 65 19
112	**Nant Dar**	**2013**						**A* B* I* P**	**38 49 29 19**
49	Nant Gelli Gethin	2012	304550	306275	288	Rough Pasture	Severn	A B I P	74 33 21 21
113	**Nant Gelli Gethin**	**2013**						**A* B* I**	**64 70 10**

Site Code	Site name	Year	E	N	Alt (m)	Land-use	Catchment	Spp	Total
								P	**21**
50	Nant Glan dwr	2012	273300	283330	285	Moor/ Heathland Rough Pasture Coniferous Woodland	Rheidol	A B I P	/ 69 52 2
114	**Nant Glan dwr**	**2013**						**A** **B** **I*** **P**	**1** **11** **48** **2**
51	Nant Helygog	2012	279110	319620	215	Moor/ Heathland/ Rough Pasture/ Coniferous Woodland	Mawddach	A B I P	20 44 30 12
115	**Nant Helygog**	**2013**						**A** **B*** **I*** **P**	**3** **20** **32** **12**
52	Nant Peiran	2012	277500	274800	280	Deciduous woodland	Ystwyth	A B I P	20 22 26 30
116	**Nant Peiran**	**2013**						**A*** **B** **I** **P**	**20** **/** **14** **30**
53	Nant Pen y Cwm	2012	305400	306200	238	Rough Pasture	Severn	A B I P	/ 47 30 37
54	Nant y foel Ddu	2012	292400	358300	381	rough pasture/ moorland	Clwyd	A B I P	/ 2 / 21
117	**Nant y foel Ddu**	**2013**						**A** **B** **I** **P**	**4** **20** **/** **21**
55	Nant y Gwryd	2012	270000	357240	210	Moor/ Heathland Rough Pasture	Conwy	A* B I P	20 69 21 2
118	**Nant y Gwryd**	**2013**						**A** **B*** **I*** **P**	**6** **20** **22** **2**
56	South Dulas	2012	291800	246900	177	Rough Pasture	Wye	A B I	/ 114 45

Site Code	Site name	Year	E	N	Alt (m)	Land-use	Catchment	Spp	Total
								P	3
119	**South Dulas**	**2013**						**A** **B** **I** **P**	**/** **50** **7** **3**
57	Upper Duhonw	2012	304400	248700	182	Rough Pasture	Wye	A B I P	/ 37 20 /
58	Upper Garth Dulas	2012	294700	253100	226	Rough Pasture	Wye	A B I P	/ 61 59 /
59	Upper Llugwy	2012	271764	358966	180	Moor/ Heathland	Conway	A* B I P	82 50 26 21
60	Wye at Scithwen	2012	308900	243700	111	Rough Pasture	Wye	A B I P	/ 114 38 3

NOTE

[1] Freshwater sites visited in 2012 were listed alphabetically by *Site name* and numbered 1 to 60 to get their *Site Code* for simplified labelling of samples. Then, 31 terrestrial sites were visited and numbered 61 to 92, and 27 freshwater sites were revisited in 2013 and numbered 93 to 119. This was to ensure all sites and years had a unique, simple site code. The terrestrial sites have not been included in ongoing analysis.

[2] Site 27 (Hafren Headstream) was an additional site due to initially sampling in slightly the wrong location when attempting to sample Site 42 (Lower Hafren Headstream).

Table A2. Table showing details of all samples barcoded using protocols in Table 2.2 and 2.3. All sequences are available on Genbank. F = sequence in the forward position only; F&R = sequenced in the forward and reverse and a consensus sequence of the two was made.

No.	Seq. ID	Species	Genbank accession number	Haplotype	F/ F&R	Note
1	109A1	*A. sulcicollis*	KU955863	H1	F	
2	109A2	*A. sulcicollis*	KU955864	H1	F	
3	96A1	*A. sulcicollis*	KU955865	H1	F	
4	96A2	*A. sulcicollis*	KU955866	H1	F	
5	108A2	*A. sulcicollis*	KU955867	H1	F	
6	104A2	*A. sulcicollis*	KU955868	H1	F	
7	107A1	*A. sulcicollis*	KU955869	H1	F	
8	95A2	*A. sulcicollis*	KU955870	H1	F	
9	94A2	*A. sulcicollis*	KU955871	H1	F	
10	93A1	*A. sulcicollis*	KU955872	H1	F	
11	93A2	*A. sulcicollis*	KU955873	H1	F	
12	113A1	*A. sulcicollis*	KU955874	H1	F	
13	113A2	*A. sulcicollis*	KU955875	H1	F	
14	112A1	*A. sulcicollis*	KU955876	H1	F	
15	112A2	*A. sulcicollis*	KU955877	H1	F	
16	108A1	*A. sulcicollis*	KU955878	H4	F	
17	116A2	*A. sulcicollis*	KU955879	H2	F	
18	95A1	*A. sulcicollis*	KU955880	H2	F	NGS
19	110A2	*A. sulcicollis*	KU955881	H2	F	
20	107A2	*A. sulcicollis*	KU955882	H2	F	
21	102A2	*A. sulcicollis*	KU955883	H5	F	
22	102A1	*A. sulcicollis*	KU955884	H6	F	
23	110A1	*A. sulcicollis*	KU955885	H7	F	
24	116A1	*A. sulcicollis*	KU955886	H8	F	
25	104A1	*A. sulcicollis*	KU955887	H9	F	
26	95A3	*A. sulcicollis*	KU955888	H1	F&R	
27	95A4	*A. sulcicollis*	KU955889	H1	F&R	
28	95A5	*A. sulcicollis*	KU955890	H3	F&R	
29	109A3	*A. sulcicollis*	KU955891	H1	F&R	
30	109A4	*A. sulcicollis*	KU955892	H1	F&R	
31	109A5	*A. sulcicollis*	KU955893	H1	F&R	
32	94A1	*A. standfussi*	KU955894		F	misidentified
33	104I2	*I. grammatica*	KU955895	H1	F	
34	101I1	*I. grammatica*	KU955896	H1	F	
35	105I1	*I. grammatica*	KU955897	H8	F	
36	101I2	*I. grammatica*	KU955898	H5	F	
37	104I1	*I. grammatica*	KU955899	H4	F	
38	105I2	*I. grammatica*	KU955900	H3	F	
39	115I2	*I. grammatica*	KU955901	H2	F	
40	106I5	*I. grammatica*	KU955902	H1	F	
41	106I4	*I. grammatica*	KU955903	H9	F	
42	93I1	*I. grammatica*	KU955904	H4	F	
43	118I2	*I. grammatica*	KU955905	H2	F	
44	98I5	*I. grammatica*	KU955906	H2	F	

No.	Seq. ID	Species	Genbank accession number	Haplotype	F/ F&R	Note
45	93I2	*I. grammatica*	KU955907	H6	F	
46	114I9	*I. grammatica*	KU955908	H1	F	
47	108I14	*I. grammatica*	KU955909	H3	F	
48	114I8	*I. grammatica*	KU955910	H1	F	
49	97I5	*I. grammatica*	KU955911	H5	F	
50	97I6	*I. grammatica*	KU955912	H1	F	
51	112I6	*I. grammatica*	KU955913	H4	F	
52	118I1	*I. grammatica*	KU955914	H2	F	
53	98I6	*I. grammatica*	KU955915	H1	F	
54	115I1	*I. grammatica*	KU955916	H10	F	
55	108I15	*I. grammatica*	KU955917	H11	F	
56	109I2	*I. grammatica*	KU955918	H2	F&R	
57	109I3	*I. grammatica*	KU955919	H7	F&R	
58	96I3	*I. grammatica*	KU955920	H1	F&R	
59	96I1	*I. grammatica*	KU955921	H3	F&R	NGS
60	96I2	*I. grammatica*	KU955922	H2	F&R	
61	109I1	*I. grammatica*	KU955923	H3	F&R	
62	102B3	*B. rhodani*	KU955924	H2	F	NGS
63	96B1	*B. rhodani*	KU955925	H1	F&R	
64	96B2	*B. rhodani*	KU955926	H1	F&R	
65	102B1	*B. rhodani*	KU955927	H2	F&R	
66	112B1	*B. rhodani*	KU955928	H2	F&R	
67	94B2	*B. rhodani*	KU955929	H2	F&R	
68	112B2	*B. rhodani*	KU955930	H3	F&R	
69	118B1	*B. rhodani*	KU955931	H2	F&R	
70	94B1	*B. rhodani*	KU955932	H2	F&R	
71	115B2	*B. rhodani*	KU955933	H1	F&R	
72	115B1	*B. rhodani*	KU955934	H12	F&R	
73	111B1	*B. rhodani*	KU955935	H4	F&R	
74	118B2	*B. rhodani*	KU955936	H2	F&R	
75	103B1	*B. rhodani*	KU955937	H1	F&R	
76	106B2	*B. rhodani*	KU955938	H1	F&R	
77	106B1	*B. rhodani*	KU955939	H1	F&R	
78	101B1	*B. rhodani*	KU955940	H13	F&R	
79	111B2	*B. rhodani*	KU955941	haplogroup 2	F&R	
80	103B2	*B. rhodani*	KU955942	haplogroup 2	F&R	
81	102B2	*B. rhodani*	KU955943	H8	F&R	
82	105B1	*B. rhodani*	KU955944	H10	F&R	
83	104B2	*B. rhodani*	KU955945	H9	F&R	
84	105B2	*B. rhodani*	KU955946	H1	F&R	
85	104B1	*B. rhodani*	KU955947	H4	F&R	
86	101B2	*B. rhodani*	KU955948	H1	F&R	
87	97B2	*B. rhodani*	KU955949	H11	F	
88	99B1	*B. rhodani*	KU955950	haplogroup 2	F	
89	99B2	*B. rhodani*	KU955951	H1	F	
90	97B1	*B. rhodani*	KU955952	haplogroup 2	F	
91	109B2	*B. rhodani*	KU955953	H6	F	
92	113B1	*B. rhodani*	KU955954	H5	F	
93	113B2	*B. rhodani*	KU955955	H5	F	

No.	Seq. ID	Species	Genbank accession number	Haplotype	F/ F&R	Note
94	109B1	*B. rhodani*	KU955956	H6	F	
95	108B1	*B. rhodani*	KU955957	H7	F	
96	119B2	*B. rhodani*	KU955958	H3	F	
97	108B2	*B. rhodani*	KU955959	H1	F	
98	119B1	*B. rhodani*	KU955960	H3	F	
99	14B1	*B. rhodani*	KU955961	haplogroup 3	F	Cryptic species
100	14B2	*B. rhodani*	KU955962	haplogroup 3	F	Cryptic species
101	14B3	*B. rhodani*	KU955963	haplogroup 3	F	Cryptic species
102	14B4	*B. rhodani*	KU955964	haplogroup 3	F	Cryptic species
103	14B5	*B. rhodani*	KU955965	haplogroup 3	F	Cryptic species
104	14B6	*B. rhodani*	KU955966	haplogroup 3	F	Cryptic species
105	14B7	*B. rhodani*	KU955967	haplogroup 3	F	Cryptic species
106	14B8	*B. rhodani*	KU955968	haplogroup 3	F	Cryptic species
107	14B9	*B. rhodani*	KU955969	haplogroup 3	F	Cryptic species
108	14B10	*B. rhodani*	KU955970	haplogroup 3	F	Cryptic species
109	14B11	*B. rhodani*	KU955971	haplogroup 3	F	Cryptic species
110	14B12	*B. rhodani*	KU955972	haplogroup 3	F	Cryptic species
111	14B13	*B. rhodani*	KU955973	haplogroup 3	F	Cryptic species
112	14B14	*B. rhodani*	KU955974	haplogroup 3	F	Cryptic species
113	14B15	*B. rhodani*	KU955975	haplogroup 3	F	Cryptic species
114	14B16	*B. rhodani*	KU955976	haplogroup 3	F	Cryptic species
115	14B17	*B. rhodani*	KU955977	haplogroup 3	F	Cryptic species
116	14B18	*B. rhodani*	KU955978	haplogroup 3	F	Cryptic species
117	14B19	*B. rhodani*	KU955979	haplogroup 3	F	Cryptic species
118	14B20	*B. rhodani*	KU955980	haplogroup 3	F	Cryptic species
119	14B21	*B. rhodani*	KU955981	haplogroup 3	F	Cryptic species
120	14B22	*B. rhodani*	KU955982	haplogroup 3	F	Cryptic species
121	14B28	*B. rhodani*	KU955983	haplogroup 3	F	Cryptic species
122	14B27	*B. rhodani*	KU955984	haplogroup 3	F	Cryptic species
123	14B23	*B. rhodani*	KU955985	haplogroup 3	F	Cryptic species
124	14B25	*B. rhodani*	KU955986	haplogroup 3	F	Cryptic species
125	14B26	*B. rhodani*	KU955987	haplogroup 3	F	Cryptic species
126	14B24	*B. rhodani*	KU955988	haplogroup 3	F	Cryptic species

Appendix B - Genomic Resource Note

Appendix B comprises of the manuscript submitted to bioRxiv (available here: Doi: http://dx.doi.org/10.1101/046227) detailing how Next Generation Sequencing (NGS) data was created and processed including full instructions of bioinformatics pipeline used to assemble NGS data for microsatellite development. Also provides links to where all the data is stored online.

Table of tables

Table B1 278

Table B2 280

Table of figures

Figure B1. 281

Author Contributions

Hannah C Macdonald	All data collection and analysis, wrote first draft and incorporated comments.
Luis Cunha	Advice and assistance through analysis of pipeline and comments on the draft.
Michael W Bruford	Advice and secured funding for NGS data generation.

Appendix B - Genomic Resource Note

Title

Development of genomic resources for four potential environmental bioindicator species: *Isoperla grammatica*, *Amphinemura sulcicollis*, *Oniscus asellus* and *Baetis rhodani*.

Authors

Hannah C Macdonald[*1], Luis Cunha[1,2] and Michael W Bruford[1]

*Corresponding author

Affiliation

[1] School of Biosciences, Cardiff University, Cardiff, UK

[2] Embrapa Florestas, Estrada da Ribeira km. 111, Colombo-PR, 83411-000, Brazil

Hannah C Macdonald

Cardiff University

The Sir Martin Evans Building

Museum Avenue

Cardiff

CF10 3AX

Email: *BurtonHC@cardiff.ac.uk*

Abstract

A low-coverage genome was generated for each of four environmental key-species of macroinvertebrate taxa for the primary purpose of microsatellite marker development. *De novo* assemblies and microsatellite markers were designed for the freshwater species *Isoperla grammatica, Amphinemura sulcicollis,* and *Baetis rhodani* but have not been completed for the common shiny woodlice *Oniscus asellus.* Here, the data is made available, and the methods and pipeline are described which led to the creation of this resource. As widespread and functionally important organisms, which are often neglected in favour of studies on vertebrates, this data will be a useful resource for further research.

Keywords: invertebrates, freshwater, terrestrail, de novo assembly

Introduction

Macroinvertebrates are widespread, often dominant and functionally important members of their environment that, coupled with their relative ease of sampling make them ideally suited for use as indicator species for biomonitoring and conservation assessment (Pfrender *et al.* 2010; Buss *et al.* 2015; Cardoni *et al.* 2015), and are recognised as such by the Water Framework Directive (2000/60/CE) (European Commission 2000). However, their use in genetic approaches is still limited, often being neglected from studies because of the lack of data (Cardoso *et al.* 2011).

Four macroinvertebrate species were sequenced for the primary purpose of developing microsatellite markers for use in population genetics; these include three freshwater invertebrates (*Amphinemura sulcicollis*, *Isoperla grammatica,* and *Baetis rhodani*) and one terrestrial soil invertebrate, the common shiny woodlice, *Oniscus asellus*. They all represent dominant, widespread species and therefore can be used as biomonitoring tools that will be effective at large spatial scales, as policy demands (Statzner and Bêche 2010). Microsatellite markers within this group are scarce, for example, within the large and diverse groups of Plecoptera and Ephemeroptera, there are only five species with between 3-13 microsatellites each, therefore this data will be a valuable and considerable resource for

future research. This data could be used for further study into these invertebrates, such as describing their mitochondrial genome (as in Stewart and Beckenbach (2006)), or studying their genome content; evolutionary analyses (e.g. divergent rates), and further investigation of their genetic features (as in Li *et al.* (2010).

Data Access

Raw data is stored in NCBI's Sequence Read Archive (SRA): NGS data for four invertebrates: Amphinemura sulcicollis, Isoperla grammatica, Baetis rhodani and Oniscus asellus (STUDY: PRJNA315680 (SRP072016)).

1. NGS sequence data (raw data sent from sequencing centres):

- *Amphinemura sulcicollis:*

SAMPLE: Amphi_NGS (SRS1349204)
EXPERIMENT: Amphi_NGS (SRX1642982)
RUN: Amphi_NGS (SRR3262386)

WTCHG_93433_274_1.fastq.gz
WTCHG_93433_274_2.fastq.gz
WTCHG_93434_274_1.fastq.gz
WTCHG_93434_274_2.fastq.gz

- *Isoperla grammatica:*

SAMPLE: Iso_NGS (SRS1351356)
EXPERIMENT: Iso_NGS (SRX1648180)
RUN: Iso_NGS (SRR3262388)

WTCHG_93433_273_1.fastq.gz
WTCHG_93433_273_2.fastq.gz
WTCHG_93434_273_1.fastq.gz
WTCHG_93434_273_2.fastq.gz

- *Baetis rhodani:*

SAMPLE: Baetis_NGS (SRS1351357)
EXPERIMENT: Baetis_NGS (SRX1648181)
RUN: Baetis_rhodani_NGS (SRR3262630)

Beatis_L3_1.fq.gz.1.gz
Beatis_L3_1.fq.gz.2.gz
Beatis_L3_2.fq.gz.1.gz
Beatis_L3_2.fq.gz.2.gz

- *Oniscus asellus:*

SAMPLE: Woodlice_NGS (SRS1351401)
EXPERIMENT: Oniscus_NGS (SRX1648318)
RUN: Oniscus_NGS (SRR3263253)

WTCHG_93433_275_1.fastq
WTCHG_93433_275_2.fastq
WTCHG_93434_275_1.fastq
WTCHG_93434_275_2.fastq

2. Each freshwater species has a CONTIG file (after *de novo* assembly) deposited at DDBJ/ENA/GenBank under the accession's listed below, any contigs under 200bp were removed.

- *Amphinemura sulcicollis:*

SUBID: SUB1394726
BioSample: SAMN04568201
Accession: LVVV00000000
Organism: Amphinemura sulcicollis Dwyryd
File name: amphi_kmer61.contig

- *Isoperla grammatica:*

SUBID: SUB1397890

BioSample: SAMN04568202

Accession: LVVW00000000

Organism: Isoperla grammatica Teifi

File name: iso_kmer61.contig

- *Baetis rhodani:*

SUBID: SUB1398024

BioSample: SAMN04568203

Accession: LVVX00000000

Organism: Baetis rhodani Tywi

File name: B_61.contig

3. Other data stored in Genbank:

- 132 Mitochondrial cytochrome c oxidase I (mtCOI) sequences for all four species (using barcoding primers from Folmer *et al.* (1994)) are available on genbank: Accession numbers KU955863-KU955994 (*Amphinemura sulcicollis* 31 sequences; *Isoperla grammatica* 29 sequences; *Baetis rhodani* 65 sequences; *Oniscus asellus* 6 sequences).
- 51 Microsatellite markers for *Isoperla grammatica*, *Amphinemura sulcicollis* and *Baetis rhodani* available on Genbank: between KR068997-KR069048 (Iso_1-18, Amp_1-21, B_1-13, respectively) and described fully in Macdonald *et al.* (2016) *in review*. Subsequent microsatellite marker development has not been completed for *Oniscus asellus*.

Meta Information

Data for the four draft genomes sequenced (Table B1) was generated at two different sequencing centres, which were compared for their cost effectiveness and yields. Libraries 1-3 (*A. sulcicollis*, *I. grammatica and O. asellus*) were sent to Oxford MRC Sequencing,

multiplexed with five other samples (eight libraries) as part of collaboration at Cardiff University. Whereas *B. rhodani* was sent at a later date to Beijing Genomic Institute (BGI), along with two other samples, these three samples were labelled by BGI and multiplexed in one lane (see Table B1 for full details). The main goals of the experiment were to develop enough genomic resources for each target species in order to retrieve enough high quality microsatellite markers.

Library

Multiple samples of each species were collected from sites around upland Wales, UK (Table B1) and stored in absolute ethanol. Genomic DNA was extracted from whole individuals using a High Pure PCR Template Preparation Kit for blood and tissue following the manufacturer's instructions (Roche Diagnostics GmbH Mannheim, Germany). All samples were treated with RNase after DNA extraction. Individual samples were identified using Sanger sequencing, with standard barcoding primers from Folmer *et al.* (1994) and by comparing the sequences with data in Genbank. To assure sample quality, quantification was assessed using a Qubit and visualised on a gel (Figure B1.). Nanodrop was used to assess contamination, where the 260/280 ratio were found to be between 1.8 and 2 and that the 260/230 ratio was between 2-2.2 across all analysed samples. The highest quantity and best quality samples were chosen; all species yielded DNA quantities required (which was 1-5µg of DNA normalized to a concentration of 50ng/µl) apart from *A. sulcicollis*, for which a vacuum concentrator had to be used. Samples showed high DNA integrity with no observed smearing on the electrophoresis gel (Figure B1.).

The samples of *A. sulcicollis*, *I. grammatica,* and *B. rhodani* were all made up of only one individual; however the *O. asellus* sample is made up of two individuals pooled. This was because allozyme loci have been used to show that two genetically distinct sub populations of *O. asellus* exist (*O. asellus* and O. *occidentails*) (Bilton *et al.* 1999) within *O. asellus,* it was thought that mixing two individuals would give the highest chance of success at developing microsatellites for the largest amount of samples. However, this meant that a *de novo* assembly could not be performed on this species.

Genomic DNA for all four samples were sent to their respective sequencing centres for library preparation (DNA was sheared, Illumina adapters were ligated, libraries were controlled for quality, normalized, pooled) and sequencing on HiSeq run (Table B1).

Processing

For each library NGS created four raw Illumina read files (two libraries, each with two pairs), which was transferred to Linux, unzipped, and the two libraries were concatenated, leaving two files of two pairs (renamed from the raw file names in section Data Access, to Amp_1.fastq & Amp_2.fastq, Iso_1.fastq & Iso_2.fastq, Woo.fastq & Woo_2.fastq, and B_1.fastq & B_2.fastq for *I. grammatica, A. sulcicollis, O. asellus and B. rhodani*, respectively, Table B1). Quality control was preformed using Trimmomatic v0.32 (Lohse M *et al.* 2012) and Musket v1.1 Musket (Yongchao Liu *et al.* 2013). Trimmomatic was used to cut adapters and other illumina-specific sequences from the reads. It was also used to remove reads of low quality and short length. In this case the threshold for quality window was set at 18 and the minimum length was 35bp (using phred33). Musket is multistage k-mer based corrector for Illumina short read data and was used to identify and remove any common Illumina errors for a higher quality *de novo* genome assembly.

For all aquatic (single sample) species SOAP *de novo* 2 (Luo *et al.* 2012) was then used to build *de novo* assemblies using short-reads. This was done in order to provide longer reads for microsatellite marker mining. Several *de novo* assemblies were run per species in order to test different Kmer values and best assembly metrics. Draft assemblies were chosen according to maximum contig and highest N50 value.

A *de novo* assembly was not attempted for the pooled sample of *O. asellus* due to the risk of chimeras, which is much higher for assemblies of mixed samples. Instead FLASH v 1.2.9 (Fast Length Adjustment of SHort reads) was used to merge paired ends creating reads of 300 bp (http://ccb.jhu.edu/software/FLASH/MANUAL [Date accessed: 02.03.16]).

PrimerPipeline (http://www.scrufster.com/primerpipeline/ [Date accessed: 02.03.16]) was then used to identify repeat regions within the data files and design forward and reverse

primers for each microsatellite. It is a windows program incorporating MISA (MIcroSAtellite identification tool, http://pgrc.ipk-gatersleben.de/misa/ [Date accessed: 02.03.16]) and Primer3 v.2.3.6 (Untergasser *et al.* 2012).

The full pipeline (including all scripts and annotations) is described in 'Appendix 1 Script_NGS' at the end of this manuscript.

Results

The NGS for all four species was very successful as the total number of reads (raw data) were very high (ranging from 71,727,142 to 123,076,504 reads), and they were of relatively high quality because the quality control sections of the pipeline did not remove too much (ranging from 0.3% and 11% of the total reads, see Table B1). *B. rhodani* data from BGI appears to be the most successful as it had the highest total reads and the lowest percentage of reads removed by quality control.

The *de novo* assemblies varied according to which kmer size was used (see Table B2) for an example of how the kmer size affected the N50 in *A. sulcicollis*). For all three species that assemblies were performed for, kmer 61 was chosen as it produced the best assemblies. *A. sulcicollis* had the best N50 at 1,543, whereas *I. grammatica* had 568, meaning that on average the *de novo* assembly for *A. sulcicollis* produced larger contigs, therefore the best assembly.

Acknowledgements

The research was funded by the NERC DURESS project (Diversity in Upland Rivers for Ecosystem Service Sustainability, NE/J01481/1) within the Biodiversity and Ecosystem Service Sustainability (BESS) Thematic Programme. HCM was funded by the Cardiff University President's Research Scholarships. Many thanks to Steve Ormerod and Hefin Jones from Cardiff University, for support and guidance.

Table B1. Details of the development of four separate libraries of macroinvertebrate using next generation sequencing.

	Library 1	**Library 2**	**Library 3**	**Library 4**
Species	*Amphinemura sulcicollis* (Stephens, 1836)	*Isoperla grammatica* (Poda, 1761)	*Oniscus asellus* (Linnaeus 1758)	*Baetis rhodani* (Pictet, 1845)
Genus	Amphinemura	Isoperla	Oniscus	Baetis
Order	Plecoptera		Isopoda	Ephemeroptera
Class	Insecta		Malacostraca	Insecta
Meta Information				
Sequencing centre	The Oxford Genomics Centre (WTCHG) / High-Throughput Genomics (Oxford, UK)			BGI (Shenzhen, China)
Platform	Illumina			
Model	HiSeq 2500, Rapid run			HiSeq 2000 (PE91)
Analysis type	DNA			
Run date	10.12.2013			13.10.2014
Library				
Strategy	Whole-genome shotgun sequencing of genomic DNA			
Shared lane	One lane (by itself)	One lane (by itself)	One lane (by itself)	One lane (shared with two other samples)
Sample type (mtDNA seq name available in genbank)	One individual (95A1)	One individual (96I1)	Two individuals mixed (71W3 and 70W1)	One individual (102B3)
Sex	Unknown			
Source of material (Taxon)	Tissue			
Sample Location#	275070E 342810N	268100E 254500N	298106E 231045N 309487E 230414N	273879E 246723N
	Catchment: Dwyryd Upland Wales, UK	Catchment: Teifi Upland Wales, UK	Brecknock Wildlife trust reserve, Upland Wales, UK	Catchment: Tywi Upland Wales, UK
Insert length	450	450	450	200
Max read length	300	300	300	200

	Library 1	**Library 2**	**Library 3**	**Library 4**
Species	*Amphinemura sulcicollis* (Stephens, 1836)	*Isoperla grammatica* (Poda, 1761)	*Oniscus asellus* (Linnaeus 1758)	*Baetis rhodani* (Pictet, 1845)
Results				
Total reads (before QC)	79,196,610	71,727,142	81,811,132	123,076,504
Total reads (after QC)	71,796,770	64,628,924	74,813,127	122,669,217
% removed	10.3	11.0	9.4	0.3
Mode of assembly	*De novo* assembly	*De novo* assembly	Flash	*De novo* assembly
Best kmer size	61	61	/	61
N50	1,543	568	/	850
Total No. of contigs	91,245	356,623	/	144,347
Total scaffold length	182,044,631	304,248,447	/	162,525,650
Longest scaffold	148,977	16,928	/	47,828

Coordinates are in EPSG:27700 - OSGB 1936 / British National Grid.

Table B2. Information on all *de novo* assemblies performed with *Amphinemura sulcicollis* data.

	Kmer_55	Kmer_61	Kmer_71	Kmer_81	Kmer_85	Kmer_91	Kmer_101
Scaffold number	88,343	91,245	89,130	87,731	88,028	87,347	92,636
In-scaffold contig number	834,812	832,323	832,173	842,707	850,059	864,994	1,170,180
Total scaffold length	184,467,063	182,044,631	180,880,141	183,600,553	181,179,087	179,755,162	172,156,150
Average scaffold length	2,088	1,995	2,029	2,092	2,058	2,057	1,858
Filled gap number	177,030	183,386	171,646	170,306	161,405	153,310	173,854
Longest scaffold	149,091	148,977	140,148	149,117	147,505	147,516	113,833
Scaffold and singleton number	606,125	580,009	605,194	626,038	645,302	671,404	950,155
Scaffold and singleton length	297,424,877	287,874,638	295,491,963	300,591,981	302,518,974	303,469,211	305,381,813
Average length	490	496	488	480	468	451	321
N50	1,550	1,543	1,482	1,525	1,466	1,448	1,141
N90	195	193	192	196	198	203	102
Weak points	0	0	0	0	0	0	0

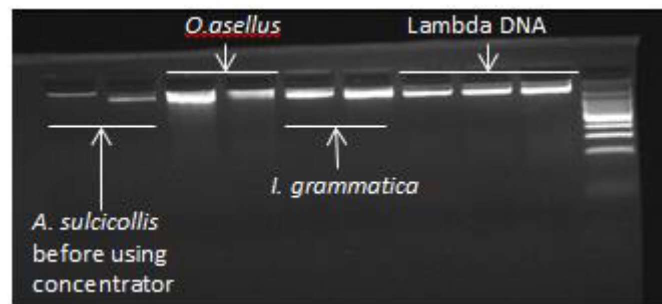


Figure B1. Photograph of a 3% ethidium-bromide stained agarose electrophoresis gel under UV light, showing genomic DNA of two individuals each species of the three species *Isoperla grammatica*, *Amphinemura sulcicollis* and *Oniscus asellus* that were sequenced first, compared to three concentrations of lambda DNA (left to right: 16.5ng/µL, 34 ng/µL and 67 ng/µL).

References

Bilton, D. T., Goode, D. and Mallet, J. (1999). Genetic differentiation and natural hybridization between two morphological forms of the common woodlouse, Oniscus asellus Linnaeus 1758. *Heredity* **82:**462-469.

Buss, D. F., Carlisle, D. M., Chon, T. S., Culp, J., Harding, J. S., Keizer-Vlek, H. E., Robinson, W. A., Strachan, S., Thirion, C. and Hughes, R. M. (2015). Stream biomonitoring using macroinvertebrates around the globe: a comparison of large-scale programs. *Environmental monitoring and assessment* **187:**4132.

Cardoni, S., Tenchini, R., Ficulle, I., Piredda, R., Simeone, M. C. and Belfiore, C. (2015). DNA barcode assessment of Mediterranean mayflies (Ephemeroptera), benchmark data for a regional reference library for rapid biomonitoring of freshwaters. *Biochemical Systematics and Ecology* **62:**36-50.

Cardoso, P., Erwin, T. L., Borges, P. A. V. and New, T. R. (2011). The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation* **144:**2647-2655.

European Commission. (2000). Directive 2000/60/EC of the European parliament and of the council of 23 october 2000 establishing a framework for community action in the field of water policy. *Official Journal of the European Union* **327:**1-72.

Folmer, O., Black, M., Hoeh, W., Lutz, R. and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular marine biology and biotechnology* **3:**294-299.

Lohse M, Bolger AM, Nagel A, Fernie AR, Lunn JE, Stitt M and B., U. (2012). RobiNA: a user-friendly, integrated software solution for RNA-Seq-based transcriptomics. Nucleic Acids Res.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y. *et al.* (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1:**18.

Macdonald, H.C,. Ormerod, S.J. and Bruford, M.W. (2016). Developing genetic capabilities for freshwater conservation using novel microsatellite markers for three stream invertebrates. *Aquatic Conservation: Marine and Freshwater Ecosystems. (In Review)*.

Pfrender, M. E., Hawkins, C. P., Bagley, M., Courtney, G. W., Creutzburg, B. R., Epler, J. H., Fend, S., Schindel, D., Ferrington, L. C., Jr., Hartzell, P. L. *et al.* (2010). Assessing macroinvertebrate biodiversity in freshwater ecosystems: advances and challenges in DNA-based approaches. *The Quarterly Review of Biology* **85:**319-340.

Statzner, B. and Bêche, L. A. (2010). Can biological invertebrate traits resolve effects of multiple stressors on running water ecosystems? *Freshwater Biology* **55:**80-119.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B., Remm, M., Rozen, S. and (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research* **1;40(15):e115**.

Yongchao Liu, Schroeder, J. and Schmidt, B. (2013). Musket: a multistage k-mer spectrum based error corrector for Illumina sequence data. *Bioinformatics* **29:**308-315.

Li, R., Fan, W., Tian, G., Zhu, H., He, L., Cai, J., Huang, Q., Cai, Q., Li, B., Bai, Y. *et al.* (2010). The sequence and *de novo* assembly of the giant panda genome. *Nature* **463:**311-317.

Stewart, J. B. and Beckenbach, A. T. (2006). Insect mitochondrial genomics 2: the complete mitochondrial genome sequence of a giant stonefly, Pteronarcys princeps, asymmetric directional mutation bias, and conserved plecopteran A+T-region elements. *Genome* **49:**815-824.

Appendix 1: NGS script

Key: `Commands in red`, annotations and instructions in black

1.0. Check and unzip

#To check data files are unaffected from download/upload use md5sum to check the unique identity of the file (compare the md5 number, it has to be exactly the same format)

```
md5sum file1 > file1_md5.txt
```

#And check the first line of the sequence

```
head filename
```

#E.g. md5sum formats for *B. rhodani* raw data files:

#aa278e7dd0de7af2e12aaf0d4ba9fc97 Beatis_L3_1.fq.gz.cut/Beatis_L3_1.fq.gz.1.gz

#1ba32f3d3a72be4877311163ce07ddc6 Beatis_L3_1.fq.gz.cut/Beatis_L3_1.fq.gz.2.gz

#26eaa5697b1c950aba2d83095f143f0c Beatis_L3_2.fq.gz.cut/Beatis_L3_2.fq.gz.1.gz

#f7fee1ecf2b928cf4504e6d48f417636 Beatis_L3_2.fq.gz.cut/Beatis_L3_2.fq.gz.2.gz

#Check per permissions, the following code changes the permissions of the file called et_trimmer.pl.

```
chmod 777 est_trimmer.pl
```

#Unzipping

#E.g. Rawdata files end in “gz” so they need to be unzipped:

#WTCHG_93433_274_2.fastq.gz

#WTCHG_93433_274_1.fastq.gz

#WTCHG_93434_274_1.fastq.gz

#WTCHG_93434_274_2.fastq.gz

#The following command will unzip everything ending in .gz in the background. The above file goes from a fastq.gz file to just a fastq file.

```
gunzip *.gz &
```

#For unzipping program files e.g. musket. If ends in tar.bz the command is:

```
tar -xvjf
```

#If ends in tar.gz the command is:

```
tar -xvzf
```

2.0. Concatenate

#I had two libraries with a forward and reverse, put the two forwards into one file and the two backs in one file, just to make it simpler. The two 1's together and the two 2's.

#The following command tells Linux to concatenate the files called 'WTCHG_93433_273_1.fastq' and 'WTCHG_93434_273_1.fastq', and name the combined file iso_1.fastq, and do it all in the background (&). Note, you have to be in the directory that the files are in or tell Linux where to find them e.g. home/c1135170/Hannah/

```
cat WTCHG_93433_273_1.fastq WTCHG_93434_273_1.fastq >
iso_1.fastq &
```

#Do the same with the other pair

```
cat WTCHG_93433_273_2.fastq WTCHG_93434_273_2.fastq >
iso_2.fastq &
```

#E.g with *B. rhodani*:

```
cat Beatis_L3_1.fq.gz.1 Beatis_L3_1.fq.gz.2 > B_1.fastq &

cat Beatis_L3_2.fq.gz.1 Beatis_L3_2.fq.gz.2 > B_2.fastq &
```

3.0. Trimmomatic

#First need to download Trimmomatic, you can find here:
http://www.usadellab.org/cms/?page=trimmomatic

#Right click, 'copy link address' for the Binary.

```
wget
```

#right click to paste:

```
http://www.usadellab.org/cms/uploads/supplementary/Trimmomatic
/Trimmomatic-0.32.zip
```

#It downloads to the folder you're in and it's called Trimmomatic-0.32.zip

```
unzip Trimmomatic-0.32.zip
```

#now just called Trimmomatic-0.32

#Now ready to run Trimmomatic

#Code means: Using Trimmomatic PE (paired end) which can be found here (pathway) do "–phred33" to these two files (B_1.fastq & B_2.fastq) then rename them trimmomatic_B_1.fastq.gz (for paired) and trimmomatic_B_1_unpaired.fastq.gz for unpaired, and the same with the other pair. The nohup at the beginning is there so I can close the window and it will still carry on running.

```
nohup java -classpath /home/c1135170/Hannah/app/Trimmomatic-
0.32/trimmomatic-0.32.jar
org.usadellab.trimmomatic.TrimmomaticPE -phred33 B_1.fastq
B_2.fastq trimmomatic_B_1.fastq.gz
trimmomatic_B_1_unpaired.fastq.gz trimmomatic_B_2.fastq.gz
trimmomatic_B_2_unpaired.fastq.gz
ILLUMINACLIP:/home/c1135170/Hannah/app/Trimmomatic-
0.32/adapters/TruSeq2-PE.fa:2:30:10 LEADING:3 TRAILING:3
SLIDINGWINDOW:4:18 MINLEN:35 &
```

#E.g. Nohup.txt

#Read Pairs: 61528668 Both Surviving: 61132622 (99.36%) Forward Only Surviving: 96495 (0.16%) Reverse Only Surviving: 298538 (0.49%) Dropped: 1013 (0.00%)

4.0 Musket

#Have to download Musket the same way as Trimmomatic, you'll find musket here:

http://musket.sourceforge.net/homepage.htm#latest

#Version used musket-1.1, right click, copy link address as with Trimmomatic:

```
wget
```

#right click to paste

```
http://sourceforge.net/projects/musket/files/musket-1.1.tar.bz
```

#When pasting link may have "/download" on the end of the link, delete this before pressing enter.

#Unzip file (see section 1.0)

#To install, go to the Musket folder and press 'make':

```
cd ../app/cd musket-1.1/ls
```

```
make
```

#Then you can remove the original zipped file

```
rm musket-1.1.tar.bz
```

#2.0. Run Musket using file outputs from Trimmomatic:

```
nohup /home/c1135170/Hannah/app/musket-1.1/musket -k 21
2192141955 -p 32 -omulti corrected -inorder
trimmomatic_B_1.fastq.gz trimmomatic_B_2.fastq.gz
trimmomatic_B_1_unpaired.fastq.gz
trimmomatic_B_2_unpaired.fastq.gz 1>out.txt 2>error.txt
```

#They are named corrected.0, corrected.1, corrected.2, and corrected.3, in order of how the files were listed in the command above. Rename the files:

```
mv corrected.0 musket_B_1.fastq

mv corrected.1 musket_B_2.fastq

mv corrected.2 musket_B_1_unpaired.fastq

mv corrected.3 musket_B_2_unpaired.fastq
```

5.0. FLASH (used for O. asellus only)

#To find Flash: http://ccb.jhu.edu/software/FLASH/

#To download:

```
wget http://sourceforge.net/projects/flashpage/files/FLASH-1.2.9.tar.gz
```

#To unzip

```
tar -zxvf FLASH-1.2.9.tar.gz
```

Command asks flash to merge paired ends <musket_woo_1> < musket_woo_1> [-m minOverlap - varied] [-M maxOverlap- 100] [-x mismatchRatio-varied] [-p phredOffset] [-o prefixOfOutputFiles] [-d pathToDirectoryForOutputFiles] [-f averageFragment Length- 300] [-s standardDeviationOfFragments- varied] [-r averageReadLength- 150].

#Several different compbinations tried to

```
../app/FLASH-1.2.9/flash musket_woo_1.fastq musket_woo_2.fastq
-m 15 -M 100 -x 0.1 -p -o merged -d -f 300 -s  50 -r 150
1>flash.out 2>flash.err &
```

#10% retained and matched

```
../app/FLASH-1.2.9/flash musket_woo_1.fastq musket_woo_2.fastq
-m 10 -M 100 -x 0.1 -o merged2 -d -f 300 -s  40 -r 150
1>flash2.out 2>flash2.err &
```

#15% retained and matched

```
../app/FLASH-1.2.9/flash musket_woo_1.fastq musket_woo_2.fastq
-m 25 -M 100 -x 0.1 -o merged3 -d -f 300 -s  40 -r 150
1>flash3.out 2>flash3.err &
```

#9.5% retained and matched

```
../app/FLASH-1.2.9/flash musket_woo_1.fastq musket_woo_2.fastq
-m 20 -M 100 -x 1 -o merged4 -d -f 300 -s  40 -r 150
1>flash4.out 2>flash4.err &
```

#100% retained and matched

```
../app/FLASH-1.2.9/flash musket_woo_1.fastq musket_woo_2.fastq
-m 20 -M 100 -o merged4 -d -f 300 -s  40 -r 150 1>flash4.out
2>flash5.err &
```

#mismatchRatio: default 0.25 . 12.78% retained and matched

#Use merged extendedfrags to feed straight into MISA

#To continue must convert fastq file to fasta file, and removes spaces at the same time, using:

```
awk 'BEGIN{a=0}{if(a==1){print;a=0}}/^@/{print;a=1}'
myFastqFile | sed 's/^@/>/' > myfastafile
```

6.0. SOAPdenovo2

#Download soapdenovo, (same way as Trimmomatic and Musket), find here:

http://sourceforge.net/projects/soapdenovo2/files/SOAPdenovo2/

#To download:

```
wget
http://sourceforge.net/projects/soapdenovo2/files/latest/downl
oad?source=files
```

#Unzip

```
tar -xvvf SOAPdenovo2-src-r240-4.tar
```

#Compile by navigating to the folder that the 'makefile' is in and type:

```
make
```

#First make config file to use with soap de novo, **green** needs to change depending on the data, especially the pathways to the musket output files so soap de novo knows where the files are. In this example the config file was named 'iso_config.text'.

max_rd_len=150

[LIB]

#average insert size

avg_ins=300

#if sequence needs to be reversed

reverse_seq=0

#in which part(s) the reads are used

asm_flags=3

#in which order the reads are used while scaffolding

rank=1

#a pair of fastq file, read 1 file should always be followed by read 2 file

q1=**/home/c1135170/Hannah/Isoplera/musket_iso_1.fastq**

q2=**/home/c1135170/Hannah/Isoplera/musket_iso_2.fastq**

q=**/home/c1135170/Hannah/Isoplera/musket_iso_1_unpaired.fastq**

q=**/home/c1135170/Hannah/Isoplera/musket_iso_2_unpaired.fastq**

#For kmers less than 63, use the following command, be in the same folder as the config file:

```
nohup ../app/soapdenovo/SOAPdenovo2-src-r240/SOAPdenovo-63mer
all -s iso_config.txt -K 55 -R -o iso_kmer55 1>iso_kmer55.log
2>iso_kmer55.err
```

Command explained: `nohup ../app/soapdenovo/SOAPdenovo2-src-r240/SOAPdenovo-63mer` (telling it where to find soapdenovo) `all -s iso_config.text` (name of the config file we made) `-K 55` (kmer size 61)`-R -o iso_kmer55 1>iso_kmer55.log 2>iso_kmer55.err` (names of the output files)

#For Kmers above 63:

```
nohup /home/c1135170/Hannah/app/SOAPdenovo2-bin-LINUX-generic-
r240/SOAPdenovo-127mer all -s iso_config.txt -K 71 -p 10 -R -o
iso_kmer71 1>kmer71.log 2>kmer71.err
```

Appendix C - PrimerPipeline Instruction Manual

Appendix C consists of the Instruction manual for the software *PrimerPipeline* (available at http://www.scrufster.com/primerpipeline), developed in collaboration with computer programmer Greg Macdonald. *PrimerPipeline* is a user-friendly program that finds microsatellites, designs primers from NGS data and displays results in a clear, informative way.

Table of figures

Figure C1 297
Figure C2 297
Figure C3. 300
Figure C4 301
Figure C5 302
Figure C6. 302
Figure C7 303
Figure C8 304
Figure C9 309
Figure C10 310
Figure C11. 312
Figure C12 312
Figure C13 313
Figure C14 313
Figure C15 314

PrimerPipeline

Instruction Manual v1.0

14/07/2015

Author Contributions

Hannah Macdonald	Concept formulating, all testing, author of manual
Greg Macdonald	Computer programming and GUI design

1.0 Getting started

1.1 Operating System: Windows 7 or later

No installation required. All you need to do is download the *PrimerPipeline.zip* folder (available via the website here) and *Extract all* the contents together to a folder anywhere on your computer. It is not recommended to do this in the root of your C drive, or in Program files, as this will require you to run *PrimerPipeline* as an administrator in order for it to write files.

If required, make a shortcut to *PrimerPipeline.exe*.

Depending on your computer's security settings, the first time you run *PrimerPipeline* you might have to click *Run Anyway*, because it won't be recognised.

1.2 Operating System: Windows XP

If you are using Windows XP, you can still use *PrimerPipeline* as above, but in addition you have to ensure you have 'Microsoft .NET Framework 4.0' installed.

To find out whether you have it already, go to:

Start > *Control Panel* > *Add/Remove Programs*

And look for 'Microsoft .NET Framework 4.0'. If you do not have it, you can download it from Microsoft here.

1.3 Contents of PrimerPipeline.zip

Contains the following 5 items:

- primer3_config (Folder containing setting files used by Primer3)
- primer3_core.exe (This is Primer3, *PrimerPipeline* runs this program in the background during the Primer3 stage of the process)

- PrimerPipeline.exe (This is the *PrimerPipeline* program itself)
- PrimerPipeline_Instruction_manul_v1.0.pdf (The document you're reading)
- example.fasta (An example fasta file of sequence data)

The first three files all need to be in the same folder for *PrimerPipeline* to work. The manual and the example file can be moved or removed if required.

Tip: The example.fasta file can be used for a practise run. The file contains 50,000 sequences, and under default settings *PrimerPipeline* finds 59 microsatellites, and designed primers around 8 of them correctly. This takes roughly 30 seconds (depending

2.0 Using PrimerPipeline

2.1 Input File

Tip: We recommend that the input file has already gone through Quality Control (to ensure ambiguous bases and errors are removed, and which program you use to do this depends on what NGS platform you have used, e.g. Musket removes common Illumina errors, available here). Ideally, also perform a *de novo* assembly (for example, with SOAPdenovo, available here). This assures that your sequences are good quality so the results can be trusted, and preforming a *de novo* assembly assemblies your sequences into larger contigs to aid microsatellite mining; the larger the sequence length, the

Format: Fasta file

Location: The file can be anywhere on your computer, but be aware that the output files from *PrimerPipeline* (see section 4) will be placed in same location as the input file.

Note 1: Make sure output files are not overwritten by mistake

If you repeat the pipeline with the same input file, make sure you either rename the input file so that the output files have different names (see section 4) or make a new folder to avoid

Example of fasta file format:

```
>2080906_length_274_cvg_7.0_tip_0
ATTAGGTTTCAAGAAATAGTCTTCGCACGCACACAATGATATTCTATTCTTGTCACGCAAAT
GCCTTGTGGAATGCCTTTAGACCCCTGTCAAAGTTTTCTTTTTGGTTTTGCTAATACAAAAT
ATTTGAAAATGCCCACAAAAATAACACTGGGGCACAGTATTAAAATATAATGGAGTCTTCGG
AAGCAAGTTTTTTTTTTTATTAAAATGATACTCCAAAGGCAGAAAACAAAGAAAATATAGCA
CACAAATGTTTAAAAATACAATTTGA

>2080908_length_274_cvg_10.0_tip_0
ATTCAATCATGCAGCAAAGTGGGAGAAGAAGCAGCCGCCTCCAGGAATACATTAGACACGCA
CCGTTGCGTCTCGTGCGTGTTGCAAACTTTTCCACTCCTCTTTTCCATTCCACGTCAAAAGT
ATTAAATTTTGTATTCGGAAAGTGGCAAAAATTTAAATTTAATCCCATTTATCCTTTTGTAG
TTGGGCCCCCAGTGTGCGCATTGGACGTATGGAAATGTAAAGTAGCGTCAATGTAGCATGAA
AACAAGCACTTGAGCCGAACACTTGC
```

See example.fasta file.

2.2 Open Input File in PrimerPipeline

- Open *PrimerPipeline.exe* by double clicking (Figure C1).
- Click the *Open files* button and find the file containing the sequences you wish to run *PrimerPipeline* on. Or drag and drop the file into the box.
- The status should now say ‘Ready to process’ (Figure C2).

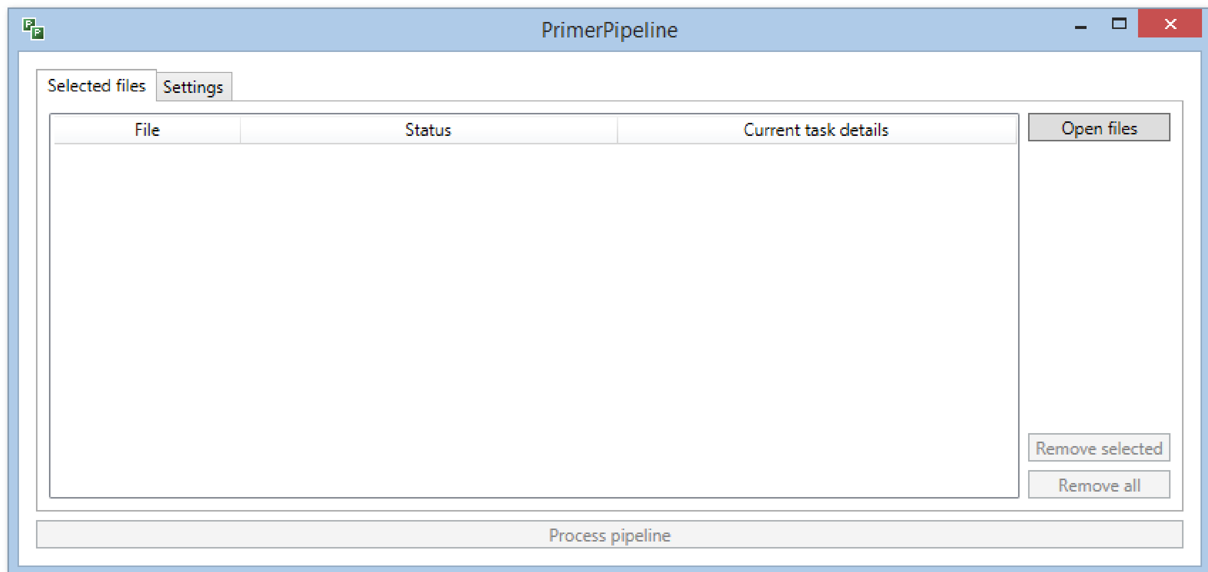


Figure C1. How *PrimerPipeline* looks when opened.

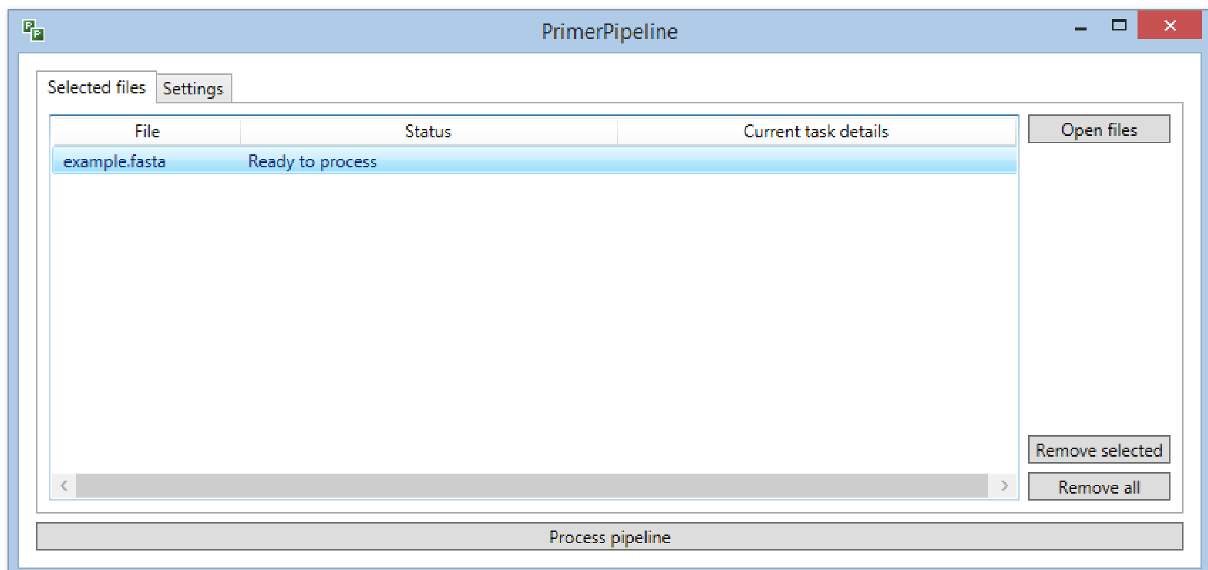


Figure C2. *PrimerPipeline* with input file loaded.

3.0 Editing Settings

Select the file you wish to run, (it will be highlighted in blue as in Figure C2) and then go to the *Settings* tab (Figure C3).

Note 2: PrimerPipeline remembers previous settings.

The first time you open *PrimerPipeline* it will load the default settings for each stage. After the first time, *PrimerPipeline* will load the settings used in the previous run.

It does this by saving a copy of the three settings files (Trim.Settings.txt, misa.ini, and primer3_v1_1_4_default_settings.txt) in the same location as '**PrimerPipeline.exe**' and replaces these files after each run). If you want to return to default settings use the *Reset to default* button in each of the three setting tabs. Also *PrimerPipeline* will return to default settings if these setting files are removed or deleted.

3.1 Sequence Trimming

- Within the *Settings* tab select *Sequence trimming* on the left hand side (Figure C3).
- To edit settings click *Edit argument*, a small window will open where you can change any of the settings (Figure C4).
- If you want to add anything to the trimming stage you can add arguments by clicking *Add argument*.
- Click *Reset to default* to return to default settings at any point. Default Settings are shown in Figure C3.
- Settings can be saved without running *PrimerPipeline* using *Save settings* and load previous settings files using *Load settings*.

Tip: *PrimerPipeline* saves settings files automatically if you run the pipeline, use *Save settings* in any of the three different settings windows (Sequence Trimming, MISA and Primer3) when you're not planning to run *PrimerPipeline* but want so save settings for later or to send to a colleague. The ability to load settings files for all sections makes it easier to: 1) repeat what a colleague has done; 2) make your runs 100% repeatable, and 3) allows you to hone your settings gradually. For example, load a settings file from a

Note 3: Further information on Trimming Settings

PrimerPipeline assumes Quality Control (QC) has already been performed however has some QC during this step.

1. 'Ambiguous trim. Number of bases =2, Window size = 200.'

Searches and removes sequences with more than 2 N's (unknown bases) in a row, it searches 200 base pairs (bp) at a time. You can edit the number of N's allowed and the window size. Increasing these numbers makes the trimming more relaxed, decreasing makes it stricter.

2. 'Remove stretches of Type A from 5' end. Minimum accepted repeat = 5, window size = 200.'

3. 'Remove stretches of Type A from 3' end. Minimum accepted repeat = 5, window size = 200.'

An artefact of some sequencing (usually older types) is stretches of one bp repeated (in default settings this is A) at the beginning (5') and end (3') of the sequence. You can edit the number of accepted repeats and the window length, again, increasing these numbers makes the trimming more relaxed, and decreasing makes it stricter.

You can add Arguments of Type G, C and T, to search for repeats of these bases, whether this is necessary depends on the type of sequencing used to acquire the data, and on the QC performed prior to using *PrimerPipeline*.

4. Cut off. Minimum value = 500, maximum sequence size = 50700.

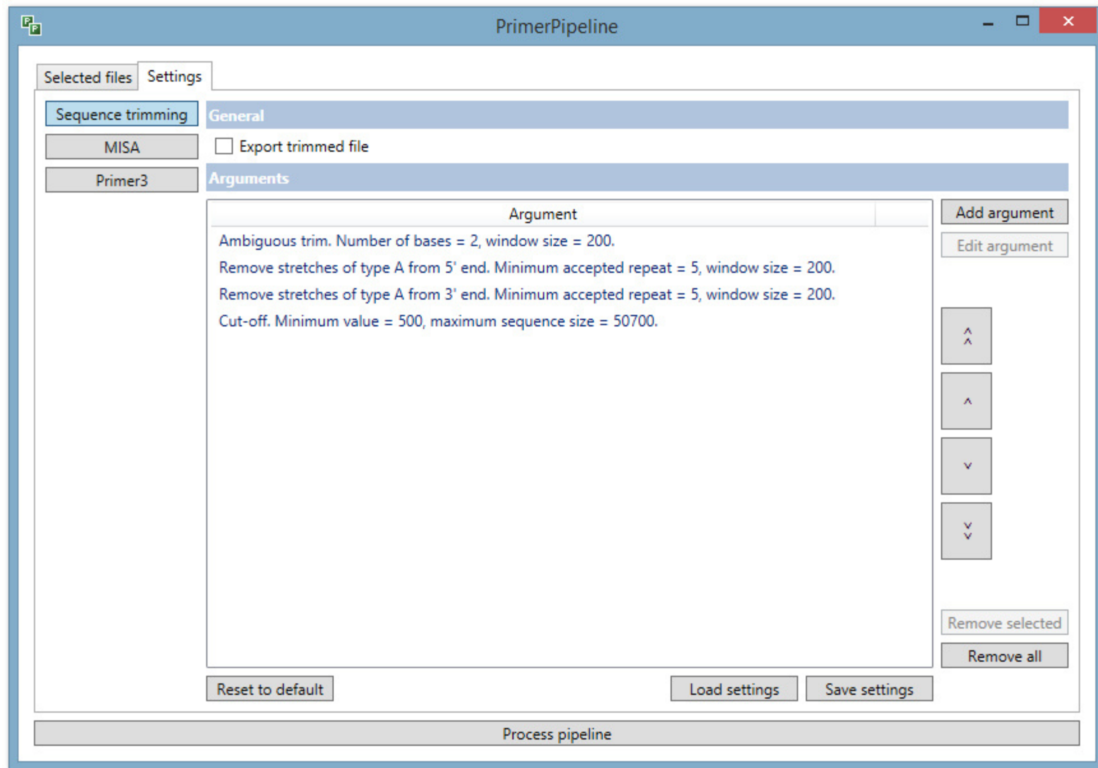


Figure C3. Settings Tab showing *Sequence trimming* selected.

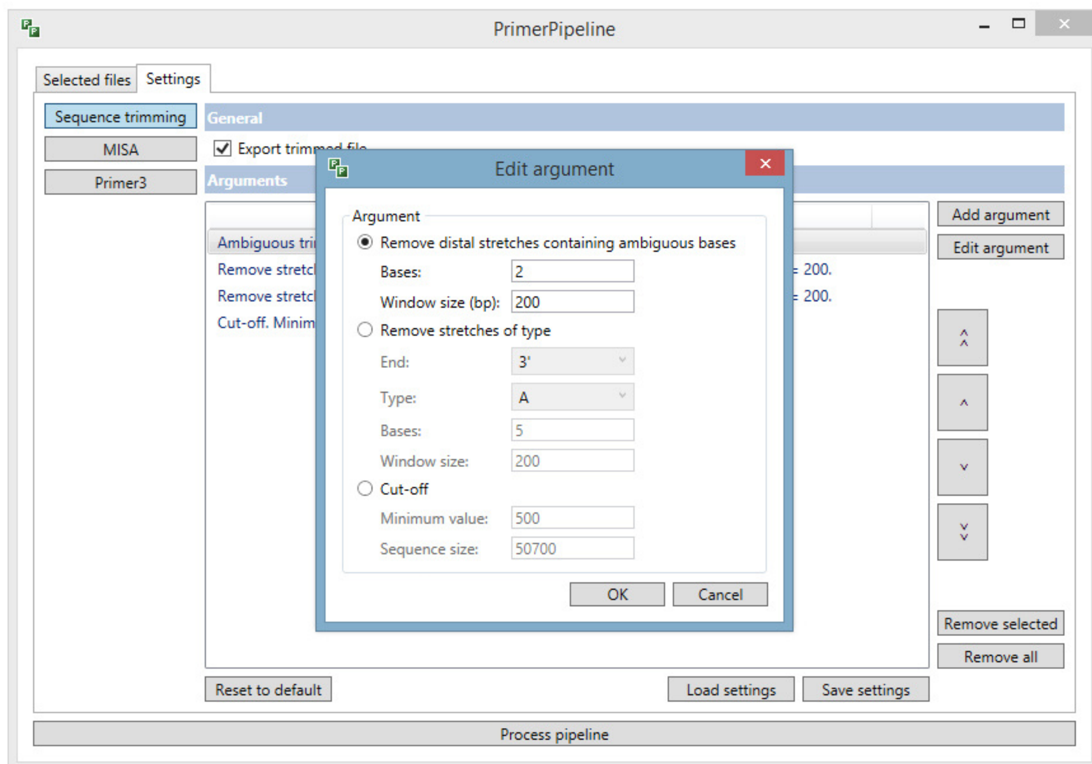


Figure C4. Sequence trimming *Settings* tab, with *Edit argument* clicked so you can edit all trimming parameters.

3.2 MISA

- Within the *Settings* tab, Select *MISA* on the left hand side (Figure C5).
- Can edit 'Interruptions' and SSR minimum repeats. Select one 'Unit size' at a time and then click *Edit definition* to change the minimum repeats.
- Press the *Reset to default* to return to default settings at any point. Default settings are shown in Figure C5.
- Again, settings can be saved using *Save settings* and settings files loaded using *Load settings*.

Note 4: Further information on MISA Settings

1. 'Interruptions: 0'

This value means the maximum difference between two SSRs allowed, i.e. the number of base pairs between SSRs. Default value 0 will allow no interruptions within a microsatellite, however if you change this setting, MISA will then find compound microsatellites e.g. Interruptions: 1 = $(AAG)_7A(AAG)_{24}$.

2. 'Unit Size – Minimum repeats'

'Unit Size' refers to the SSR/microsatellite type, e.g Unit size 2 = $(CA)_{23,}$ dinucleotide.

Unit size 3 = $(TGA)_{12}$ trinucleotide ect.

The minimum repeats refers to the length of the microsatellite, for example, 'Unit size 2,

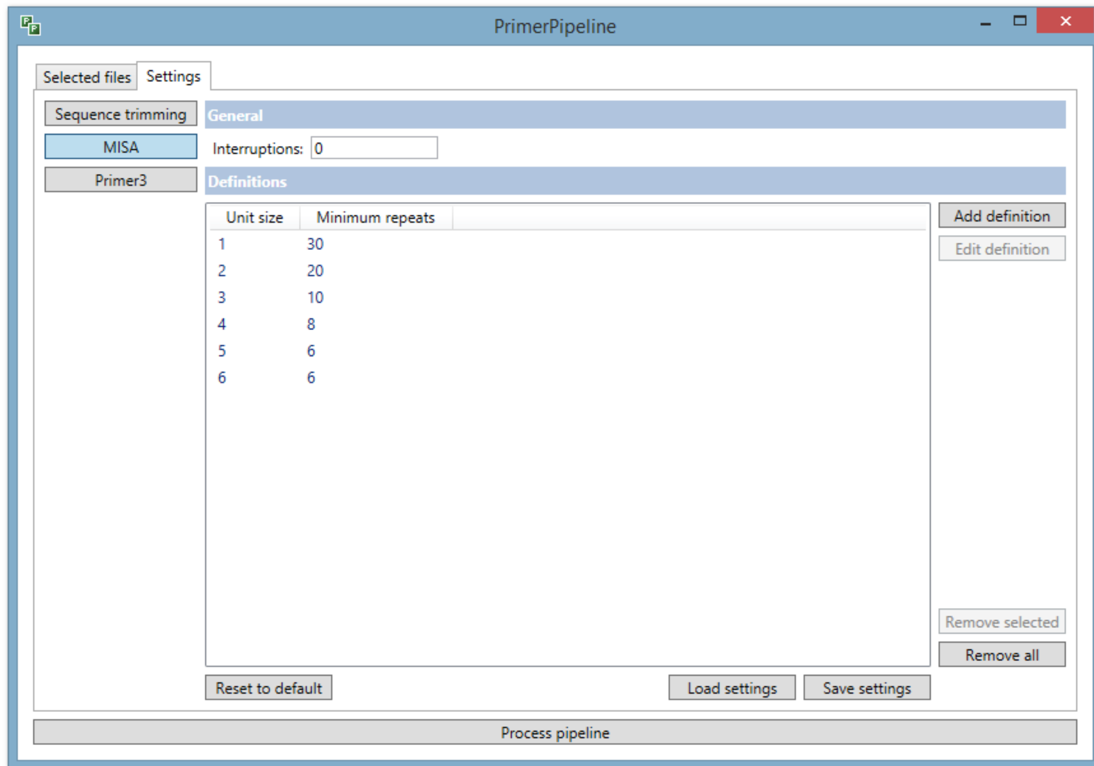


Figure C5. *Settings* tab, with *MISA* selected, allowing you to edit MISA settings.

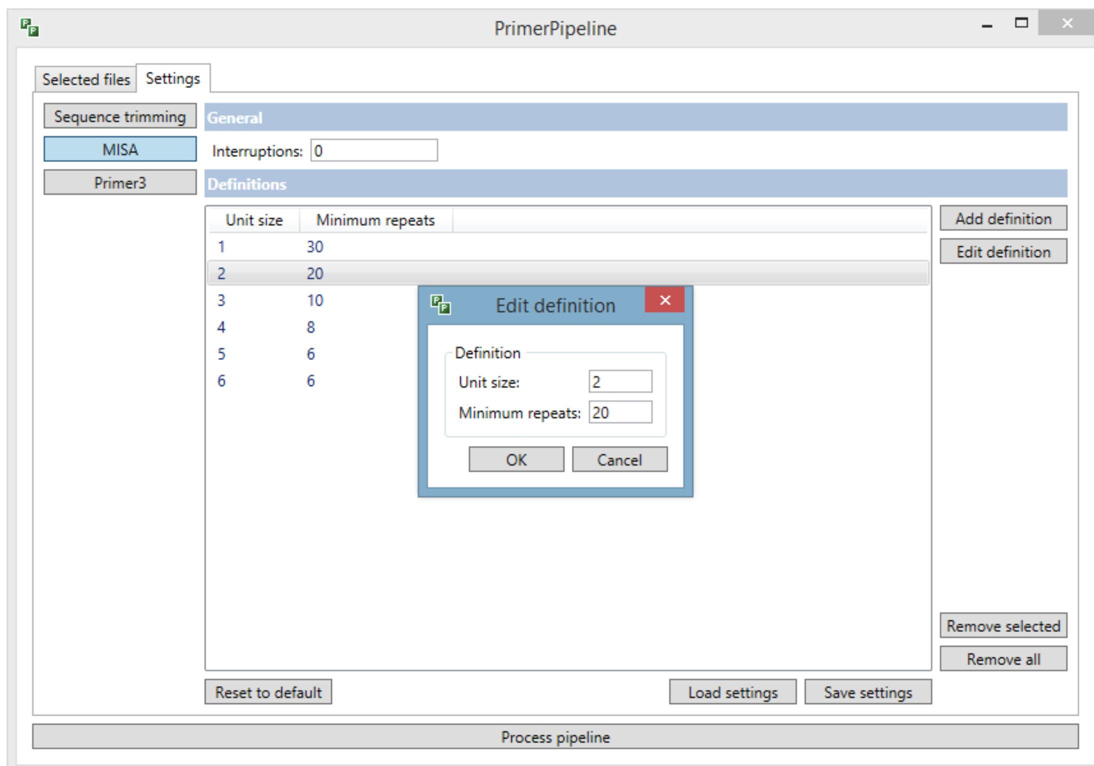


Figure C6. MISA settings tab, with *Edit definition* clicked so you can edit that unit size minimum repeats.

3.3 Primer3

- Within the *Settings* tab, Select *Primer3* on the left hand side (Figure C7)
- Can edit any of the Primer3 settings directly on this tab
- Press the *Reset to default* to return to default settings at any point, some of which are shown in Figure C7.
- If you wish to see the advanced settings Tick the box at the bottom 'Include advanced settings' and the advanced settings will be added to the list in grey (Figure C8).
- Again, settings can be saved using *Save settings* and settings files loaded using *Load settings*.

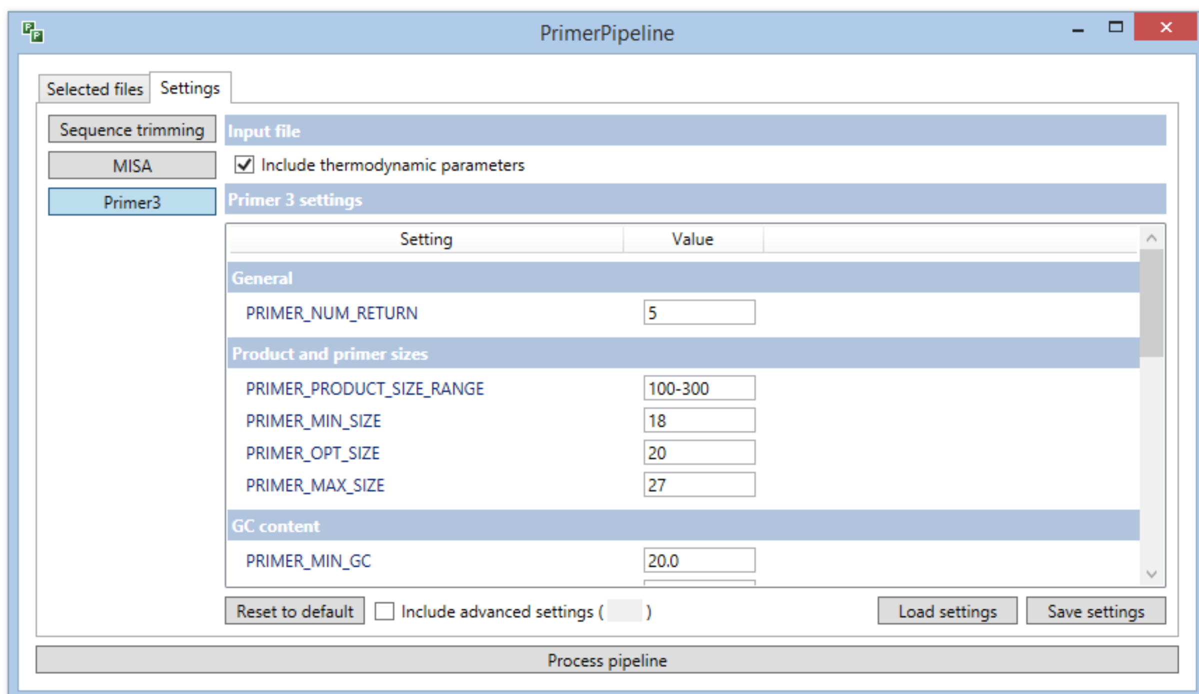


Figure C7. *Settings* tab, with *Primer3* selected, allowing you to edit Primer3 settings (without advanced settings displayed).

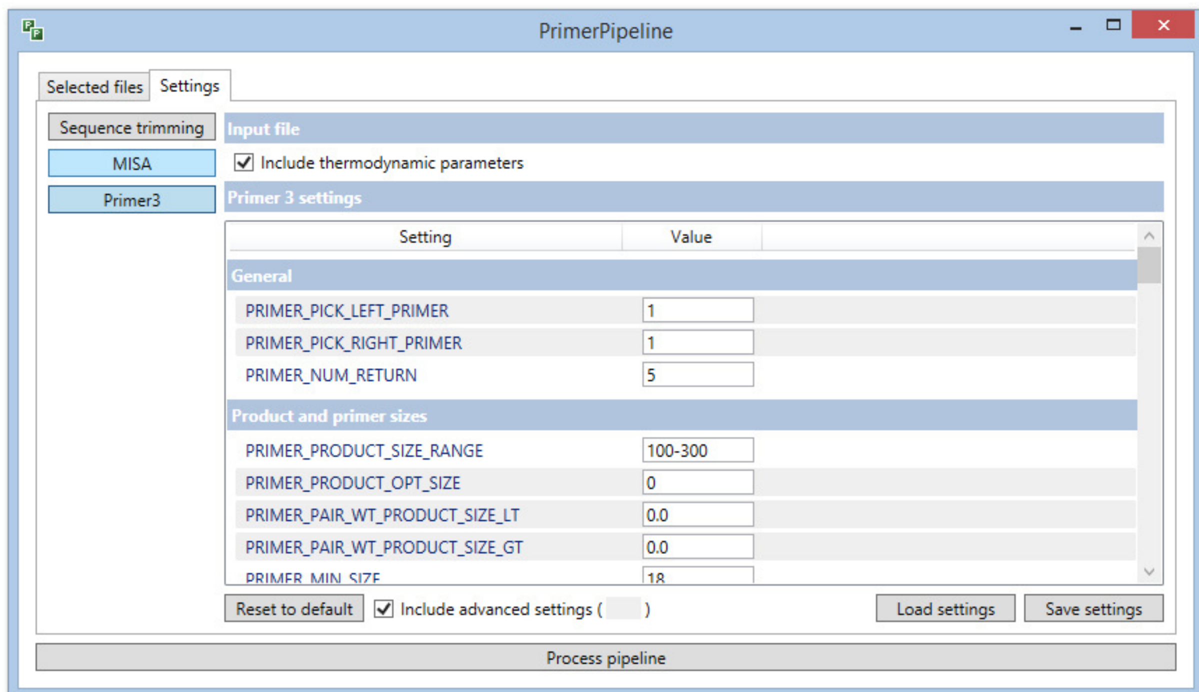


Figure C8. Settings tab, with Primer3 selected, allowing you to edit Primer3 settings (with advanced settings displayed).

Note 5: Further information on Primer3 Settings

All setting names in *PrimerPipeline* have been kept consistent with Primer3 so you can refer to Primer3's website here for more information.

There are too many settings in Primer3 to explain them all fully in this manual, however, for all settings explanations are available within *PrimerPipeline*. If you hover over the setting name a tip box will appear with a short description of the setting, however some are too long to present fully in this way, so if you read 'See Primer3 website for more details', then refer to the Primer3 website for a longer explanation.

For simple navigation of Primer3 settings, they have been gathered into seven groups:

'General'
'Product and primer sizes'
'GC content'
'Melting temperature'
'PCR conditions' (advanced only)
'Self-binding (primer-dimer and hairpins)'
'PolyX and other'

'Include thermodynamic parameters' tick box at the top. This tick box specifies the path to the directory that contains all the parameter files used by the thermodynamic approach ('PRIMER_THERMODYNAMIC_PARAMETERS_PATH' in Primer3). If this is unticked, Primer3 will not be able to locate the files it needs.

'Include advanced settings' tick box at the bottom. In order to make PrimerPipeline even more user-friendly, many of the Primer3 settings that will not be needed by the average user have been placed in the advanced settings. However, if you wish to change more settings, click 'advanced' at the bottom for the full functionality of Primer3.

Tip: There is a lot of advice online about designing primers which you can read; however here are some of my suggestions of settings to change to design better quality primers:

1. 'PRIMER_MIN_SIZE', 'PRIMER_OPT_SIZE' and 'PRIMER_MAX_SIZE'

   I advise primers to be between 18-30 bases, optimum 20, this will make sure the primers are specific and increase amplification success.

2. 'PRIMER_MIN_GC', 'PRIMER_OPT_GC_PERCENT' and 'PRIMER_MAX_GC'

   I recommend that the primer GC% be between 40-60%, with the optimum 50%. This is due to fact that there are three hydrogen bonds between G and C, whereas there is only two between A and T. This means the bonds between G and C are stronger, and a good percentage of C's and G's across your primer will ensure the primers will bind affectively and increase amplification success.

3. 'PRIMER_GC_CLAMP'

   For the same reason as above, it is good to have a GC clamp at the end of your primer, meaning the primer ends in a G or C (3' end). Default is 0, meaning not all primers will have one, you can change this to 1 if you want a GC clamp.

4. 'PRIMER_PAIR_MAX_DIFF_TM'

   This is the maximum acceptable difference between the melting temperatures of the left and right primers. Default is 100, however for PCR it is better that the annealing temperature of your primers to be similar. I advise you to change this to 2.

5. 'PRIMER_MAX_POLY_X'

   This is the maximum allowable length of a mononucleotide repeat within a primer, for example AAAAAA. I advise a maximum repeat of 3. This reduces the chance of slippage occurring during annealing step of the PCR reaction.

4.0 Output Files from PrimerPipeline

During each run, *PrimerPipeline* creates nine files. These files will be placed in the same location as the original input file.

Files are named after the original input file, for example, when using the input file name example.fasta the files will be named as follows:

4.1 Trimming

1. example_TrimSettings.txt (The trimming settings file so that for every run of *PrimerPipeline* there is a record of what settings were used (a duplicate TrimSettings.txt settings file is created in the same folder as *PrimerPipeline.exe*)).

4.2 MISA

2. example_misa.ini (MISA settings file. PrimerPipeline will read and write your MISA settings to this file (a duplicate misa.ini settings file is created in the same folder as *PrimerPipeline.exe*)).
3. example.misa (The output file from MISA, containing details of the microsatellites found).
4. example.misa_statistics (The MISA statistics file containing details of the MISA step, including how many files were examined).
5. example.p3in (Input file for Primer3).

Note 6: Running Primer3 only

Drag the Primer3 input file, example.p3in, into *PrimerPipeline* if you want to re-run the Primer3 stage only, *PrimerPipeline* will then ask for the example.misa file, direct it to this file, and then you can adjust primer3 settings and click *Process Pipeline* to re-run Primer3 and results stage.

4.3 Primer3

6. example.p3out (Primer3 output file, containing details of all Primers designed)
7. example.Primer3_ErrorLog.txt (Primer3 error log file)
8. example.Primer3_Settings.txt (Primer3 Settings file, (a duplicate Primer3_Settings.txt settings file is created in the same folder as *PrimerPipeline.exe*)).

4.4 Results

9. example.FinalResults.csv (This is the Results file that can be opened in *PrimerPipeline* and also excel)

5.0 Running the Pipeline

- When happy with the settings for all three stages, return to the *Selected Files* tab (Figure C2)
- Select your file so that *Process pipeline* button becomes enabled (blue), then click *Process pipeline* button to start.
- The pipeline will now run until results are ready.
- In the 'Status' column you will always be able to check the progress of the pipeline, it will state which step (of 3) it is on and the percentage complete.

Note 7: Move to next task/ cancel process

At any point during the process you can 'Move to next task' or 'Cancel process'. To do this right click on the running process and the options will pop up like in Figure C9.

If you click 'Move to next task' during 'Step 1/3 – Searching for microsatellites' *PrimerPipeline* will continue with 'Step 2/3 – Running Primer3' will the microsatellites already found.

If you click 'Move to next task' during 'Step 2/3 – Running Primer3' *PrimerPipeline* make the results file with the primers designed so far.

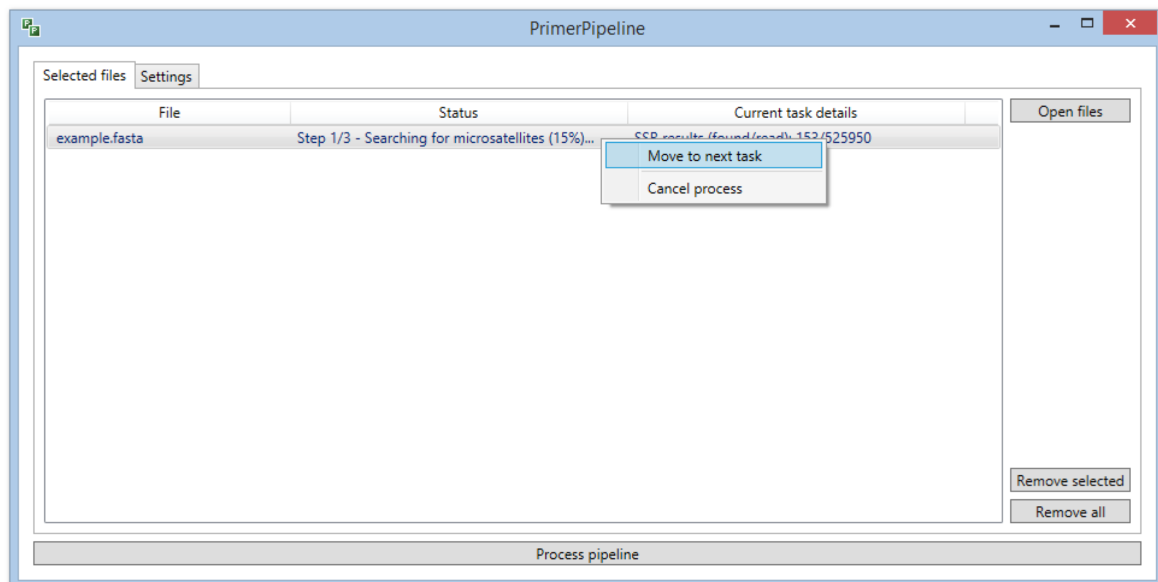


Figure C9. While *PrimerPipeline* is running you can 'Move to next task' or 'Cancel process' at any time.

5.1 Step 1.

This step incorporates Sequence trimming and MISA (finding microsatellites), and when complete makes a Primer3 input file for the next step.

- Within the 'Current task details' column the 'SSR results (found/read)' are shown and updated every 50 sequences, therefore is constantly updating (Figure C10).

- Within the 'Status' column the percentage complete compares the amount of sequences read to the total number of sequences in the input file (Figure C10).

Note 8: Current task details for Step 1

In the 'Current task details' column 'Read' = the number of sequences read by *PrimerPipeline* so far (Read for the Trimming stage, not necessarily searched for SSR's, this depends on your

Tip: If your sequences are listed in size order, smallest to largest sequence, (like outputs from SOAPdenovo are) then it will take proportionately longer per sequence to complete, therefore , during 'Searching for microsatellites' and 'Running Primer3' the

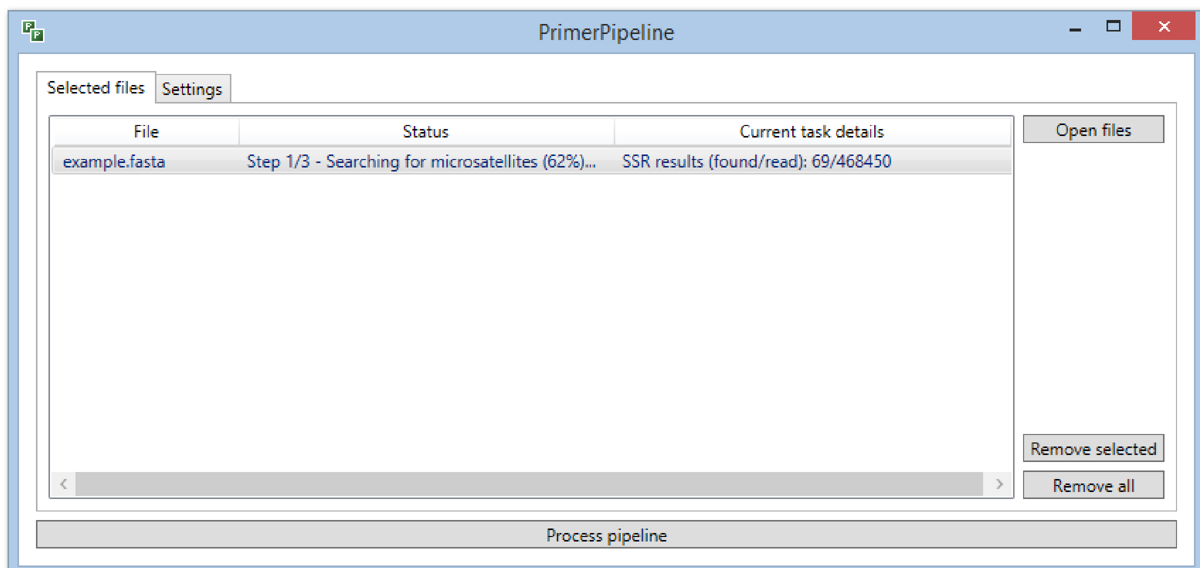


Figure C10. How *PrimerPipeline* looks whilst running step 1/3 – Searching for microsatellites.

5.2 Step 2.

Running Primer3 (Figure C11 and 12)

- During the Primer3 step the ‘Current task details’ and percentage complete in ‘Status’ updates every 30 seconds. This was a decision based on a compromise to keep the user informed during the process but not to significantly increase running time. So for the first 30 seconds of Primer3 running, the ‘Current task details’ will say ‘(updates every 30 seconds)’ as in Figure C11.
- After 30 seconds the ‘Current task details’ will show ‘Results found (valid/total)’ (Figure C12).
- In the ‘Status’ column the percentage is calculated by comparing the total amount of sequences in the Primer3 input file to the amount of sequences so far in the Primer3 output file.

Note 9: Current task details for Step 2

In the ‘Current task details’ column ‘total’ is the number of sequences (that contain an SSR) that have been read so far and, ‘valid’, are the number of results that Primer3 has been able to

5.3 Step 3.

Creating Results file.

- This is a very quick step, therefore is usually not shown on the display.
- When *PrimerPipeline* has finished (Figure C13) you can right-click on the file and go to ‘view results’ which will take you to the *Results* window (Figure C14).

5.4 Time Taken

Steps 1 and 2 are the longest steps; with time taken depending on your computer specifications, your input file and the settings that are chosen (e.g. number of sequences being read, the number of sequences that go through to the MISA stage, number of microsatellites found, and the number of primers that can be designed). Therefore it is

difficult to predict the total time taken for each process, (however if in a hurry, please see Note 7).

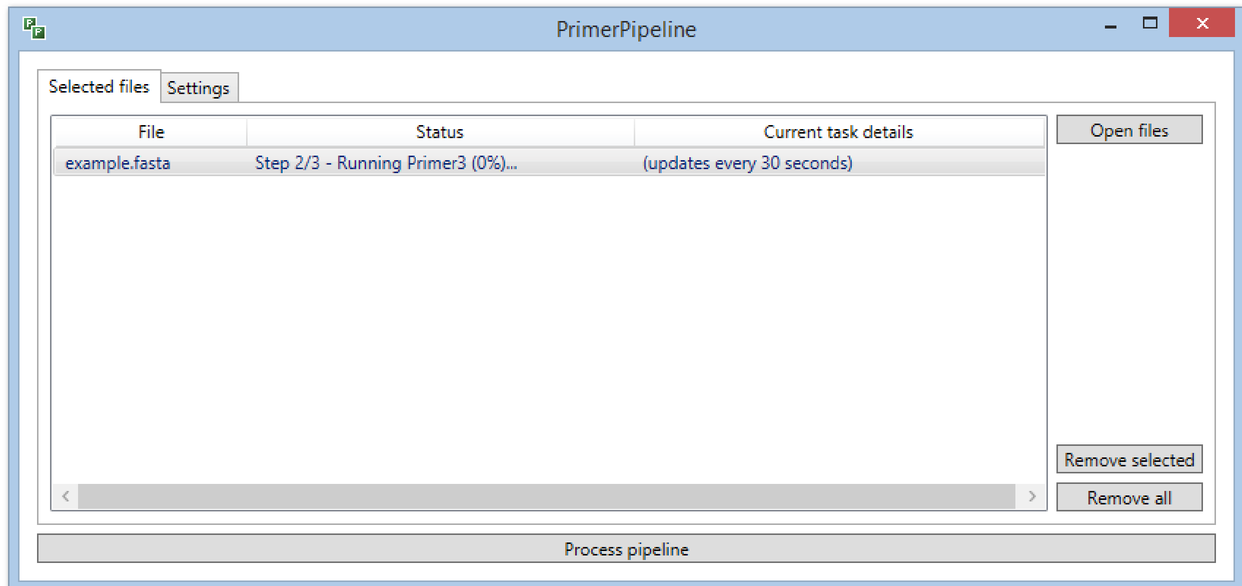


Figure C11. How *PrimerPipeline* looks for the first 30 seconds of step 2/3 – Running Primer3.

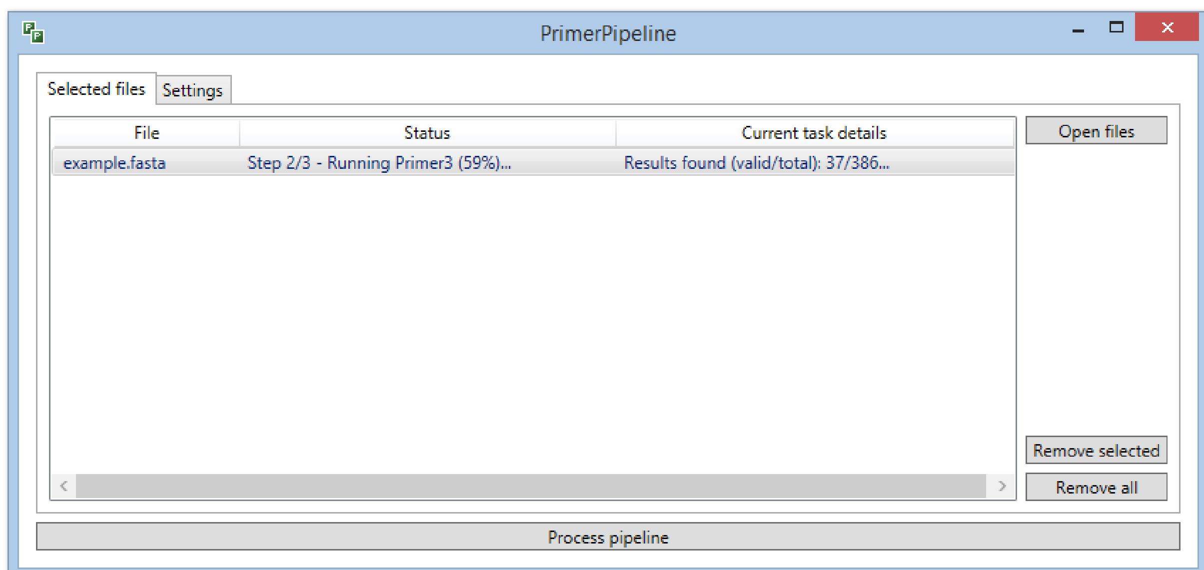


Figure C12. How *PrimerPipeline* looks when running step 2/3 – Running Primer3.

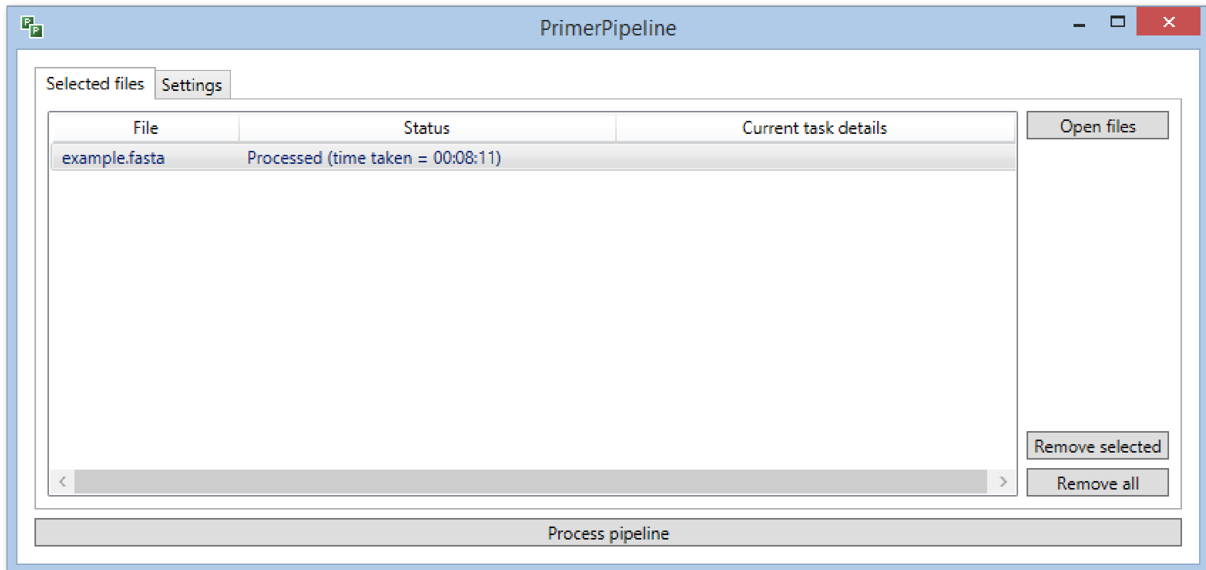


Figure C13. How *PrimerPipeline* looks when the pipeline has finished.

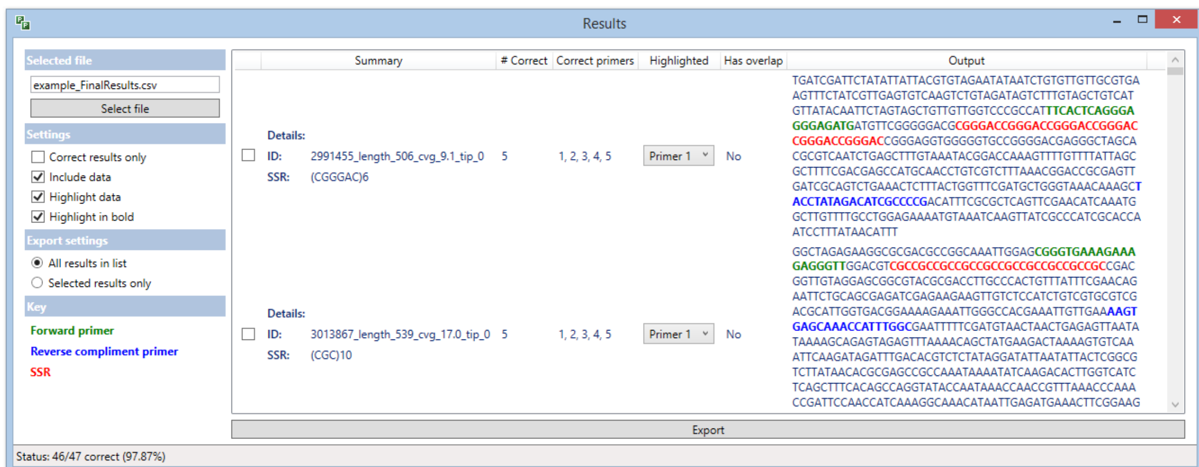


Figure C14. *PrimerPipeline* Results Window.

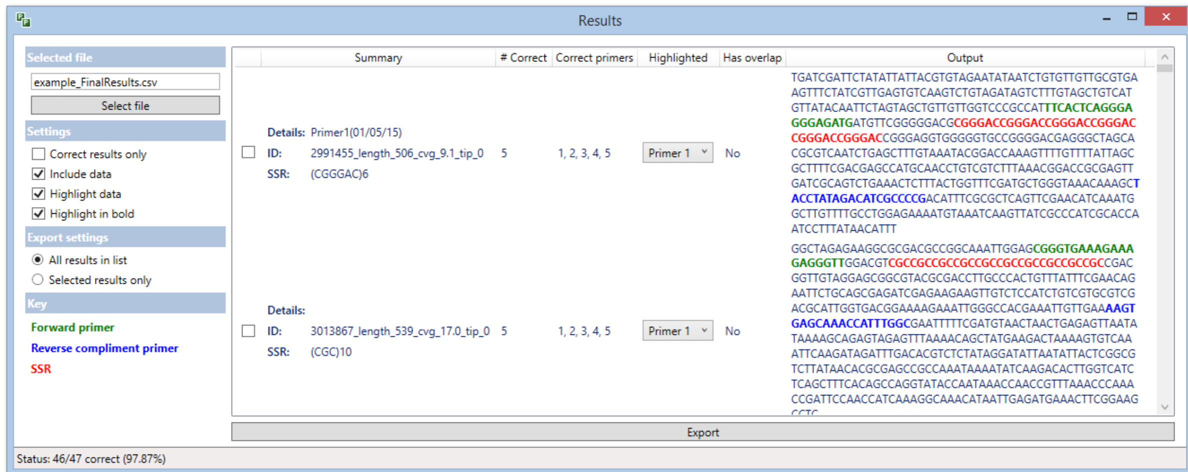


Figure C15. *PrimerPipeline* Results Window displaying information added in the 'Details' column of the FinalResults.csv.

6.0 Viewing results

When the pipeline has finished, the status will say 'Processed' (Figure C13). Right-click on the file, and click *View results*, this will open the 'Results' window (Figure C14).

- **Viewing.** You can view the results with or without the Data (Data means the sequence in the output column) by selecting or deselecting *Include data*, and you can chose whether to have the data highlighted or not, again by selecting or deselecting *Highlight data* and *Highlight in bold*.

Tip: We recommend viewing the results with *Include Data, Highlight data* and *Highlight in bold* all selected because seeing the sequence with primers and SSR clearly defined allows you to check the flanking regions and pick the primers you will test in an easy,

- **Ensuring correct results**. The status at the bottom of the 'Results' window (Figure C14) shows the results from an automatic check. The program has checked that the microsatellite and primer positions are in the correct order and do not overlap with each other. So you know that no errors in Primer3 have been made. If you see that a few primers are not correct, you can select *Correct results only* in the 'Settings', and click the *Export* button at the bottom (Figure C14), *PrimerPipeline* will ask you to 'Save as' to export all the correct sequences into a new .csv file.
- **Exporting.** If you wish to make a new .csv file with selected sequences only, click the boxes to the Left of the 'Summary' column, of the chosen sequences and then in *Export Settings*, click *Selected results only* before clicking the *Export* button at the bottom of the page.
- **Copying.** You can copy and paste sequences with formatting (with the forward and reverse primers and SSR highlighted) into Word, powerpoint, ect, if you wish.
- **Using the results file in Excel.** It is useful to open example.FinalResults.csv file in excel to edit, it is very flexible and allows you to plan your next steps in the microsatellite development. For example:

  1) To order the sequences by SSR type or length.

2) To add a more user-friendly name for the primer pairs that you will order (a 'Details' column in the FinalResults.csv file is felt empty for this purpose, see Figure C15).

3) If you're only interested in certain SSRs, you can make a new .csv file with only dinucleotide microsatellites or with a certain size range so you can open a smaller, more manageable number of sequences in *PrimerPipeline* 'Results' window.

- **Reopening Results files in** ***PrimerPipeline*****.** To reopen a file in *PrimerPipeline* save as .csv and drag into *PrimerPipeline*, it will recognise the file as a results file and immediately say 'Processed' in the 'Status' column, allowing you to right click on the file and click on *View Results*.

6.1 Results Window: Each column explained

1. Summary

Containing:

Details: Is the empty 'Details' column in the .csv file, allowing you to give the primers a new name or add any other extra information. If you write it in the column in excel and save as '.csv' it can be displayed by *PrimerPipeline*. If you haven't added your own names, this will be empty like in Figure C14, if you add a detail; it will look like Figure C15.

ID: Is the ID of the sequence from the original input file uploaded.

SSR: The SSR type and number of repeats e.g. (TG)24

2. # Correct: The number of primer pairs that are correct (out of however many you chose in the Primer3 settings 'Number of primers to return', default is 5). Correct means forward primer first, SRR then reverse primer with no overlap between the three.

3. Correct Primers: States which primer pairs (called 1 to 5) are correct.

4. Highlighted: This column has a drop down menu to enable you to change which primer pair is shown overlaid on the sequence, for each microsatellite. All show Primer pair 1 as default because Primer3 sorts primer pairs by quality, 1 being the highest quality. However if you are looking for a different product size for multiplexing then you may wish to use one of the other Primer pairs.

5. Has Overlap: Tells you if there are any overlap between the microsatellite and each primer. If you select *Correct only* under 'Settings', or if the status says 100% correct, then this will be 'No'. If there are overlap it will say 'Yes', avoid these primers if you can.

6. Output: This column shows the sequence in normal text, and if *Highlight data* and *Highlight in bold* are both selected the microsatellite (SSR) will be overlaid in **red**, and forward primer in **green** and the reverse compliment primer in **blue** (see key on Figure12). This allows you to easily view and check your primers. This column is only there if *Include data* is highlighted in 'Settings'.

7.0 Tips

1. Not finding enough microsatellites?

Consider relaxing the settings. To see where you need to change your settings first check the 'example.misa_statiz`stics' to see how many sequences were examined and compare this number to the total number of sequences in the original input file. If there is a large difference, it means that the *Trimming* step is possibly too strict, and isn't letting enough sequences through to the MISA stage. Consider decreasing you minimum sequence and increasing your maximum sequence.

If the trimming settings are fine, in the MISA settings consider decreasing the minimum repeat so that smaller microsatellites can be found or allowing interruptions so that compound microsatellites can be found.

2. Not finding enough Primers?

If you are finding plenty of microsatellites but are not getting enough primers, it may be due to the position of the microsatellite in the sequence. If the microsatellite is at the very beginning or very end of the sequence there is no room for Primer3 to design a primers around it. Check this by looking at the 'example.misa' file and looking at the start and end column. This shows where the microsatellites are located in the sequence.

There are a lot of settings during the Primer 3 stage which you can adjust, it is always a compromise between quality and quantity of primers. Look into more detail at the Primer3 settings and consider making them more relaxed. To read more about Primer3 settings please visit Primer3's website here (see Note 2 to run Primer3 only, without having to re-run Trimming and microsatellite mining steps).

8.0 Acknowledgments

We'd like to thank the creators of MISA and Primer3 for allowing us incorporate their amazing programs into ours.

I'd also like to thank all my colleagues in my lab for help and guidance, and a special thank you to my supervisors at Cardiff University for not being too hard on me for completely failing in my first attempt at bioinformatics, when I spend 6 months of my PhD trying to optimise primers that were incorrect. It turned out there was an error in the Linux pipeline I was using which meant that Primer3 designed primers randomly in the sequence, not around the SSR. We created *PrimerPipeline* to help others avoid such pitfalls when first embarking in bioinformatics.

9.0 References

9.1 MISA

For original MISA and trimmer pearl scripts:

Program name: misa.pl
Author: Thomas Thiel
Release date: 14/12/01 (version 1.0)
Website: http://pgrc.ipk-gatersleben.de/misa/

9.2 Primer3

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M and Rozen SG. Primer3--new capabilities and interfaces. Nucleic Acids Res. 2012 Aug 1;40(15):e115.

The paper is available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3424584/

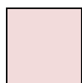
Source code available at http://sourceforge.net/projects/primer3/

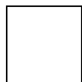
Website: http://primer3.ut

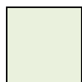
Appendix D - Microsatellite Development

Appendix D contains details of all correct microsatellite primers tested for each species during microsatellite development.

In the tables within this appendix, each microsatellite is colour coded relating to the outcome of testing each primer. The meaning of each colour is shown below:


Tested but not continued with


Chosen for publishing but not used in ongoing analysis


Chosen for publishing and used in ongoing analysis

Microsatellites chosen for publishing were renamed, as shown in the tables, and their sequences were submitted to Genbank.

Table of tables

Table D1 321
Table D2 330
Table D3 341

Amphinemura sulcicollis

Table D1. Details of microsatellite development for *Amphinemura sulcicollis*. Microsatellites were initially named Amp_1 to Amp_75, however Amp_1 to 50 were incorrect due to an error in primer design. This table details Amp_51 to Amp_75, containing the microsatellite number (*No.*) in order of testing; original *Contig ID* from Next Generation Sequencing, the *Name*, *SSR* (Simple Sequence Repeat), *Forward primer*, Annealing temperature (*Tm*) of the forward primer, *Reverse primer* and corresponding *Tm*, the *Product size* of the amplified region, and details of the *Outcome* from testing this microsatellite. The corresponding sequence can be found below each microsatellite with the forward primer highlighted in green, the SSR highlighted in red and reverse primer highlighted in blue and all underlined.

<table>
<tr><th>No.</th><th>Details</th><th>SSR</th><th>Forward primer (5'-3')</th><th>Tm (°C)</th><th>Reverse primer</th><th>Tm (°C)</th><th>Product size</th><th>Outcome</th></tr>
<tr><td rowspan="2">1</td><td>ID: 3251325_length_3685_cvg_19.5
Name: Amp_51</td><td>(GT)28</td><td>ATCCGCTCAATCACACTGCA</td><td>60</td><td>CCTATCGAGTATCGCGCAGG</td><td>60</td><td>274</td><td>No samples successfully amplified in 1st round of analysis -rejected</td></tr>
<tr><td colspan="8">GGCGTATCCGCTCAATCACACTGCACCAAAATTCACAGGTCGCGTTTTGTGTGTGTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTAAGTGTGTGTGTGTGTGATGCTTAGTACCTTTAAGTAGTCGAGATCACTTGCAAAGTTAGACAGTTCAAATATCTTGCAGCTCGTTGCCGGCAAAGTGGCATAGTATCTCACCAACTAATTGGATTACATTACTTGACACTCGCCTGCGCGATACTCGATAGGCTCAAAAACAAGCGAAACATACTGAGATCTTGTTGGGATCTTACTGGGATCTTGGCTAGGAGCTTGCCAGCCACACGGTGAGCCCACATGGCACAGAGCA</td></tr>
<tr><td rowspan="2">2</td><td>ID: 3253285_length_4017_cvg_18.2
Name: Amp_52
New name: Amp_13
Genbank #: KR069027</td><td>(CT)27</td><td>TCACACGCTCTATACACCGC</td><td>60</td><td>TTTCTACCCTCTGTGTCGCC</td><td>59</td><td>151</td><td>Through to 2nd round of analysis, but not found to be in HWE - accepted for publication and ongoing analysis.</td></tr>
<tr><td colspan="8">GACGATTTTTCCACTTCCACAACTTTTCCGCCATCAACAAGCTAGAAGGTCATCACTGACAAACATTTTCCAAGAACCAACACCATGGGCTACATCACCGTCACACGCTCTATACACCGCGTTTTCCCATCCAACAAGAACGTCATTTACTTTCCTCTCTATTAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCGAGGCAGTGCCGGCGACACAGAGGGTAGAAAAAATACTTCCTTTGTCTCCTTTAATGCATTTGATTAAATAATTCACATTTCGAGCCTATTATTCTTGATAAGGTTTTGCGTAAAGTGTCAGAAGCAGAAA</td></tr>
</table>

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
3	**ID:** 3237201_length_2407_cvg_18.8 **Name:** Amp_53	(CA)26	ATTCCACCCATCACAGCT GT	60	ACATACAGTGACAGTCCC GTT	59	185	No samples successfully amplified in 1st round of analysis -rejected
	GATAGGTACGTTTTATAATTTTGCTTCTTAGTTTGCGACATTTTTAATGTTGATCACATACAATTACGTGCATGAAAACGATTGTAAAATGCTTGAACAC**ATTCCACCCATCACAGCTGT**GTATGCTGATAATTATTTAAAGTCAC**CACACACACACACACACACACACACACACACACACACACACACACACACACA**CCCTCACACACACACACACCAACATAAAAGTTAATCACAACTGCGACAAATTTTCTAGAAAGTTAA**AACGGGACTGTCACTGTATGT**ATACACACACGACAATTTAGCCTGATAAATTACCTCAAAGCTAGACGAC							
4	**ID:** 3197869_length_1335_cvg_16 **Name:** Amp_54 **New name:** Amp_18 **Genbank #:** KR069032	(AG)25	CATCCCCATACGGCGGTT AA	60	TTGCTCTCTCCCTTTCCCCT	60	171	Through to 2nd round of analysis, but not found to be in HWE- accepted for publication but not continuing for ongoing analysis.
	ACGAGACACGCAATGTCACAACAAATTCTCGCAAATTACAAACAGCAGCGACGTTCAGACCATCAAAGGGTCGCGAGGTCAAAACACCATTGTCATCGCG**CATCCCCATACGGCGGTTAA**CAAGCGCCGAACAATTTGAAAGGTGAGGAATGCAGATTTGCTGCAATGTCGTCGCCGTCGTCGCGGTGA**AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG**AACGCGAGAGGG**AGGGGAAAGGGAGAGAGCAA**GCCGCCCACCCGGATGGCCCCGCGCCGAGAACTAGAGCGCCTGGCAGAGTGCAGACACGCCCTAGCCAGACCCGTGAAGGTCATCTCGCGGCAAAAGATT							
5	**ID:** 3176935_length_1107_cvg_12.9 **Name:** Amp_55 **New name:** Amp_2 **Genbank #:** KR069016	(CA)24	GTGTGAAGTGGAGGCTG TCA	60	ACAGCGCAAACTAAATCG CTC	60	233	Through to 2nd round of analysis, and found to be in HWE in both populations tested – accepted for publication and ongoing analysis.
	GTGCGCTCAGAATCGCAGCGCTGCCTTTCACCATCGGCACGGCGTGCTTTCTTTCAGGCGTCGCTTTGAAGGCGACGCCCGGCGCCCGACGCCGGCGAGT**GTGTGAAGTGGAGGCTGTCA**TCCGAGATAACGTGATCATCTAGGCGTGCGCTGAACTGCTTTGACGAGGACACCCCGAACGCGATGGATCCGATGGAACGGATTGATGATTGAAAGGATTTACTTTAATACACCCAGCACTGCATG**CACACACACACACACACACACACACACACACACACACACACACACACACA**TAGATACTTGCACACAGA**GAGCGATTTAGTTTGCGCTGT**TTTGCACCATTACGCGCTCGGTTTGTGAACCAATCCGCTTTGAATTTACTTTGAGACTGAGTAGCAGGATGTCTTTTGTTTTTGCGTTTAATGAATTGGA							

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
6	**ID:** 3260393_length_6163_cvg_20.1 **Name:** Amp_56 **New name:** Amp_21 **Genbank #:** KR069035	(TG)24	AGACGCAGCGAATTTAG AGGT	60	TGCAAATGGTCACGTTACT GTT	59	249	Through to 2nd round of analysis, but not found to be in HWE- accepted for publication but not continuing for ongoing analysis.
	CCCGCCGACCTCAGAGCAGTGGGCATCCCCTCCCTCAATGATCCGAGGCCGAGGGGGTAACTAGCTTACAGATATAGGAATATAGGCCGTGATTGGAATT**AGACGCAGCGAATTTAGAGGT**AGTTACCCTTGCAGGTGGATTTTTAATTTTAGGTATTTGTGTGTGTGTGTC**TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTG**ATTTTCAAAACAGTTTTTGTTTACGATATAACTTTGCATAATATGTGTGTGTAAAGAAAGATGATTTATAAACCATGTTATGCTTTTTATTATTTAGTAATATTTAT**AACAGTAACGTGACCATTTGCA**ATTACAACCACTTTCAATTTATTACAGAACTCGGTTGAAATTTGAAAAATCTTTTGGTTTGTAGTTTCTATAGGTGGGTTTACCATAAGGTGAGCCGGAT							
7	**ID:** 3250749_length_3601_cvg_19.1 **Name:** Amp_57 **New name:** Amp_1 **Genbank #:** KR069015	(AG)23	TCAACCTATGGATCTGG AAACGA	59	TGCATTTCCTTTCTCTTTCC AGC	60	199	Through to 2nd round of analysis, and found to be in HWE in both populations tested – accepted for publication but not continuing for ongoing analysis.
	ACTGCTCTCATATCTTGAATGTGCATTTGTTTCAAGATCGGATAATGTTAAATGCTCACAGTTGTAGATTTAAAGTACCCAGCATTCTGATGTATTTTTT**TCAACCTATGGATCTGGAAACGA**TAAAATCAGAATAACGA**AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG**AAAGAGAGAGAAAATAAATAAATAAATAATTCTACGGCTATCCAAATACATGGTTGACCATATTAATCTTTATTATAAAATAAAATGAAA**GCTGGAAAGAGAAAGGAAATGCA**GCTTACAGTGAAAAAAAGAATATTGGGCTGCCACAGTACTTGTAGATGGAACGAAACGTTTCGGCGTCTGCCTTCGTCAGTTTGATGAGAAATGATGTCA							
8	**ID:** 3257949_length_5134_cvg_20.5 **Name:** Amp_58 **New name:** Amp_3 **Genbank #:** KR069017	(TG)23	GCCGTGCAGATCTCCAT ACT	60	ACTCACCTGCCTCGAACAA G	60	185	Through to 2nd round of analysis, and found to be in HWE in both populations tested – accepted for publication and ongoing analysis.

<table>
<tr><th>No.</th><th>Details</th><th>SSR</th><th>Forward primer (5'-3')</th><th>Tm (°C)</th><th>Reverse primer</th><th>Tm (°C)</th><th>Product size</th><th>Outcome</th></tr>
<tr><td></td><td colspan="8">TTTATCTCATCTCATGAAAACAGAAACCTGATGTTTATAATTTTTCAAATATAGTTACCCTAATAATAGAACAAATTCGCACGAAACCTTGTTGTTTTTGGCCGTGCAGATCTCCATACTTCCTAAAGTAAGCCCCTTTATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTACAAATTTTATTGGTGCTTTAAACGCAACGAGGACAATTTGACTGGAACTAATTTAAAGATAAAGTCTGGCAACAAGGCTTGTTCGAGGCAGGTGAGTAGAAACTAAAACGCGTTGCCATCTGTCCTGTTCCTCGCAAGCAGCTACCGGCAAAAGTGCGTATACTAGTGCCTATGGCTGGCTGGTACCACAGACTGGT</td></tr>
<tr><td rowspan="2">9</td><td>ID: 3257095_length_4889_cvg_17.8
Name: Amp_59
New name: Amp_11
Genbank #: KR069025</td><td>(GT)23</td><td>CATTTTGCTGCCCAGTTGCT</td><td>60</td><td>AAAAGCCGCGAACACACAAA</td><td>60</td><td>250</td><td>Through to 2nd round of analysis, and found to be in HWE in one population – accepted for publication and ongoing analysis.</td></tr>
<tr><td colspan="8">TATGTGTCATGAAGAAAATGCCTGGTTTTGAAAATAGTAGGTTTAGATCTGCAAGAATTGCAATTCAACTTTGCCTTTCAACTTGAACCTGGGAAGTGGCCATTTTGCTGCCCAGTTGCTTAGAAAAATAATAGACACAGCGTTTAAGATAAATACAGGTGCTTTTTTCAATTCTTGTATACATAGAAATAACTCTGGAAGGTAAAGATATTCCGTTCATGTGTGATATTCATTGACTCTTTCAATTACACTGATCCAGAATGCTCATCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGCGTTTGTATGTTTGTTTGTGTGTTCGCGGCTTTTCGCTCGCAGCGTAGCGGCCAAACAAAAGCACGGATTGCGTTGTAATTTTTAGGGAACCTTTTAAGGTTAGATACCCATTGTTTTTGAGGACGATCCGAGC</td></tr>
<tr><td rowspan="2">10</td><td>ID: 3260569_length_6259_cvg_18.2
Name: Amp_60
New name: Amp_5
Genbank #: KR069019</td><td>(AG)22</td><td>CCAGGGAGGTGATTGCAAGT</td><td>60</td><td>TCCGGTGGACTGCACAATTT</td><td>60</td><td>189</td><td>Through to 2nd round of analysis, and found to be in HWE in both populations tested – accepted for publication and ongoing analysis.</td></tr>
<tr><td colspan="8">CCCATGGGCTACCCAGGGAGGTGATTGCAAGTATTTTTGTAGTTTGTAAGGTGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAAGAGAGAGAGACAGCCGCTCGTGGCCACGGCTGGCAGCGTGAGAGATAAGACACCAAGAGAGAAATGAGAGACAGCACCAAAATTGTGCAGTCCACCGGATTCGGCACCAGCATCTTCGATCTCAAGCAACAGTCTTGTGAATGCAGTAAAATTTCTATTCTCCAGCCAATACATAATTAGACTGATAATTTGAAATTTG</td></tr>
<tr><td>11</td><td>ID: 3256387_length_4696_cvg_17.2
Name: Amp_61
New name: Amp_9
Genbank #: KR069023</td><td>(GA)22</td><td>TCAAGCGGTCTCCTTTGCTC</td><td>60</td><td>CCCTTTTCACGCATCGCAAT</td><td>60</td><td>204</td><td>Through to 2nd round of analysis and found to be in HWE in both populations – accepted for publication and ongoing analysis.</td></tr>
</table>

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
	ATGCGACAAATCGTGGAGTTTTCCTCGACGACGGAAAGCCGAGGGTGCTGTACGAGAATTTATCGTGATCTCTATGCTCTTTTAACATGTCGCACTTCTCTCAAGCGGTCTCCTTTGCTCTTGTAGCCGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGGAGGGAGAGAGAGGCCCAAACTTTACATGACACGGGAGTTTGTCATGACTGTGTTTTGGCTAATAAACTATTTCTGTCGTGGTGCAAGGATTTATCAGACCCAACATTGCGATGCGTGAAAAGGGAATGTAATGTCACACTTCAAGTACTGCCTTGGATGATTGATTAGAAGATATAAAATTTTAGCATTTTTTGCCAAGATATCAGAATATATTTAAGATTTTT							
12	**ID:** 3251643_length_3736_cvg_23.3 **Name:** Amp_62	(AC)21	CAGAGCGCGAACAAACAGG	60	GCACAATTGACCTTCTTTCATGC	59	275	No samples successfully amplified in 1st round of analysis -rejected
	TGGATTTATTGAAAATTGAATTAGGTTCAAGTTGAGAACACAAACACCTGCCCTCCAAATTTGGTTCAAATCGGTGCAGGCGTTTGGCCGCCAGAGCGCGAACAAACAGGCGAAACACACACACACACACACACACACACACACACACACACACACAAAAGTGCTTCTTTTCTACCGTATGTGAAATTGAAGATTTCGCTTGCTACGCTCATCAATTAGAGTAGGTTATTAGTCGGAATTCAAGTAAAATAGGTTAGTCTCATTTTTGTAATTTTTATTGTAATAGTTTATTTTTCAATTTCTTTGCAGAAACACTCTTTAAGTTGGTTACAACAAAATACGAGCATGAAAGAAGGTCAATTGTGCATAAAATACATCATGATAAAATTAAAACCAAATACCAATATACAGTCAATTTATACAGATAAATCTATACAATTCTGTAATTATGTATATATTATTATAT							
13	**ID:** 3196705_length_1320_cvg_11.6 **Name:** Amp_63 **New name:** Amp_8 **Genbank #:** KR069022	(GA)21	CGGATCGGATGCAGTGGTTA	60	GTCGGGCAAGTGACTAGAGG	60	125	Through to 2nd round of analysis, and found to be in HWE in both populations tested – accepted for publication and ongoing analysis.
	GGCAAGCTGGGTCACTGAACTGCTGAAGAGGACCCGCCACTGCCTTCTCCGGCGACCGACTCGAGCGGACTGCGAGGAGTGAGGACTGACCATGATTCGGCGGATCGGATGCAGTGGTTACCTGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCTTATACTCGCGGTCGACAAACCTGAGGGTGGCGCCTCTAGTCACTTGCCCGACGATTGGGCAACATAACAAAATGAAAGCTTAGCGTATGCTCCAAAGCTTTAAATTTTGATAAATAAGTACCCTAGGAAAGCTGGACGTATTTGTTTGTTTG							
14	**ID:** 3192841 **Name:** Amp_64 **New name:** Amp_15 **Genbank #:** KR069029	(TG)21	GTGGGATAAGCGCCTAGAGTT	60	GTTAGGGTCGGCATGGAACT	60	235	Through to 2nd round of analysis, and found to be in HWE in one population – accepted for publication but not continuing for ongoing analysis.

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
	GACTATTGAAACGCACACGACCGCGGTTAACAACGCAAGTTCGTATAAAGCCAGCTGGCATCGCAACGCAACGATAAAAACACTCTGGCCGGATCGCGATGTGGGATAAGCGCCTAGAGTTTTTAATTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGCAAACTCGAAGATCGAAGCCGACTCGCAGACCCTACTCGTGAAATCCCAGCCTAGCCTAACCAATCCAAATCCAGTTCTGCAGAGAAACTGCGTGGTGAAACCCCAACCAGGCCTTGAACGGCTGTGACTGAAACCCCCCGTCACGAAGTTCCATGCCGACCCTAACCCTAATTCTAAGAACCCCAGCTAATCTAACTTAACTCGGCCCGCCAGCGCATCTGCACGGGCTTCGTTGGCTTCCCCGGAATGCAGAAAACATCCCGAGA							
15	**ID:** 3257263 **Name:** Amp_65 **New name:** Amp_14 **Genbank #:** KR069028	(CA)21	GCAACATTCAGGTCAACGCA	60	GGCAGGAGAACACAACACCT	60	178	Through to 2nd round of analysis, but not found to be in HWE- accepted for publication but not continuing for ongoing analysis.
	AACTCTAACAAATATTCGGTAACGAAAAAGCTCTTGATCCATAAATATTTTTAACGCCACAACCAAGGGCTAAATTTTCTTCCCACGTGTGCCAAGTTTGGCAACATTCAGGTCAACGCAAACTCAAACGCACACACACACACACACACACACACACACACACACACACACAATTCCATCGGTAAAGGCATAGACTATAGATAAAGGTAAAGGTAAAGGTAAAGGTAAAGGTAAAGGTAAAGGTAAAGGTAAAGGTAAAGGTGTTGTGTTCTCCTGCCTGTTAACAACGTAACAAAGCAAAACAGGTCGACTTTATTCGAAATTGCAACCATCACTTCTAACACCCATCTCAAAAGGAGCAACTGCAAATCCCCAAGG							
16	**ID:** 3197241 **Name:** Amp_66 **New name:** Amp_4 **Genbank #:** KR069018	(TC)20	GGTCGACGCTGATAGTTTTCC	59	AGAAGGATTATGGGAGGAGGAGA	60	152	Through to 2nd round of analysis, and found to be in HWE in both populations tested – accepted for publication and ongoing analysis.
	CCGCGAGCCAACCAGCGTCGCCGAAGCGACGAAAGAGTGACTATGAGATCCGGTCGACGCTGATAGTTTTCCTATGAGAGCGCCAGTGTCTTGCCCAACAATTGTGATCACAGCCCATGCCCATCATCAATCAATTACGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTTCTCCTCCTCCCATAATCCTTCTCTATCTTACAAGCTAATTAGCAACCTGTGCGACCTCTTTATGCACACACTGGTAAAGCCAATCGGAGATTTGTTTATTTCAGGGAACCCAATGTCATAGG							
17	**ID:** 3212711 **Name:** Amp_67 **New name:** Amp_16 **Genbank #:** KR069030	(AC)20	TATCCAGCAGACATCGCCAG	60	GGGACCCTTGTGAGTCGGTA	61	250	Through to 2nd round of analysis, but not found to be in HWE- accepted for publication but not continuing for ongoing analysis.

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
	GTGGATTTTGAACTCTGTGTTCTTAGGCACAGCACAAACCTACAACAGTGCTGCCGTGGTGGTCACTTTAATCAACTTATACGCTAGAAAAACTACGGCATATCCAGCAGACATCGCCAGTGTTTGCTCAGTTAAAGTGAGCAAAATCTGTTCACAGCTTGTCTAATTGTATTGTTAACTAAACTAAAGCCAACCCTTCTCACTTCACATGGCAAAACACGCATACACACACACACACACACACACACACACACACACACACACACAGATAAATAAGCTAGCTCGAGTCGTTCACATATCAAACTACACACGACTCACTCCACAGAATTAATTACCGACTCACAAGGGTCCCCCCCCCC							
18	**ID:** 3171045 **Name:** Amp_68 **New name:** Amp_7 **Genbank #:** KR069021	(GT)22	TAACGCACTGTGTACGCACT	60	CCCAAATTGTGATTGGCCGT	59	278	Through to 2nd round of analysis, and found to be in HWE in both populations tested – accepted for publication but not continuing for ongoing analysis.
	GCCCGGGATCGAGACAAGCTCACAAGCATAACGCACTGTGTACGCACTTTCTATCTTAACGACGCACTTTCGCCGCTGCGTTTATCGCCGCTTCATCTCGAGCACGCATCATGCGTCTAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTAGAGAAAGTGCGTGCTTTCGAATGTGCGCGTGTCGTTATTTTTCATGCGGCTCACACACATGAACATTTTGCGTTGTGTGTTTGTCGTACGCGGCCGACTAGATTCGCTGTTCACAGTGGCGCACGGCCAATCACAATTTGGGAGAATTTTGTTTGAATATTTGTAGGTCTTAAAGACCTGGTAGTCTGAGTCTCATATCACAGTTGAAGCAAAATCTTACAGAGTTGACTTCAAAGATGAAG							
19	**ID:** 3151381 **Name:** Amp_69 **New name:** Amp_19 **Genbank #:** KR069033	(TG)24	ATCCGCTCAATCACACTGCA	60	CTGAAATTTCGCCAGGTGCC	60	177	Through to 2nd round of analysis, but not found to be in HWE- accepted for publication but not continuing for ongoing analysis.
	GGCGTATCCGCTCAATCACACTGCACCAAAATTCACAGGTCGCGTTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTTTGTTGTGACCAGAACCCATGAGCTCAAATTTTCTAGGGATTGGAGCAGTGCTGAGAGTGATTTTCGGCACCTGGCGAAATTTCAGCCCTAACCTCGGATTGAAATAAAAACACGTTTTGCAGGTGTAACTTGTAAGATCCGGGCAGTACTTTTTTACACATAGGGTCTTTAAAAGGCCTTTTTGA							
20	**ID:** 3182603 **Name:** Amp_70 **New name:** Amp_12 **Genbank #:** KR069026	(CGC)18	CAAGTTGGAAGCTCGGGTCT	60	GTTTTTAGTGGGCCCATCGC	60	200	Through to 2nd round of analysis, but not found to be in HWE and low success rate- accepted for publication but not continuing for ongoing analysis.

<table>
<tr><th>No.</th><th>Details</th><th>SSR</th><th>Forward primer (5'-3')</th><th>Tm (°C)</th><th>Reverse primer</th><th>Tm (°C)</th><th>Product size</th><th>Outcome</th></tr>
<tr><td></td><td colspan="8">GGCCAAAAACTTTTTGCAAGCATAGCGAGTTAACATTACGCAATCTCGTCGATCAGAGAGCATTTCCTCGAGGTATTTTTCTATTCGCGGTGATTCGTATCAAGTTGGAAGCTCGGGTCTCGGTCTTTGAATGGCTTCCAGTTCGCAATCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCTGTTGGGTGGTGGGGGTAGGAGTGCCTCGCAAACACACACTAGTTTATTGAGCCGGGGCCGCGAAGGGCTGGGGGGCGCGATGGGCCCACTAAAAACCTACCCCGCGCAGCGCAGCGCGGCTCGCGGCTCGTGGCGGCACAGATAAGACGCAACGATCAATTTAATTACGCAAGTTTTTCCATTCGCTTAGAAATGA</td></tr>
<tr><td>21</td><td>ID: 2754993
Name: Amp_71</td><td>(GT)25</td><td>CGTAATTCCGTGCAGTACAGC</td><td>61</td><td>TGACTGACGCGCTAAGTTTG</td><td>60</td><td>205</td><td>Poor amplification - rejected</td></tr>
<tr><td></td><td colspan="8">CAATTCCTGAGAAATTAGATTATCGTAATTCCGTGCAGTACAGCTTACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGCTTACCAGTGGCGTGCAGTCCATGGAAGCAAAGAAAGTACTGCTATACCTTAGCAATGCCGTGGTCGGGCGCTTTTCAACCGACGATGCATATTAAGTTGTGAGAGCACAAACTTAGCGCGTCAGTCATATAAATGTTGCTTTTCAAATTATTCTGTGCAAATACTCATGGTGAGTACTAGCGGCGGCTGTCGCCTTTGCACAGAAG</td></tr>
<tr><td>22</td><td>ID: 2825633
Name: Amp_72
New name: Amp_6
Genbank #: KR069020</td><td>(AC)21</td><td>TCGGTTTCACGTGCAGTTAG</td><td>60</td><td>TGTCAATACGTTCCACTGCG</td><td>61</td><td>249</td><td>Successfully amplified for two sites, in HWE for both populations – accepted for publication and ongoing analysis.</td></tr>
<tr><td></td><td colspan="8">TTTTGCCCGCCACTCTCGGTTTCACGTGCAGTTAGCACATACACACACACACACACACACACACACACACACACACACACACATCCTGGGGAAGCGAGTTAAAAGTTTATCTAAACGTGGAAGGAAGCGGAGAACATATAACAAATCAACTTAGAAACAGTTAAGTTAGACTGCGCTTTAAAACACAAAGAAGGTAAAGTTCGCAAGGGGATTCGGTTGATTGAATTTCGCCGCGGGCGCTTGGCGCAGTGGAACGTATTGACACACGAGAAAGAAAAGTGAAATAATGAATTTATCGGGAAAACCTCTAGCCGCCGGCGTGGTAATTATAAACGCTGTGGGCATACACTTT</td></tr>
<tr><td>23</td><td>ID: 3024733
Name: Amp_73
New name: Amp_17
Genbank #: KR069031</td><td>(AC)25</td><td>TGCCCTTACTGTGTTTGTCG</td><td>58</td><td>AGTGCGAATATGGGCCTGTC</td><td>60</td><td>188</td><td>Successfully amplified for two sites, out of HWE for both populations – accepted for publication but not continuing for ongoing analysis.</td></tr>
<tr><td></td><td colspan="8">TCTTATTCTTATGCTTAAATAGCACAAAATTCCATGTTTTCAAACCTTTCCGTTTTTCCGTTTTTTCTGTATTTTAATATTTTTCCGCGACGTTTTCCGGATATTTTTGCCCTTACTGTGTTTGTCGATCACACAAACACACACACACACACACACACACACACACACACACACACACACACACACAAAATCTCATCAAAATGCTAGTAGCCAGAGCCACCAATCGACCAAGTAAACGTAAATTATACTCACACTCACTCATGGACGAGGACTGCGACAGGCCCATATTCGCACTCGTAATGAACTGAATCTTTTTTTCTGTAACCAGCGCGACAGTGATCGCCCCTCCC</td></tr>
</table>

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
24	**ID:** 2913891 **Name:** Amp_74 **New name:** Amp_20 **Genbank #:** KR069034	(TC)25	TCCTTTCCACGTCTCGGTC	61	TGGCCTGCTCCTAGAAGTTC	60	152	Successfully amplified for two sites, out of HWE for both populations – accepted for publication but not continuing for ongoing analysis.
	TCCTTTCCACGTCTCGGTCGTTT**TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC**GTGTTGCTCGCGTCCGACGGCAGTTGATATGTGGATTCGCGGTCCCCGATCTCGCGAGC**GAACTTCTAGGAGCAGGCCA**GGAGCAGGTCAGGAGCAGGCCAGGAGCAGGCCAGGAGCAGGTCAGGAGCAGGTCAGGAGCAGGTCAGGAGCAGGCCAGGAGCAGGCGAGGAGGTGGCGAG							
25	**ID:** 2973107 **Name:** Amp_75 **New name:** Amp_10 **Genbank #:** KR069024	(GA)21	AGAGATAGCGGCGTATCACG	60	TTTCCTGTCCCTCACACTCC	60	248	Successfully amplified for two sites, in HWE for both populations – accepted for publication and ongoing analysis.
	GGCATCGTCGACATCGAGTTGGACAGCCAAGGCCATTCCACGGGAATGTTTAGCGTAAGTATTATTATTTTTTACAGTAAAGTTCTCTGATAATTG**AGAGATAGCGGCGTATCACG**GGAACGCTGTCATCTAT**GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA**GGGAGTACTAAATAGAAGTTGCTATTAGACTTAAAGAGAGTTAGAGGGGAGAGGGGAGAGGGAGAAGCCAAAACTGAAACGAAAGCCACGGCTTTTCGAAAATTATCGGCGATGTCAACAACTTCATTAATTTTTTACCATGACTGCCT**GGAGTGTGAGGGACAGGAAA**ATAGAATTTCAGGCTCCTCTCGCTTTCCTCCAAAGAATATCCTGACCCAAAGATGAAGACATCTTCCTGTTGGTTATAAGTAGCTCCGGTCGTTGATCCT							

Isoperla grammatica

Table D2. Details of microsatellite development for *Isoperla grammatica*. Microsatellites were initially named Iso_1 to Iso_65, however Iso_1 to 44 (apart from Iso_12, 15, 18 & 22) were incorrect due to an error in primer design. This table details Iso_12, 15, 18, 22 that were correct despite the error, Iso_3(5), 8(2), 14(3) that were tested again with a different primer set correcting the error (as Primer3 gives you 5 sets of forward and reverse primers per SSR) and Iso_45-65. This table contains the microsatellite number (*No.*) in order of testing; original *Contig ID* from Next Generation Sequencing, the *Name*, *SSR* (Simple Sequence Repeat), *Forward primer*, Annealing temperature (*Tm*) of the forward primer, *Reverse primer* and corresponding *Tm*, the *Product size* of the amplified region, and details of the *Outcome* from testing this microsatellite. The corresponding sequence can be found below each microsatellite with the forward primer highlighted in green, the SSR highlighted in red and reverse primer highlighted in blue and all underlined.

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
1	**ID:** 13685736_length_1337_cvg_4.7_tip_0 **Name:** Iso_12 **New name:** Iso_13 **Genbank #:** KR069009	(AG)23	AATACACGAGCTGGCCCAAA	60	AGGCACCCAATCAGTGATCG	60	210	Through to 2nd round of analysis, but not found to be in HWE - accepted for publication but not continuing for ongoing analysis.
	TTGTGTAATTCTTAAGGCGCTTTGCCCTGGGGAAGTCACATTAATAAAATATATAATATTCACTTATTCTATTGCGTTCACCCACTGTGCTGGGTTAACA**AATACACGAGCTGGCCCAAA**TATGAACAAAGAGAGAGTGAGAGAGACTAGAGACTAAAGAGAGATAGATGGATCAATGTCTTTATTTTCAGGCGTGATTA**AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG**AATGCGAGCGAAGCACAGGCGCATTCTGCAATCATCTATAAATA**CGATCACTGATTGGGTGCCT**AACACATGAACTAGTCACCTCGTACGACATGCAGAGCGCCCTCTCCCGATAAAGCGAAGCACTAACGAAATGAAACCTACACGGTCGTCGTCCTAACCTT							
2	**ID:** 13398137_length_723_cvg_4.4_tip_0 **Name:** Iso_15	(CT)21	GTCTGCACAGCATGTTTCGG	60	AGCGATGGGTCTGACAGTTG	60	242	After 1st round of analysis, low success rate, and no heterozygotes - rejected

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
	AAGTTCTAACGGAATAAGTTAGACACTTCGAAACGAAAGTATATTTCGAGCGAGACAAATGGACAAAGTTCTAGTTCTCGTTTGTGGGTCTGTCTAATCG**GTCTGCACAGCATGTTTCGG**TGTAGTTTCGGTGTACACTTTCAAA**CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**CGCTCTACTCCTCCCCAAACTCTTCACTTATATCTCGTTGTTTTATATAATACCAATACTGCAACGATGATGATATCTTGTACAGGATTCGAACCCACGCCCTTTGATTCCGGTTGGTTTGGATCTGATCCCCT**CAACTGTCAGACCCATCGCT**GTCGAGCTCCCTACATTAGTACTTGCGTACTTGGCATTAGCTAGCGTATGATCCCTCATTGCTGGAATACAGATGAGACATCT							
3	**ID:** 13619189_length_1045_cvg_4.1_tip_1 **Name:** Iso_18 **New name:** Iso_10 **Genbank #:** KR069006	(AG)26	CGATCTGGGCAGGACAAGAG	60	CCAAGAACTTCCAGCCCCT T	60	215	Through to 2nd round of analysis, and found to be in HWE in one population – accepted for publication and ongoing analysis. Size Range: 187-291 No. Loci: 36
	AACTGGTGGCTAACTGAAGACGACAGCGGTGTCGAGTCAAACAGCTGAATGTTGTTGTGTATCAGTATGTTAGCAAATGCTATGCTCGGTAGGCGAAAGT**CGATCTGGGCAGGACAAGAG**CATGGAAAGGCTGTCGGATTAAACAGATAGAACGA**AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG**GGACAATGGGATGATATTTATTAGATTTGCTTTGAAAAAGAATGAGTTTTCTGACCAAGACCGAACTAAGACGCTAGACATAAATTTG**AAGGGGCTGGAAGTTCTTGG**GTCTTTTGTTCAATCAAGACCACAGCTATGCAACGGTCGGCAAGAGGGTTATCACACAAGGTGGCGCCCTTCGACTTTTCAATTTCGACAAAATGGTGGT							
4	**ID:** 13187499_length_589_cvg_4.0_tip_0 **Name:** Iso_22 **New name:** Iso_17 **Genbank #:** KR069013	(AG)25	CGCCGGGGTCGTCATAATAA	60	GTGGCCCCGAATCATCTCT T	60	196	Through to 2nd round of analysis, but not found to be in HWE- accepted for publication but not continuing for ongoing analysis.
	GCATCACTATCACACCCACGTGATAAGAGCGCATCACTATCACACCCACGTGATAAGAGCGTATCACTTCACAT**CGCCGGGGTCGTCATAATAA**GAGTCCAATAGAGGGTTGGTGGGCATTACGCCGCAAAGAGGGCCGGCATGAATGTGAATGTGA**AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG**ACATTAAAGTTAAAAAAGTGAGCAAGACCATGATAAAAGACCA**AAGAGATGATTCGGGGCCAC**TGGTACCATAAATCCCTCCCTCCCTGATGTTCTTCCTCACATCTCACTCGTTCTTCTTTAATGCATGGGAGGGCCCTCCCAGCTTGGAAACGCCACTGTA							

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
5	**ID:** 13595231_length_985_cvg_6.4_tip_1 **Name:** Iso_3(5) **New name:** Iso_1 **Genbank #:** KR068997	(CT)22	ACTTTCCATCCATGCGCTCA	60	GACTTTGGGTGACACCGTGA	60	270	Through to 2nd round of analysis, and found to be in HWE in both populations tested – accepted for publication and ongoing analysis. Size range: 228-386 No. Loci: 43
	CGTCCGTAGAACGACAAGAGAAGAGGGCCTTGTGTTTGTTGGCTCTAGGTATCGCTACACAAATAGTGCGGCACACGCGTTTCC**ACTTTCCATCCATGCGCTCA**CCGACAGTCAAGCGGAGGAAGCAAACCCACTTCCCCGTCCAGTGACGTCACTACAGATAACTGTGACGTCACATGCGGTGGTTGTATCCTCACTTCACAACACTGTAATG**CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**ACGACAGTAACTGTTTCTAGGACTCGGATAGATAGATTACCTACATAAGGGGTCACCCCTCAGGAATATTTTTGTA**TCACGGTGTCACCCAAAGTC**GGTGCACTCTCAGCACAGATTTACACATTACAGTTACACCGAGCGTACTCTCAGTGCAGACGAGACACCGTTTGAGCACATTGTCTATCTTTACTATAGC							
6	**ID:** 13723986_length_1898_cvg_5.9_tip_1 **Name:** Iso_8(2)	(GA)20	GTGTTTCTCGGCCAGGGTTA	60	ATTTATCGTCGCCCAGCTGT	60	205	1st round of analysis was in HWE but 2nd round had multiple peaks - rejected
	CTCCCCCCTACTTCCCCGTCGTCGCCGTATCT**GTGTTTCTCGGCCAGGGTTA**AATAATGAGCGAGCTCGATTATTAAAACTATCTTCATCATAAACTCGCTACTTGGTTGTTTTCTACCCTTGGAAATAACAAGGGTAATGAAATGCGCTGTGTGAGAGCGAGAGAGC**GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA**GGTGCACAC**ACAGCTGGGCGACGATAAAT**TTTAATGAAAACAAGGACATGAACCCGAGGTGACCTTCGGAGTTCACACTTTCAAGGTCACTAAACTTCCAACTTAGTTTCTGAACAATTGTTTTTTGAT							
7	**ID:** 13638145_length_1103_cvg_4.5_tip_0 **Name:** Iso_14(3) **New name:** Iso_5 Genbank #: KR069001	(AG)21	AGCTCACAGCCTTGTCAACA	60	GTACCCGTATAGGTCGCACG	60	223	Through to 2nd round of analysis, and found to be in HWE in one population –accepted for publication and ongoing analysis. Size range: 206-336 No. Loci: 52

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
	TACGTCCTCTAAGAAAGACACATGTTTCTTGCGTATACAGACGTTTATCCCGACGTTTCTTACAGTTTATCCCGTACAGCGGTTGTATAAACTCTCACAA**AGCTCACAGCCTTGTCAACA**ACTTGAAACGAATTCTTCAAGCAGAACAAATCCCTATCTTGTCTTGCGTGAAAATATCGTCAACACTTATTTAAAATACTACACAGGCTGTTCCGAAACGGAGGGAGAGAGAGAA**AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG**AATTTGTTTCCGATGCGGCGAACTCA**CGTGCGACCTATACGGGTAC**ACATACGGAAGTTAATAGCATTTGCATAGTTAATGCGGTGCTCATCAACACTAAGGTTTAACAAACTGATCGTAAACCTTGCCCCGATTCAGTAAGCCGC							
8	**ID:** 13509499_length_839_cvg_2.8_tip_1 **Name:** Iso_45(5) **New name:** Iso_2 **Genbank #:** KR068998	(TC)26	TAGGCCTACTTACCGACCGT	60	ACCACTACCTCCTGTCTCCC	60	206	Through to 2nd round of analysis, and found to be in HWE in both populations tested – accepted for publication and ongoing analysis. Size range: 178 – 386 No. Loci: 60
	TAATAGTAAAATCGCCAGCCAAGTAGTTTCGAGTGATACAATAATAATACAAACATCTGGAGAGGATTTCAGTAAGCTCGCTAACATAATGCTGTTCTAA**TAGGCCTACTTACCGACCGT**GTATATATACGGGTATTATAT**TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC**CACATCTCTCCACCTCGGTCCATCATTGAAAGTATCTCCATCCTTGTCGCACTCGCTCGAACGTTGTGGGCAATACTTACATATCTCACAACG**GGGAGACAGGAGGTAGTGGT**AACCCTATAATTATAAGCACTACGTCATAATGCTATCGTGTCGGCACATCATGCACCCCCCCTCTCCCCGCCCCCCACGCCCCGGGCGCCAGGGTACTGA							
9	**ID:** 13525693_length_862_cvg_7.0_tip_1 **Name:** Iso_46(4)	(AG)26	AAGGTGGGATTAAGGCCAGC	60	AAGAAGTTCGGCAGCCATGT	60	274	Failed to amplify-rejected
	TAATGGTCTTTATTG**AAGGTGGGATTAAGGCCAGC**AAGGCCTTATCTTCCATCCAACCTTTAACAAATAATAC**AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG**TTGAGTACGAAAGAAAGATGAGAACCAACAGCTTACGTTAAAGAAGTTCGACAGCCGTGTAGTTGGAGATAACATGGATATTTACAATAACAATAGCAATTCATGATGGAATGTTTCGGCTTACGCAGCCTTGATCACCCAACT**ACATGGCTGCCGAACTTCTT**TAAGGTAAGCGGTTGGTTCCTATCTTTCTTTCCCACTCAACAATAAAATAGACGACAGTAATATAAATGATATAGTCTTAGGGTTTAATTAAAACAACTT							

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
10	**ID:** 13239497_length_616_cvg_3.5_tip_1 **Name:** Iso_47-3 **New name:** Iso_3 **Genbank #:** KR068999	(CT)24	CAGCGACAATAGCCTCGGAT	60	CGAACAAGGCGGTGTGAATG	60	145	Through to 2nd round of analysis, and found to be in HWE in both populations tested – accepted for publication and ongoing analysis. Size range: 132-238 No. Loci: 49
	GATACATCTTGTATATTTTATTGGTGTTCTGCTCTACATATATACAAGTCTACTAATCCAGTGGTTAGGACTGTACGCGCGATTTCCCGCATCCTATTAA**CAGCGACAATAGCCTCGGAT**AATTGAGTTTGTCGACAGCTCTTCTAGTTCTAAATATGGACACAGGATGCGTG**CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**CGCC**CATTCACACCGCCTTGTTCG**TTAAATAAGAACACTCGACGAGTCGACGCTAAGCAATGGCTGTGACATTTTTACAC							
11	**ID:** 13286386_length_643_cvg_2.5_tip_1 **Name:** Iso_48-3 **New name:** Iso_9 **Genbank #:** KR069005	(TC)24	TACTATCGTGGTTCGCGCTC	60	GCAGGCCTCGAAACGAATTG	60	247	Through to 2nd round of analysis, and found to be in HWE in one population – accepted for publication and ongoing analysis Size Range: 224-452 No. Loci: 70
	GCGAGGAAACGTTTTGTGCCTCACTCCCGTATCTGTCCTGTCTCAAGACATGTTTCATACTGATGCAAAATATTCGTCTCGACTCGGCCCGCTGTCATCA**TACTATCGTGGTTCGCGCTC**ACATGGATATGCGCCACTTCGAGCA**TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC**GCGTGCGCACATCCAGTGTTTTACTCCGCACAAACTGCCTTCTTGTGTACTGGGATCATCGTTATCTGAGTTACTTGTGCGCGCACATGGAAACAGATAAATTGTGCAGCCTACACTAAATAAGACCAATTTAG**CAATTCGTTTCGAGGCCTGC**TCAGTGTTCACCTTTTTATTTGGCTCACTGCGTTTC							
12	**ID:** 13623575_length_1057_cvg_2.2_tip_1 **Name:** Iso_49-2 **New name:** Iso_14 **Genbank #:** KR069010	(AG)24	GTCGCTTCTGTCGTCAGGAA	60	AACATAGGTGGCTTCGGCAA	60	181	Through to 2nd round of analysis, but not found to be in HWE- accepted for publication and but not ongoing analysis.

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
	TTGTAGGTTTTAAGGGGTCGTATAAATATCTGTCAGTTTTGCGTTTAAACCATTGAATGGGTGAAATTGCTACGGTTGAAATACCTGCTGTGAACTTAGG**GTCGCTTCTGTCGTCAGGAA**TCCAGACAGTGGAATGTGACGCTGCAATTGTGTTGTTTTGATTTAGAGAGAAGGGGATAGAGAAAGAAGAGGAT**AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG**AAATAGAGAATTTGCGTCA**TTGCCGAAGCCACCTATGTT**AATTTCAGACGAGCCATCCAGCCAGCCAGCCAGGCAGCCAAATAACAATAACAAACATATGAATGCCAGCTCACTGTATAATGCATATTATTAACCGGCC							
13	**ID:** 13704996_length_1523_cvg_3.3_tip_1 **Name:** Iso_50-5 **New name:** Iso_6 **Genbank #:** KR069002	(CT)24	TGTGAGAGCAACTGCACAGT	60	ACACCGCTTCGAACACAAGA	60	270	Through to 2nd round of analysis, and found to be in HWE in one population – accepted for publication and ongoing analysis. Size Range: 260-508 No. Loci: 59
	TTAACGTTCTCTACTGCGAATAATCTTTTGTGTTTTAATGGAATAATGCCGAAATAATGAATAATTGTTATCAAATCATTGACCTCCAGGTGCAACATTC**TGTGAGAGCAACTGCACAGT**TAAG**CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**CCTGTCTTCACGCCGTAGGGCGTAGACTGTACTAGAACATAATAATGTTCATTTCTATGGATTTTCTGTTAGTTTATATTCCACTGATCTCCCTTTAGGATCTGGCTCATGACGGTAGCAGAGCTAGCTTAAGTTTTTCACTGCTGTCTTCCTCGATACCGCTCTAGATTGTTCACAA**TCTTGTGTTCGAAGCGGTGT**TACTAGTGAGTTATTAGTACAGTTAATGGCGCATTACCGGCTAGGCTGGGTCATTAGCACATTTGCAATTATTTTCTGACGGGGAAAGTAAAACAAACGC							
14	**ID:** 13348834_length_684_cvg_3.6_tip_1 **Name:** Iso_51-5	(TC)23	CACCACTGCTGACGAGGATT	60	CGACGTACTAGCGAGGTTCTC	60	238	Failed to amplify-rejected
	GCGCATCTTTACGGTTAGCGCTGTTGC**CACCACTGCTGACGAGGATT**CAACGTTCAATGATAAACCAACCAAGCTCATTCGTATAACATCTGTCAGTGTTGAAAACTCTGGT**TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC**TTCACACTCTCCTCGTCAAATGGATCCTTTGATAACCGGAACTCTTTGTGTCAGCAACACTGCAATATGTCAGCATTTTCTCTTGA**GAGAACCTCGCTAGTACGTCG**ATATTTTTACTTAATTAAAACTGCGACTTCTTTTATCACTTATCATTATCAAGATACTTCCCAATGTTGCTACAATTTAACAAAATTGTTAATAGAGTAC							

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
15	**ID:** 13576593_length_947_cvg_2.8_tip_1 **Name:** Iso_52-3 **New name:** Iso_8 **Genbank #:** KR069004	(GA)23	CGTGTTAAACGCGCTCGATA	59	CTGCCTGGTCACTGTGTTCA	60	263	Through to 2nd round of analysis, but not found to be in HWE (depends on test)- accepted for publication and ongoing analysis. Size Range: 256-414 No. Loci: 51
	CTACGTTCTTGCTAAGTGAGTAATCACACATCAATAACGACTTCATACAATTATTTCCTTAATTGGTGTTGATTAGAGAGCATTAACATTTGTTTGATGTCGTGTTAAACGCGCTCGATATTCAACATCACGGTGATATTATTATCGAGGGGGTAACGTTACAAGTTTCACAAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGACAACCAATCTTGGCTTCCACAGATTGTTTGAATTGCCAAATCCGATATTCAGAATGTCGACTTATTCACCACTCTCTTCTCTTCTCTCCTCGTATCGAATCAGCTGCTTCACCAGGCGCCAATGAACACAGTGACCAGGCAGCCACTCCGAT							
16	**ID:** 12974368_length_504_cvg_2.1_tip_1 **Name:** Iso_53-3 **New name:** Iso_15 **Genbank #:** KR069011	(GA)22	AGACCAGCTCAATCGGGTTG	60	GTCTAGCTAGCGGGAGGGAA	60	232	Through to 2nd round of analysis, and found to be in HWE in one population – accepted for publication but not continuing for ongoing analysis.
	ACTGTTTACAATAAATTAACATGTTATTGGCTGTTTGAACAAATGTGTTCTTTTAGCTCGCGTGTTCGTGTTCGGCTGGCATTTCGCAATGTAAATCTGCAGACCAGCTCAATCGGGTTGTTCTAGAAAGCAATTAAAAGTCAATAGAAAGAACGAGCGCGAGAGAGAACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAAGAAAGCAATCACATAACACATCTGAATCCACCATTTTCCCATCCTTCTGGTATTTCACAAACATCAATGCATCACGACTATATATATACACGTATATTCCCTCCCGCTAGCTAGACCCTATGACATCAGATCGTCACAAGCAAATATGTGTGTATATATTTGTGCGACGGCTGCTCCGAATGGAACCAGCCGAACAAC							
17	**ID:** 13629105_length_1073_cvg_3.0_tip_0 **Name:** Iso_54-1 **New name:** Iso_12 **Genbank #:** KR069008	(AG)22	TAACGTGCCATGACGTCACA	60	TAGGGTCCCTGAGGAGAAGC	60	210	Through to 2nd round of analysis, and found to be in HWE in one population – accepted for publication but not continuing for ongoing analysis.

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
	TGTCTGAAATCATCACTGGATAGCATCGCTATCGCGTGCGACAATGACCTACGACATATATATGGTTGAATGTGGAATATACGGCTTAATAACGTGCCATGACGTCACA GAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGTAGCACAATACAAATTGCATACATTGCACAACCGTTTCTATTTGCTCACTCCAA GCTTTTCATGTCCTTACTCTTCGCGATTCCCGAACTAAATCTCTCTAGCATGTGTTAGATTTACCTGCTTCTCCTCAGGGACCCTATTCTGTAACTTTGTCATTCAATGAAT TCAGCTTGAATTAATGTAACAATAGAATGGAAACGCTCGCTTCGCTCGCTTAGCTTTACAACACTTTCATTTCT							
18	**ID:** 12964466_length_501_cvg_1.1_tip_1 **Name:** Iso_55-1	(AG)21	CGGAGCTACGAGACACAGAC	60	TGCCTGTTGTACGGATGCTT	60	220	No samples successfully amplified in 1st round of analysis -rejected
	ATTATCATACACTTACTTGTTTTCTATCAACGCGGACGTAGATTGACTCAAGAGACGTTTGAGACATAGCAACGAACCAACGATGCTCGGCTTTTCTACACGGAGCTAC GAGACACAGACAGAGAGACAGAGAGTCAGACAGACCGACAGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGGTGGAGAACAAATTT ACTCCGTTCGCTGCTTTGAATTAATGGCATCAACCCTTAATCGACCAGAGAAGACAAACATTCAAATTAACAACCCCCTTCCGTCATCGCAAGCATCCGTACAACAGGC AGACAGCCCGGGGAGCCAGGGGGTGATATTACACGTCACGCGTCACGCGTCACACGTCACACGTCACGCGTCACACGTCACACGCCACACGTCACACGTCA							
19	**ID:** 13226837_length_609_cvg_3.2_tip_1 **Name:** Iso_56-3	(TC)21	AGGAAACTCGTTACACGCGT	60	AGCTTGTCATGTTCCCACGT	60	132	1st round of analysis was in HWE but 2nd round had multiple peaks - rejected
	GGCTTGAAACGTATATATTTCTTCTCTTTTTATTATTTTTGTTAGGAAACTCGTTACACGCGTTGCGATTATTTTCATTTCTCAGTAGAATGTGCCTAAATGCGGTCTCTCT CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCCGCGTGTCACGTGGGAACATGACAAGCTGCGCCCTACATTCAACCCTACAACGGGGGTAGTTAGTATCTAGTTTGG TTTAGCGAGGGATAATGAAAAACGTTCCCCGCCCGATCTGTGTTGCTGTATT							
20	**ID:** 13428983_length_750_cvg_3.2_tip_1 **Name:** Iso_57-3 **New name:** Iso_16 **Genbank #:** KR069012	(GA)21	TTCCCCTAGCCTCCGATCAA	60	TTTCGGGAAGGTTGTTGGGT	60	198	Through to 2nd round of analysis, and found to be in HWE in one population – accepted for publication but not continuing for ongoing analysis.
	TATTTCACCCAGCAAGGTACATATCACAAGTGCTGGATCGTGCCTTGGGCTGATTCAAGCTGAGAAGTGAAGAAAGATTGGCATCATTGATTCTTCCCCTAGCCTCCGA TCAACACTTTTCTCATCTTTCACGAGAGAGCGAGAGACAGAGCGAGAGAGAGAGAGCGAGAGACAGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA GAGAGAGACATACAGCGGTATAAACCACCAGAGACAGTGAATGGCCGCGAAGAGTATAAACCCAACAACCTTCCCGAAAAGATTGAAAACGAAAGTCCTTTATTTTG TTTCCCCATAAACTGGTAGTTTACCCTATTTTCTCTGCGTAGTAAACGAATATGTATTTATGAATACCATCA							

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
21	**ID:** 13521065_length_855_cvg_3.0_tip_0 **Name:** Iso_58-4 **New name:** Iso_11 **Genbank #:** KR069007	(TC)21	CCACTTTGCCTTATGCGTCG	60	GAGAAAGGGAGAGAGCGCAG	60	165	Through to 2nd round of analysis, and found to be in HWE in one population – accepted for publication but not continuing for ongoing analysis.
	AAGTTACTATGTCAGAGGATGGAGGGGTAGAAGGGAAAGCTATCGCGAAGGTAGACTGGGACTACGTTCAATATGGGTTTATTTCTTCTCCCCTTCCTTC**CCACTTTGCCTTATGCGTCG**TCTGCAAGGATGATCCAATCCACACACGATACCGCACTACATCGTAGACGTGATAATT**TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC**ACACACACATACACACACCCTCATC**CTGCGCTCTCTCCCTTTCTC**CCTTTTTCTCTCTCTCTTGTAACTGAGTGGGTGTAGACTATGTACTTTTTACTATGACTGACTATATGTGACTATATCGTATACACTAGCAAATGTATTT							
22	**ID:** 13664003_length_1207_cvg_3.4_tip_1 **Name:** Iso_59-1	(AG)21	GCAGGTCGACTTTTCGGGTA	60	GCCACGTGATTGCGTCATAC	60	145	No samples successfully amplified in 1st round of analysis -rejected
	GCAGAAACTCGAACTTTAGGCGCGATACCTTTGGCGGTAAAGGCGATAAGATAACGCGGTATCAGTACACGTGTTAGGGGTTATCTACTGTTGAGTTTCT**GCAGGTCGACTTTTCGGGTA**TCAACATCACCCATTCATCGTCTCGATTAGCATCGTGAAGAGAGAA**AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG**CGCTAGCGAGATAGGAT**GTATGACGCAATCACGTGGC**GCGCTCTATGGAGCGTGGCTCGCGACAGGTCGGCAATGTCCGGTCACGCCCGGTCGGGTTCGGGGCGCGTGCCTGCCTCGACGCGGCCTCGTCCCACGAC							
23	**ID:** 13065885_length_537_cvg_3.6_tip_1 **Name:** Iso_60-4	(GA)20	AGGGGGAAAACAGGAGGAGA	60	TTGTCGGTGGGAGTGAACTG	60	164	No samples successfully amplified in 1st round of analysis -rejected
	GGCGCAACTTTTTTGGAATTTCTTATTGATATCTTTGCCGTACGAATATTTGAAACGTGTTGTCCTTGTTTTTAATTAAAAACAGTGGGGGAAAGAGAG**AGGGGGAAAACAGGAGGAGA**GAGAGAAAGAAAAGCGGAGCCGGTCTTCCGTCACGACGGGGTAAAATCAAACGCCAATCATTTTGGGCGAGAC**GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA**ATTGCATGCTC**CAGTTCACTCCCACCGACAA**AGCCGTTACTTCAACCTAGTTGCGTCTCAACTCTGTCTGTCTACATTCTTGCCACCGTCACTTCCTTCACACTTGCAGGGATTCGGCTGTGTTGATGTTC							

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
24	**ID:** 10071274_length_164_cvg_2.5_tip_0 **Name:** Iso_61 **New name:** Iso_7 **Genbank #:** KR069003	(CT)22	TGTGTTCCGGATGCAAGTAG	60	GATGCAAACGCTCGATAAAC	58	131	Successfully amplified for two sites, in HWE for one population – accepted for publication and ongoing analysis. Size Range: 120-266 No. Loci: 60
	TCAGG**TGTGTTCCGGATGCAAGTAG**CACATGTGC**CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**CGCACACAGCATCAAGATCCTCACGTACGAGGATCATA**GTTTATCGAGCGTTTGCATC**TGGCGAATTATTTAAGTACAGCCGATCT							
25	**ID:** 10519683_length_190_cvg_1.0_tip_1 **Name:** Iso_62 **New name:** Iso_18 **Genbank #:** KR069014	(GA)21	TGAATGATGGTGTTCGAAGTG	60	TATTGACAGCACCGACGTTG	61	174	Successfully amplified for two sites, and found to be in HWE in one population – accepted for publication but not continuing for ongoing analysis.
	CGATCAA**TGAATGATGGTGTTCGAAGTG**TATGAATAGAATAGACCATCGCACCT**GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA**GCATCATTACAATGCATGTCTAAGAATAACTCAACTGTCACCCAGCTGTCAACTTTCTGGACAAA**CAACGTCGGTGCTGTCAATA**AATGACGTC							
26	**ID:** 11337450_length_237_cvg_6.0_tip_1 **Name:** Iso_63	(AG)25	AGAGACGAAGCGATTGAACC	59	GAAGGAGCTTCAGGGATTCC	60	183	Poor amplification - rejected
	GGG**AGAGACGAAGCGATTGAACC**CTCGCTAGGCGGGGCTGGTCAAGAGA**AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG**ACTTACCACTTGTGAACCACCGCCTGGTTGTGGAATGGAGCGAGGTAGCTTCCCACCGCCCTCGCAA**GGAATCCCTGAAGCTCCTTC**ATGTACAACGAACACTGGCCGCTCGGACTTGCGTCCCTGCAAACGCAC							
27	**ID:** 11881420_length_291_cvg_1.9_tip_0 **Name:** Iso_64	(CT)23	TCTGGAGGTCCCAAGTGTTC	60	AATTAATGTATCCGGCGTGC	60	156	Poor amplification - rejected
	GTCTCGTCATTCCG**TCTGGAGGTCCCAAGTGTTC**AGCGCATAGGAACG**CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**CGGCGGGGATCGAGCCGCAGAACACAAAGCATACGCAACACTTGTTTACAAACAAT**GCACGCCGGATACATTAATT**AAACATGACAGACCTCACAG							

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
28	**ID:** 12763102_length_443_cvg_3.0_tip_1 **Name:** Iso_65 **New name:** Iso_4 **Genbank #:** KR069000	(TC)23	AACTGATCGCGAACTCCAAG	60	GTTCTTCACAGTCGGCTTCC	60	168	Successfully amplified for two sites, in HWE for both populations – accepted for publication and ongoing analysis. Size range: 160-356 No. loci: 60
	TATATGGGGCACAGACATAGACACAGACCAGCCGTGCAAATCAGCGATGTAAGTGA**AACTGATCGCGAACTCCAAG**TT**TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC**TGCTCCGTTGCAAACATAAATATATTCAAAAGCAAGCAAGGCAAGCGGGCATCCATCTTGTAGCGCGAGTCAAGTGGAGA**GGAAGCCGACTGTGAAGAAC**AAATGACGGTGTCAAGACGCTGTGTTTCGCCTGTCCCGAAACCGGATATAGATGGCGTTGATGTCCCGACCGGAAATCGCGTCGATTTGTTCAAACATGG							

Baetis rhodani

Table D3. Details of microsatellite development for *Baetis rhodani*. *B. rhodani* was the only species used in this thesis that already had microsatellites developed. No's 1 to 13 shows microsatellites that were previously designed by Williams (2003). The microsatellites that were developed were initially named B_1 - B_15. This table contains the microsatellite number (*No.*) in order of testing; original *Contig ID* from Next Generation Sequencing (B_1 - B_15 only), the *Name*, *SSR* (Simple Sequence Repeat), *Forward primer*, Annealing temperature (*Tm*) of the forward primer (B_1-B_15 only), *Reverse primer* and corresponding *Tm*, (for previously developed microsatellites the Tm is not per primer, it represents the published annealing temperature); the *Product size* of the amplified region; and details of the *Outcome* from testing this microsatellite. If known the corresponding sequence can be found below each microsatellite with the forward primer highlighted in green, the SSR highlighted in red and reverse primer highlighted in blue and all underlined.

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
1	**Name:** Brh-1 (Williams *et al.* 2002) **Genbank #:** AY081168	(CA)11	GTGCGTTACATACGAAT AAGTG		CGTGTTAGAGAAAATGAG CC	61	233	Tested: 173/182 HWE: 2/9 Used for ongoing analysis
	CGAA**GTGCGTTACATACGAATAAGTG**TACGTAGCAAGAGAGGAGTACGTGCCGCCGGAGCGCAGCCCGAGAGCTACTTTGCGGTGGTGGTATAGTGGTGGTGCGAC GTTMTTGCACGTTGCATACTCGGCAGGCGGAATGTTCG**CACACACACACACACACACACA**GCACCGCCACCTCAACTCTCTACGTATTAGTATGTATATTGTATGCACG CC**GGCTCATTTTCTCTAACACG**GCGCGCATTAGGCCGCACAACAAACTGCCGCCAGGAGACTTTATACACACCGAGTCTTTTGTTTACTCGCCCCGTGCTCACACACACC ATCGGCTGAATGGCGCGTACACAAACGTGCAT							
2	**Name:** Brh-2 (Williams *et al.*, 2002) **Genbank #:** AY081169	(CT)18	ATTGCCGACAAAACTCG C		CGTCTCATTGATATACCCT GC	61	177	Tested: 178/182 HWE: 6/9 Chosen for ongoing analysis
	TGCAAAGAGCACCGTGTTTATCTTGTCACATTTAAACG**ATTGCCGACAAAACTCGC**TGGGGATTTCTTGGCAAACACGCAGCCAATGAATAGAGCCGTATGTCTCTCTG CTGCGTGGGCGGTTCTCGTCGCTCCGACTC**CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**CCCGTGTTTGTCTAGCCGC**GCAGGGTATATCAATGAGACG**CGCAT AACCCGATTATAGCGGTACCAATCTCTGT							

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
	Name: *Brh*-2-IK (redesigned by (Alp *et al.* 2012)	(CT)18	TTTCTTGGCAAACACGCAG		ACAGAGATTGGTACCGCTATAA	61	187	Not tested because same microsatellite as above and that was successful
	TGCAAAGAGCACCGTGTTTATCTTGTCACATTTAAACGATTGCCGACAAAACTCGCTGGGGA**TTTCTTGGCAAACACGCAG**CCAATGAATAGAGCCGTATGTCTCTCTGCTGCGTGGGCGGTTCTCGTCGCTCCGACTC**CTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT**CCCGTGTTTGTCTAGCCGCGCAGGGTATATCAATGAGACGCGCATAACCCGA**TTATAGCGGTACCAATCTCTGT**							
3	**Name:** Brh-3 (Williams *et al.*, 2002) **Genbank #:** AY081170	(CA)12	ATGCAGTGAATGAGCCGC		CACTTTGAGCCAATCTTAATAGAGC	57	234	Tested: 178/182 HWE: 1/9 Used for ongoing analysis
	GTCTCGT**ATGCAGTGAATGAGCCGC**ACCGGCGTAAAATTAGAATTTACCTCATTCAAACT**CACACACACACACACACACACACA**CGTCGACTTTCCTGGTGTGTACCACTCCTGTGGAGATTTTATACTTTCAATCTGCAATTCGGCGCCAATTGTGATTCACGTTTTTAAACGGCGAGTGCGTGATTAATAGTATGGTGTATGAACTACCTATTT**GCTCTATTAAGATTGGCTCAAAGTG**CGCGCACTGTTTATATTATGGC							
	Name: *Brh*-3-IK (redesigned Alp *et al.* 2012)	(CA)12	GGCGTAAAATTAGAATTTACCTCAT		ATCACGCACTCGCCGTTTA	57	159	Not tested because same microsatellite as above and that was successful
	GTCTCGTATGCAGTGAATGAGCCGCACC**GGCGTAAAATTAGAATTTACCTCAT**TCAAACT**CACACACACACACACACACACACA**CGTCGACTTTCCTGGTGTGTACCACTCCTGTGGAGATTTTATACTTTCAATCTGCAATTCGGCGCCAATTGTGATTCACGTTTT**TAAACGGCGAGTGCGTGAT**TAATAGTATGGTGTATGAACTACCTATTTGCTCTATTAAGATTGGCTCAAAGTGCGCGCACTGTTTATATTATGGC							
4	**Name:** Brh-4 (Williams *et al.*, 2002) **Genbank #:** AY081171	(GA)21	TAGCCCCATTTGCCTTTG		GGTTTTATGATGTGTTCGGTTG	61	213	Tested: 182/182 HWE: 9/9 Chosen for ongoing analysis
	AAAAGTTTTTCCCAGGGATAATCTGCCGGCTGACACAGC**TAGCCCCATTTGCCTTTG**TTTTATTTCGGCAGCGCGAAAAGGAGCGAGCGAGAGCTCTCCTCCGCTCTGCCA**GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA**GTGATATAAAATGTCAGGCTAAAATGCCTATATAGTTGGCCGCTCTTGTTCTTCTGCAGCCGAGCGGCGCATGCATG**CAACCGAACACATCATAAAACC**GAATTGGAAGCGACATT							

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
	Name: *Brh*-4-IK (redesigned Alp *et al.* 2012)	(GA)21	GAAAAGGAGCGAGCGAGAG		GAAGAACAAGAGCGGCCAAC	61	133	Not tested because same microsatellite as above and that was successful
	AAAAGTTTTTCCCAGGGATAATCTGCCGGCTGACACAGCTAGCCCCATTTGCCTTTGTTTTATTTCGGCAGCGCGAAAAGGAGCGAGCGAGAGCTCTCCTCCGCTCTGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGTGATATAAAATGTCAGGCTAAAATGCCTATATAGTTGGCCGCTCTTGTTCTTCTGCAGCCGAGCGGCGCATGCATGCAACCGAACACATCATAAAACCGAATTGGAAGCGACATT							
5	**Name:** Brh-5 (Williams *et al.*, 2002) **Genbank #:** AY081172	(TC)15	AAACGGTCTCTCTCTGTCTG		CTGTCAAAACGAACAAACC	61	220	Tested: 181/182 HWE: 8/9 Chosen for ongoing analysis
	CGAGATAACAAAAAAGTGAATTGCACAACGAGCTGTTGGAAAACGGTCTCTCTCTGTCTGTCTTTCTCAATTTATTTATATATTCGTGCCATCAATTTCATCCTTGGTGCACACACGCGGAGACACGAGAGATAATATAAGAGAGATGAAACGCATCGAGTTATGCTCTCTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCCCTCATCATGTGGCTGCGGCCTTGGCTGCATGGTTACATGGTTTGTTCGTTTTGACAGATAAGCGCGCGCAGAGTGAAAAGAGAGGAGCGCGCTTAAGAGATTAGTTTGGCTTGCAACACACTTCATT							
6	**Name:** Brh-6 (Williams *et al.*, 2002) **Genbank #:** AY081173	(CT)26	GGATGGATGGATGGATGC		CAAAGTTGGTCAAAGGCG	57	151	Tested: 182/182 HWE: 8/9 Used for ongoing analysis
	GATACGACACACTCGGCGCTTGCCAACCCGCCCACAAAAGCAAACCCGGGCGCCAGCACCAGAATTCGCGCTGGATGGATGGATGGATGGATGCGGCACCGCCGCGGCCGTGATTAATGATTGGAGTGACGCGCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGGCATAATTATATTCATGCGGCGCCTTTGACCAACTTTGATTAAGCCACCGCAAACTTTTGCCGGGCCC							
	Name: *Brh*-6-IK (redesigned Alp *et al.* 2012)	(CT)26	CGCCCACAAAAGCAAACC		TGCGGTGGCTTAATCAAAGT	57	213	Not tested because same microsatellite as above and that was successful
	GATACGACACACTCGGCGCTTGCCAACCCGCCCACAAAAGCAAACCCGGGCGCCAGCACCAGAATTCGCGCTGGATGGATGGATGGATGGATGCGGCACCGCCGCGGCCGTGATTAATGATTGGAGTGACGCGCACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGGCATAATTATATTCATGCGGCGCCTTTGACCAACTTTGATTAAGCCACCGCAAACTTTTGCCGGGCCC							

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
7	**Name:** Brh-7 (Williams *et al.*, 2002) **Genbank #:** AY081174	(CA)15	TGTGTGTAACAAGAAAA CGC		GATTCGTCCCTTATGTATT AGC	57	174	Tested: 181/182 HWE: 7/9 Used for ongoing analysis
	AAAGAGAGACGATGTGTGTAACAAGAAAACGCGTCGCGCATCAGCCGCGGCGGCGGTCGCCGCCGCCGTCGTCGTCACACACACACACACACACACACACACACACAGAAGCTGCCGAGAGCGAGTTTTGGCTCTCTAGATATGCGTATAGAGAGAGGGTGAGATTGCTAATACATAAGGGACGAATCCGCCGCCGCCGCTCCACCGCTGCTCAACAGGGAGATGACAAAATCC							
8	**Name:** G7 (Williams 2003) No Sequence	(GT)12	TGCAAACTGACATGAAT GC		ACATTGATGCCCTCAAGC		177	Tested at a range of Tm and rejected due to poor amplification
9	**Name:** G4 (Williams 2003) Sequence Unknown	(GT)24	TGTTTAGCTGCTTGTCGG		CACAATGCACATGACACTT CGC		126	Tested at a range of Tm and rejected due to poor amplification
10	**Name:** A4 (Williams 2003) Sequence Unknown	(GT)24	GGATTCGTCCCTTATGTA TTCGC		CCGACAGTTTAGATTGCTT TCCTC		217	Tested at a range of Tm and rejected due to poor amplification
15	**Name:** F11 (Williams 2003) Sequence Unknown	(CA)12	AATTGAAAGCATCCGTG C		TTCACTGGACATCGTCGC		174	Tested at a range of Tm and rejected due to poor amplification
11	**Name:** H2 (Williams 2003) Sequence Unknown	(CA)13	GAGACGAAGGACAAAG AAGTCGC		TGGGAAGTTGGAAGGTGT TGC		171	Tested at a range of Tm and rejected due to poor amplification
12	**Name:** G2 (Williams 2003) Sequence Unknown	(GT)16	TTTCCAGCAAATGAGAG C		TCGTCGGCAAACTTAATG		107	Tested at a range of Tm and rejected due to poor amplification
13	**Name:** G3 (Williams 2003) Sequence Unknown	(CA)12	TGCCCAAGAGACAAGAG AG		GTGTGTGTTTGCCTCACG		142	Tested at a range of Tm and rejected due to poor amplification
Newly Designed Microsatellites								

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
14	**ID:** 2513711 length 511 cvg_5.4_tip_0 **Name:** B_1 **New name:** B_12 **Genbank #:** KR069047	(GA)28	AATTCACATTTGCGTGATGG	59	TCATCATTCTTCTTGCTGTCAAA	60	274	Success rate variable, out of HWE for both populations- accepted for publication but not ongoing analysis
	TCAGCCGTGGTAATAATTCACATTTGCGTGATGGCCCCTATGAGCTTTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAATCTCGCCGGGAATCCAATTTTCACGCACTTTCGATAGGGGTCAAATGGATGCTTCAATTTTACGAACTGTTTTTTAATAGTTTATTCATATAAAGTCTTGATAATACCCATGTCAAAACTATCCTTTTCCAAATTTAAGCTACTATTTTTAGCAAAGTATTTGACAGCAAGAAGAATGATGAT							
15	**ID:** 1750168 length 204 cvg_4.7_tip_0 **Name:** B_2 **New name:** B_1 **Genbank #:** KR069036	(CT)26	GCGCGCTCGAATTTATAATT	59	TCGTACATAATAATAATCGGGCA	58	148	In HWE in both populations – accepted for publication and ongoing analysis
	TATCTCCGGCGGGAGAATTATTGCGCGCTCGAATTTATAATTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTGTGGCGAGAAAGAGCTAGCGCCTTATCTCTTTCAAGAAGCTGCGCGCGCTGCCCGATTATTATTATGTACGACTGTAAATTATACTGCTCCCGCAAAAGTAT							
16	**ID:** 2212503 length 319 cvg_10.1_tip_0 **Name:** B_3	(CT)25	GCGAGCGAGAGAGAGAGAGA	60	TTGAATGACGTCTGAATGCG	61	207	Gel. Failed to amplify - rejected
	TTTCCAGCCACGCGAGCGTCTCTCCCGTTCGCGCTCGCTCGCTGCTTGCACACACAATGTGAACTCTTGTGAGCGAGCGAGAGAGAGAGAGAGAGCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCGCTCACTCACTCAGTCTGCCAGCGAGCGCTCTCGCTACTTACACTAATAGCGTTGTCAGTCAGGGCACAAATGTCAACATCACTTTTTATCTCGCTCTCTTTCTTCCCCAGCCGCATTCAGACGTCATTCAAGCAACTTTTTCATTCAATTTTCCCCTCTTTAAATGTGTG							
17	**ID:** 1743242 length 203 cvg_6.7_tip_0 **Name:** B_4 **New name:** B_9 **Genbank #:** KR069044	(GA)24	TCGCACCAGAACAAGACATC	60	CGCCTCACACTTTGAGTGAA	60	168	In HWE in both populations – accepted for publication but not ongoing analysis
	TTACATTTTCGCACCAGAACAAGACATCTAAATCTCATTAGCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGGAAACTTGTGAATGCTTGAGTTGCCGAATCCACAAGTTTTTTCAACTTTCCGTGCGACAAATTTCACTCAAAGTGTGAGGCGTTATGGATTTTTTGTCGTGGCGTT							

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
18	**ID:** 2089702 length 276 cvg_7.9_tip_0 **Name:** B_5 **New name:** B_3 **Genbank #:** KR069038	(GT)24	GGCGCCGAATTATTATC ATC	59	TTATAAACGGCGCCAAGA TC	60	206	In HWE in both populations – accepted for publication and ongoing analysis
	CAGAC**GGCGCCGAATTATTATCATC**ATCATCAACACTGCCG**GTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT**GAATTTGTCAGCTTGCATG TATATGTACATGCACCAGTCTGAGCGGATTACACGCCGCTATACACGCACGCATGGGCCGCCTTGTTATTACTCTCTGCGTCT**GATCTTGGCGCCGTTTATAA**TCGTGTC TTAAGCCTGCGAAATTTAGCGACTGCATTAACTGTGCTGCATCAACGGACACGCCGC							
19	**ID:** 2830165 length 2602 cvg_8.3_tip_0 **Name:** B_6	(TC)24	CGCGTTGTTGATTTCATG TT	60	GTCACCACCATTGCATCAA G	60	233	Gel. Failed to amplify - rejected
	TCAGTGGAATTGGGTCGAGCATGCAAATGCACCAGCG**CGCGTTGTTGATTTCATGTT**GCTACTCTACTTCTTATACAACTGCCTT**TCTCTCTCTCTCTCTCTCTCTCTCTCT** **CTCTCTCTCTCTCTCTCTCTC**CATCAAATGCGTTGTTAAGGGGAATTTCTGCAAATGAAAGATACTAAAAGTATTCTTTGAGCTAAAAGTAGGCATATTTCTGGTTTCTGG ACAAGATAAGATACATTTTTATAATTT**CTTGATGCAATGGTGGTGAC**CACTGACCAGATAGAATATATAGTAGGAATGGACAAAACTTTCAGTTTTTGGTCCAATGGTC GACCGTAAAAATAATTTGTTGAAACAAAAAGTCACATTTAAGTTAACTTTGTTTCCTATTTGCAATATTTTGTTCGATTTTAAAAATACAAAAATTATATTTTCAATGTTAG CCTCACTTTTTATACAAATTATA							
20	**ID:** 2060120 length 268 cvg_7.0_tip_0 **Name:** B_7 **New name:** B_7 **Genbank #:** KR069042	(GA)23	GCAGAAAGAGCAGCGCT TTA	61	GGCACGGTGTTTTAATTGC T	60	255	In HWE in one of the two populations – accepted for publication and ongoing analysis
	T**GCAGAAAGAGCAGCGCTTTA**ATGAAACTGGACTTGTGTGATCAACATC**GAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA**GCGAGAG TGTGGGTTGCGCAATTAGCATAAGCACGTCGTGGATAAACGTTTTAAATGTCACAGGGGCTGCAACATTATACAATCTCTGGCAATTATACAGATCGGTATCATCCCGG CCGAGACAATATGCACGAGCCGAAA**AGCAATTAAAACACCGTGCC**GAGCGATGCCG							
21	**ID:** 1831274 length 217 cvg_5.4_tip_0 **Name:** B_8 **New name:** B_4 **Genbank #:** KR069039	(TC)22	CCAATGCCAAATATGTTC CC	60	ACGACACGCACTCTTGGA A	60	181	In HWE in both populations – accepted for publication and ongoing analysis
	TGAATATTTTGAATTCTTGACCATTTGAC**CCAATGCCAAATATGTTCCC**AAATACGCCATCAAGACTCTT**TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC** ACTTACAGTTGCAAATAATGGAAGGAATTGTGCAGTGCCTATACACACCATTGTGCGCAACGCCAGATGTTTTTATT**TTCCAAGAGTGCGTGTCGT**GCCGCC							

No.	Details	SSR	Forward primer (5'-3')	Tm (°C)	Reverse primer	Tm (°C)	Product size	Outcome
22	**ID:** 2002316 length 253 cvg_5.8_tip_0 **Name:** B_9 **New name:** B_2 **Genbank #:** KR069037	(AG)22	GCACACAGCATCAAGCA AAT	60	GCATATGCAGCCAGTCTTC A	60	234	In HWE in both populations – accepted for publication and ongoing analysis
	GCAGCGCACACAGCATCAAGCAAATAGCTTAGACTTTGCATGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAACGAGTGTGTGTGTG CGCTGGGGGAGGAAATTGTAAAGAGAGAGAGAGGGTGCAGAGCGGTGTGCGCGCTGTCTTATTTCCCAACCCCACAGCTAGCGCACGCCGGGATTAAGTTTTTCC GGGGGCTGCTGAAGACTGGCTGCATATGCGAATGAGCGCAGC							
23	**ID:** 2095284 length 278 cvg_8.0_tip_0 **Name:** B_10 **New name:** B_11 **Genbank #:** KR069046	(GA)22	GCGTTTTCCTCCATCCCT	60	TATTATGTCGCCGGCAAAA T	60	189	Out of HWE in both populations – accepted for publication but not ongoing analysis
	CCTGCGGCGGTTCCGGCCGCGCGCAGCAACCCTCGTCGCGTTTTCCTCCATCCCTCCGGCATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA GAGAGCGATTACCAGATAGCTGCATCAAAATCACGCGGTGATGAATGCGCCGCCGAATTTACTTTTCCACACGGCCAAGCAGATTATTTTGTGTCACTGCACATTTTGC CGGCGACATAATATAAAGAGTGAAT							
24	**ID:** 1684850_length_195_cvg_4.0_ti p_0 **Name:** B_11 **New name:** B_5 **Genbank #:** KR069040	(CA)26	GATTCGTGTTATGCATTG CG	60	ATAGCCGACGAGCATCAA AC	60	165	In HWE in one of the two populations – accepted for publication and ongoing analysis
	TTAGCTTCGTGTGCATGTTGGGCGATTCGTGTTATGCATTGCGCGCACACACACACACACACACACACACACACACACACACACACACACACACACACACGTGAATAGA AAATACGCGCGGCAGTGGTAATTAAATTTGGATGATGCAGTTGGTGCATTTAGTTATTTTTGTTTGATGCTCGTCGGCTATGAAAAG							
25	**ID:** 1532350_length_176_cvg_5.0_ti p_0 **Name:** B_12 **New name:** B_8 **Genbank #:** KR069043	(GA)23	CGGTGCTCCCTCTTACTC TC	59	TTTAATTAACGGACGCCCT C	59	143	In HWE in one of the two populations – accepted for publication but not ongoing analysis

<table>
<tr><th>No.</th><th>Details</th><th>SSR</th><th>Forward primer (5'-3')</th><th>Tm (°C)</th><th>Reverse primer</th><th>Tm (°C)</th><th>Product size</th><th>Outcome</th></tr>
<tr><td></td><td colspan="8">GTCTCCGGCGCGCGCGGTGCTCCCTCTTACTCTCCTATTAGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGCTGCTCAAATATAATT
AAAACAAGCGCGTCTCTGCGCTTACACTCGAGGAGGGCGTCCGTTAATTAAAACCCCAGCACACTGTAT</td></tr>
<tr><td>26</td><td>ID: 2630505_length_664_cvg_8.3_tip_1
Name: B_13
New name: B_10
Genbank #: KR069045</td><td>(GA)20</td><td>GCTTGAGAATTGCATTCACG</td><td>59</td><td>CGGTGAGCAAATGGTGTATG</td><td>60</td><td>158</td><td>In HWE in one of the two populations – accepted for publication but not ongoing analysis</td></tr>
<tr><td></td><td colspan="8">CGGATGCTTGAGAATTGCATTCACGTTTCCTCGCGCGCCATGAGCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAGTGTGAATTATTTAT
GTGTTTGGCGCGAGAGTGGAAAGGCTTAAAAGGCACACACATACACCATTTGCTCACCGGGCAAACGAGCCAACGAGACGATGCGTAAAGAGAGGTGCGGAAAATT
TATATATACGATTAACAGTTCTTTGATGGTTGTTTACATTGACACTGCCCTTTTGAAATGTGTCTTGTATATGTGGACGTGCGTGCGCAAGCAGCATGCAACAAATTGAT
TGATAAACGATTGAGGA</td></tr>
<tr><td>27</td><td>ID: 2836013_length_4915_cvg_7.7_tip_1
Name: B_14
New name: B_13
Genbank #: KR069048</td><td>(GA)20</td><td>CTTGACTTATTGCACGAAATCC</td><td>59</td><td>CCACGCCTATATGATCTCCC</td><td>59</td><td>180</td><td>Out of HWE in both populations – accepted for publication but not ongoing analysis</td></tr>
<tr><td></td><td colspan="8">TTATTTAAACATACTATAGTCCCTGGAATCGATTAAAATTACAAACTTGACTTATTGCACGAAATCCTAATTACCTAGCGGGAGAGAGAGAGAGAGAGAGAGAGAGA
GAGAGAGAGAGAGAGGAAAACACCCAGGCAGTTTGGAGTTCAATCCAAGAGTAAATAATAGTAAATCTTATCTTTATCTTTATTTCACACAAAGAGAGGGGAGATCA
TATAGGCGTGGAGTGGGGAACATATAGAGGTACGGAGGCGCAAGTGGCGCGCACAATAATAATACCTGATAATATAATTTGTACACCGTTGGTGTTTAATAAAAGTT
TAATAATAATAATATAAT</td></tr>
<tr><td>28</td><td>ID: 1591520_length_183_cvg_4.7_tip_0
Name: B_15
New name: B_6
Genbank #: KR069041</td><td>(CT)20</td><td>ACTTGTATTCCACGCTGCG</td><td>60</td><td>TCGGGTTGCAAGATACACAG</td><td>60</td><td>146</td><td>In HWE in one of the two populations – accepted for publication but not ongoing analysis</td></tr>
<tr><td></td><td colspan="8">CACCAACTGCTCGACGCGTCCAAAATTACCAGCGACTTGTATTCCACGCTGCGCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCGCCGCTGAAGATAAAG
CGAAAAGAGACCATTTTTGCCTCGTCGCCGGGCCTCATTTCGCATTCTCTGTGTATCTTGCAACCCGATCT</td></tr>
</table>

References

Alp, M., Keller, I., Westram, A. M. and Robinson, C. T. (2012). How river structure and biological traits influence gene flow: A population genetic study of two stream invertebrates with differing dispersal abilities. *Freshwater Biology* 57:969-981.

Williams, H. (2003). The genetic structure and dispersal of the mayfly baetis rhodani in relation to acidication and inter-catchment distance. Thesis Ph.D., Cardiff University.

Williams, H. C., Wilcock, H. R. and Bruford, M. W. (2002). Microsatellite loci for the mayfly Baetis rhodani (Baetidae, Ephemeroptera). *Molecular Ecology Notes* 2:411-412.

Appendix E - Aquatic Conservation Manuscript

Appendix E consists of the manuscript submitted to *Aquatic Conservation: Marine and Freshwater Ecosystems*, describing the final microsatellites designed for each species.

Table of tables

Table E1......369

Table E2......376

Author Contributions

Hannah C Macdonald	All data collection and analysis, wrote first draft of manuscript and incorporated comments
Steve J Ormerod	Comments and advice on manuscript
Michael W Bruford	Comments and advice on manuscript

Appendix E - Aquatic Conservation Manuscript

RESEARCH ARTICLE

Title

Enhancing capacity for freshwater conservation at the genetic level: a demonstration using three stream macroinvertebrates

Authors

Hannah C Macdonald*, Steve J Ormerod and Michael W Bruford

* Corresponding author.

Affiliation

School of Biosciences, Cardiff University, Cardiff, UK

Cardiff University

The Sir Martin Evans Building

Museum Avenue

Cardiff

CF10 3AX

Email: BurtonHC@cardiff.ac.uk

Abstract

1. Species diversity is declining more rapidly in freshwater ecosystems than in any other, but the consequences for genetic diversity, and hence evolutionary potential, are poorly understood. In part this reflects limited use and development of modern molecular tools and genetic approaches to address conservation questions in rivers, lakes and wetlands. As widespread, diverse and functionally important organisms, freshwater macroinvertebrates are ideal candidates for genetic approaches to reveal, for example, the conservation consequences of demographic histories and past disturbances. However, the availability of microsatellite markers for this group is very limited.

2. Using next generation sequencing, microsatellite markers were developed for *Isoperla grammatica* (Poda, 1761), *Amphinemura sulcicollis* (Stephens, 1836) and *Baetis rhodani* (Pictet, 1843) to enable conservation genetic investigations of these widespread invertebrate species. Fifty-two robust microsatellite loci were developed (18, 21 and 13 per species), all with high levels of allelic diversity (7-27, 3-16, 5-13 alleles per loci, respectively).

3. These tools will allow assessment of genetic structure, dispersal and demographic resilience in these model species as a function of environmental change and variation, thereby aiding freshwater monitoring and conservation. We urge further capacity building to support genetic applications to the conservation biology of other aquatic organisms.

Keywords: biodiversity, invertebrates, genetics, monitoring, river, stream

Introduction

Organisms in freshwater ecosystems are the most threatened worldwide, with biodiversity declining at a greater rate than in terrestrial or marine systems due to over-exploitation, pollution, flow modification, species invasion and habitat degradation (Postel and Richter 2003; Xenopoulos *et al.* 2005; Dudgeon *et al.* 2006). The effects of climate change will exacerbate these effects further (Heino *et al.* 2009; Dohet *et al.* 2015), creating major pressure on freshwaters and their conservation. While assessments of the resulting risks in freshwater ecosystems are increasingly widespread, overwhelmingly they are based on traditional abundance and species compositional data (e.g. Chessman 2015; Storey 2015) and largely ignore indicators of genetic biodiversity. Few risk assessments take advantage of molecular markers that can provide more cost effective and reliable indicators of demographic change than some traditional approaches (Schwartz *et al.* 2007). The application of genetic methods can aid in monitoring the presence of a species through DNA barcoding (Cardoni *et al.* 2015), and determine current population structure (e.g. connectivity or isolation) and longer-term changes in abundance and distribution (effective population size and genealogy) through the use of microsatellite loci (Theissinger *et al.* 2011). Without genetic techniques this would be labour intensive and may require sampling over long time periods which could be expensive and, for invertebrates, could be quite inaccurate. For example, traditional methods have inferred that mayfly species such as *Baetis rhodani* remain close to the stream channel from which they emerged throughout their lifespan (Petersen *et al.* 2004), however in contrast genetic studies reveal that this species has very high levels of dispersal and migration over large geographic distances (Alp *et al.* 2012).

Macroinvertebrates are among the most commonly used organisms for biomonitoring and conservation assessment in freshwater environments (Pfrender *et al.* 2010; Buss *et al.* 2015; Cardoni *et al.* 2015). They are recognised as such by the Water Framework Directive (2000/60/CE; European Commission 2000), reflecting their diversity, ease of sampling and functional importance. There is increasing recognition of the need to appraise genetic diversity as part of conservation strategies to compliment traditional studies (Taberlet *et al.* 2012; Crook *et al.* 2015). Such genetic assessments can be used for inferring demographic

resilience and flow within and among catchments, key features for inferring biotic response. Microsatellite markers (also known as simple sequence repeats / SSRs) are the most widely used molecular marker in population genetics due to their codominance and hypervariability among individuals (Liu *et al.* 2015). Microsatellites are frequently used successfully to aid in biodiversity conservation in the freshwater ecosystem, including wetland birds (Corrêa *et al.* 2015), fish (Raeymaekers *et al.* 2005; Abdul-Muneer 2014; Junge *et al.* 2014) and a range of invertebrates (Jones *et al.* 2015; Lopes-Lima *et al.* 2015; Pérez-Portela *et al.* 2015). However, the resources available for invertebrates are still limited when compared to vertebrates, due largely to the requirement to develop species-specific microsatellites. This constraint is now diminishing as the cost of Next Generation Sequencing (NGS) declines, while increased sequence output is accelerating microsatellite locus discovery (Yu *et al.* 2011; Fernandez-Silva *et al.* 2013).

Here, the specific aim was to produce sets of polymorphic microsatellite markers for three species of aquatic macroinvertebrate; *Isoperla grammatica* (Poda, 1761), *Amphinemura sulcicollis* (Stephens, 1836) and *Baetis rhodani* (Pictet, 1843), while addressing a broader goal of illustrating the potential for genetic approaches to freshwater organism conservation. These species were chosen due to patterns of co-occurrence yet are widespread enough to act as model organisms to study the genetic effects of a range of ecological pressures (Alp *et al.* 2012). Wide ranging species with large population sizes are also the most likely to reveal historical population processes (Whiteley *et al.* 2006). Their broad geographical range can be used to compare rivers and streams of contrasting types, while also meeting policy needs, which increasingly requires biomonitoring and conservation policies over large spatial scales (Statzner and Bêche 2010). All three species have aquatic larval and terrestrial adult stages, allowing a focus on dispersal and connectivity. Simultaneously, however, the species differ in size, functional group and trophic level: *A. sulcicollis* is predominantly a shredder; *B. rhodani* is a generalist detritivore or scraper/grazer whereas *I. grammatica* is a predator. Such functional contrasts potentially allow an examination of differing ecological processes through the lens of population genetic diversity and structure. Finally, the three species were chosen to fill a clear gap in available knowledge: *I. grammatica* and *A. sulcicollis* have no known nuclear genetic markers available. *B. rhodani* has seven microsatellite loci described (Williams *et al.* 2002),

but additional microsatellites are required as it is unknown how many unlinked loci are needed to distinguish between recently diverged populations and to increase statistical power (Haasl and Payseur 2011; Putman and Carbone 2014). Many studies choose a minimum of ten polymorphic loci (Cervini *et al.* 2006; Östergren *et al.* 2015).

Methods

Samples of each species were collected from sites in the Tywi, Wye, Glaslyn and Mawddach catchments in upland Wales, UK (Table E2), and stored in absolute ethanol. Genomic DNA was extracted from one individual per species using a High Pure PCR Template Preparation Kit for blood and tissue following the manufacturer's instructions (Roche Diagnostics GmbH Mannheim, Germany). All samples were treated with RNase after DNA extraction. *I. grammatica* and A. *sulcicollis* were sequenced on an Illumina HiSeq 2500 Rapid run at The Oxford Genomics Centre using the 300 paired-end (PE) mode. *B. rhodani* was sequenced later on an Illumina HiSeq 2000 PE91 at BGI (Shenzhen, China). The data obtained (36, 40 and 62 x 10^6 PE reads were produced for *I. grammatica*, *A. sulcicollis* and B. *rhodani*, respectively) were quality-controlled using the programs Trimmomatic (Bolger *et al.* 2014) and Musket (Liu *et al.* 2013), and assembled using SOAPdenovo2 (Luo *et al.* 2012). Microsatellite repeat sequences were located and primers designed using *PrimerPipeline* (Macdonald and Macdonald 2016) which incorporates the microsatellite identification tool, MISA (Thiel 2002) and Primer3 (Untergasser *et al.* 2012). Twenty-eight, 25 and 15 primer pairs (*I. grammatica*, *A. sulcicollis* and B. *rhodani*, respectively) were selected for screening using 36-40 samples from two sites in Wales (Table E1, Table E2; please see Supporting Information and Macdonald *et al* (2016) for full details on the pipeline; raw data is also available through the Short Read Archive (SRA STUDY: PRJNA315680 (SRP072016), and all assemblies have been deposited at DDBJ/ENA/GenBank (accessions LVVV00000000, LVVW00000000 and LVVX00000000).

All forward primers were tagged with an M13 tail (5'-AGGGTTTTCCCAGTCACGACGTT-3') at the 5' end. The use of the M13 tail method (Boutin-Ganache *et al.* 2001) negates the need for fluorescently labelled primers, significantly reducing the cost of genotyping (Schuelke

2000). The 5 µL PCR volume contained: 2.5 µl of QIAGEN Multiplex mix, 0.5 µl Q solution, 0.5 µM of forward and reverse primers, and fluorescent dye with complimentary M13 tail attached (TAM, HEX or FAM), and 1 µL DNA template (~50 ng). The following PCR conditions were used: initial denaturation at 94°C for 3 min, followed by 35 cycles of 94°C for 30 s, 60°C for 45 s and 72°C for 1 min, with a final extension of 72°C for 10 min. Samples were run on an ABI 3730 DNA Sequencer and analysed with GeneMarker 2.2.0 (SoftGenetics) using GS-500 (LIZ) as a size standard.

The number of alleles, heterozygosity and Hardy–Weinberg equilibrium (HWE) were estimated per locus for two populations of each species in GenAlEx (Peakall and Smouse 2012). Significances for multiple tests were adjusted using Bonferroni correction (Rice 1989). Polymorphism information content (PIC) was calculated per population using Cervus v.3.0.7 (Kalinowski *et al.* 2007). The presence of null alleles, allelic dropout and scoring errors were ascertained using Micro-Checker v2.2.3 (Van Oosterhout *et al.* 2004). Linkage disequilibrium (LD) was estimated using GENEPOP 4.2 (Raymond and Rousset 1995).

Results

Based on screening for successful PCR amplification, polymorphisms and accurate genotyping, a set of 18, 21 and 13 robust polymorphic microsatellites were identified for *I. grammatica*, *A. sulcicollis* and B. *rhodani* respectively. High levels of allelic diversity and PIC were observed for all three species (Na = 7-27, 3-16, 5-13, average 17, 9, 9; PIC= 0.67-0.95, 0.36-0.90, 0.42-0.88, average 0.88, 0.73, 0.72; Table E1). Some loci within each species showed significant departure from HWE for both populations (Iso_11-18, Amp_16-21 and B_11-13; Table E1) and evidence of null alleles (Iso_7-18, Amp_10-21 and B_9-13). There was no significant LD or scoring errors found for any of the loci.

Discussion

The highly polymorphic microsatellite loci characterised here are powerful tools with the potential to assess how environmental change and variation in upland rivers affects demographic structure in three indicator species. The development of these – or similar tools for other species, could aid freshwater conservation considerably by facilitating the appraisal of fine-scale demographic change within species, and by helping to identify important processes affecting whole communities.

Heterozygosity deficit and null alleles, as found in these species, have been commonly observed within insects (Chapuis and Estoup 2007; Brownlow *et al*, 2008), particularly within Baetidae (Alp *et al.* 2012). It has been largely explained by the patchy recruitment hypothesis, where individuals sampled are the offspring of a small number of breeding individuals (Alp *et al.* 2012), as opposed to a random sample of local genotypes. Heterozygosity deficit can also be caused by the Wahlund effect (inadvertently sampling distinct populations) or by statistical error from analysing small samples with highly polymorphic markers. No clear evidence in favour of any of the above explanations was found, and it may therefore be the case that some or all factors contributed.

Ultimately, all conservation efforts to protect and restore biodiversity depend on species and population monitoring to appraise distribution patterns and population size (Thomsen and Willerslev 2015). However, microsatellites such as those presented here can expand such monitoring to the genetic level, thereby revealing key attributes of genetic health, such as inbreeding and/or genetic drift. The effects of changing environmental conditions can be investigated by measuring the respective changes in population size and allele frequencies, particularly in species with short generation times (van Straalen and Timmermans 2002; Hoffmann and Willi 2008), potentially revealing important elements that might otherwise be overlooked during or following environmental fluctuations in river habitats.

More generally, invertebrates are often neglected in genetic approaches to biodiversity conservation because basic data (including taxonomy, species distributions and abundances, how these change temporarily and spatially, and other biodiversity informatics) remain scarce (Cardoso *et al.* 2011). These 51 new microsatellites for three species of freshwater invertebrate therefore represent a considerable resource for researchers and conservation

biologists working on freshwater ecosystems. To the best of our knowledge, within the large and diverse groups of Plecoptera and Ephemeroptera, there are only five species with between 3-13 microsatellites each (*Ameletus inopinatus* (Theissinger *et al.* 2008; Taubmann *et al.* 2011; Theissinger *et al.* 2011); *Siphlonisca aerodromia* (Gibbs *et al.* 1998), *Arcynopteryx compacta* (Theissinger *et al.* 2009; Theissinger *et al.* 2013), *Brachyptera braueri* (Geismar and Nowak 2013), and *Dinocras cephalotes* (Elbrecht *et al.* 2014)). Since *I. grammatica*, *A. sulcicollis* and *B. rhodani* have 18-21 microsatellites each (including primers by Williams *et al.* (2002)), they have the highest number of microsatellites available within their respective orders, which makes them suitable for a wide variety of analysis, including applications which benefit from a larger amount of data e.g. Approximate Bayesian Computation (ABC).

Recent studies exemplify some of the current limitations in genetic applications to freshwater invertebrate conservation. For example, Chessman (2015) investigated how the traits of different species affect their resistance and resilience to drought by studying the post-drought macroinvertebrate assemblages. However, no information was available to assess the genetic consequence for these assemblage changes, and Chessman (2015) himself calls for a revival of research on freshwater invertebrates to predict the effects of drought or other climate change effects. Similarly, the growth of trait-based approaches in freshwater conservation and assessment reveals how some organisms are systematically affected by global change, yet any parallel genetic consequences are unclear (Larsen and Ormerod 2010). Dudgeon *et al.* (2006) said that "*Developing effective conservation and management strategies for freshwater biodiversity requires documenting declines and extinctions and understanding the underlying causes*". This can be achieved by utilising a molecular approach; resistance, resilience and systematic biodiversity loss can be explored in more detail through investigating demographic history. This would shed light on invertebrate response to past disturbances (e.g. Estoup *et al.* 2004), as well as allowing inference on other species that share given localities. In contrast, conservation studies that include a genetic component are allowing improved understanding of key issues, revealing problems that could not be identified by ecological approaches alone (Schwartz *et al.* 2007).

Microsatellites have been utilised in a variety of ways to aid freshwater conservation in a wide range of taxa. For example, Ashikaga *et al.* (2015) recently used microsatellites to

define the population structure of the freshwater fish, *Brycon orbignyanus*: similar information for invertebrates could establish priority areas for conservation. Lopes-Lima *et al.* (2015) were able to differentiate among populations of the mussel, *Anodonta anatinato* allowing conservation strategies to benefit from genetic diversity and phylogeographic information. Genetic data from other aquatic invertebrates, such as the mayfly *Ameletus inopinatus*, have been used to develop climate change scenarios to aid future conservation efforts (Taubmann *et al.* 2011). Studies of common species, such as those presented here, might aid the conservation of a larger range of species through a comparative approach (Whiteley *et al.* 2006). Whiteley *et al.* (2006) studied the common species of mountain whitefish, *Prosopium williamsoni*, and found that patterns of genetic differentiation in this species were representative of other native species, providing a useful conservation unit guide that reflected the genetic differentiation of multiple species. At present, conservation efforts in freshwater ecosystems are focussed overwhelmingly on higher levels of organisation including species, populations, habitats and emergent properties such as ecosystem functions or services (Ormerod 2014). As pressures on freshwater environments intensify, we envisage the expansion in the applications of genetic data to aid in conservation assessment. We therefore urge further capacity building to support genetic applications to the conservation biology of macroinvertebrates and other aquatic organisms.

Acknowledgements

The research was funded by the NERC DURESS project (Diversity in Upland Rivers for Ecosystem Service Sustainability, NE/J01481/1) within the Biodiversity and Ecosystem Service Sustainability (BESS) Thematic Programme. HCM was funded by the Cardiff University President's Research Scholarships. Many thanks to Hefin Jones from Cardiff University for support and guidance, and to the two anonymous referees for their useful comments and feedback.

Supporting Information

Table E2. Details of pipeline leading to primer design and site information.

References

Abdul-Muneer, P. M. (2014). Application of microsatellite markers in conservation genetics and fisheries management: recent advances in population structure analysis and conservation strategies. *Genetics Research International* **2014:**11.

Alp, M., Keller, I., Westram, A. M. and Robinson, C. T. (2012). How river structure and biological traits influence gene flow: a population genetic study of two stream invertebrates with differing dispersal abilities. *Freshwater Biology* **57:**969-981.

Ashikaga, F. Y., Orsi, M. L., Oliveira, C., Senhorini, J. A. and Foresti, F. (2015). The endangered species *Brycon orbignyanus*: genetic analysis and definition of priority areas for conservation. *Environmental Biology of Fishes* **98:**1845-1855.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30:**2114-2120.

Boutin-Ganache, I., Raposo, M., Raymond, M. and Deschepper, C. F. (2001). M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. *Biotechniques* **31:**24-26, 28.

Brownlow, R. J., Dawson, D. A., Horsburgh, G. J., Bell, J. J. and Fish, J. D. (2008). A method for genotype validation and primer assessment in heterozygote-deficient species, as demonstrated in the prosobranch mollusc *Hydrobia ulvae*. *BMC Genetics* **9**:1-9.

Buss, D. F., Carlisle, D. M., Chon, T. S., Culp, J., Harding, J. S., Keizer-Vlek, H. E., Robinson, W. A., Strachan, S., Thirion, C. and Hughes, R. M. (2015). Stream biomonitoring using

macroinvertebrates around the globe: a comparison of large-scale programs. *Environmental monitoring and assessment* **187:**4132.

Cardoni, S., Tenchini, R., Ficulle, I., Piredda, R., Simeone, M. C. and Belfiore, C. (2015). DNA barcode assessment of Mediterranean mayflies (Ephemeroptera), benchmark data for a regional reference library for rapid biomonitoring of freshwaters. *Biochemical Systematics and Ecology* **62:**36-50.

Cardoso, P., Erwin, T. L., Borges, P. A. V. and New, T. R. (2011). The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation* **144:**2647-2655.

Cervini, M., Henrique-Silva, F., Mortari, N. and Matheucci Jr, E. (2006). Genetic variability of 10 microsatellite markers in the characterization of Brazilian Nellore cattle (*Bos indicus*). *Genetics and Molecular Biology* **29:**486-490.

Chapuis, M.-P. and Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution* **24:**621-631.

Chessman, B. C. (2015). Relationships between lotic macroinvertebrate traits and responses to extreme drought. *Freshwater Biology* **60:**50-63.

Corrêa, T. C., Del Lama, S. N., De Souza, J. R. and Miño, C. I. (2015). Genetic structuring among populations of the great egret, *Ardea alba egretta*, in major Brazilian wetlands. *Aquatic Conservation: Marine and Freshwater Ecosystems*. Doi: 10.1002/aqc.2588.

Crook, D. A., Lowe, W. H., Allendorf, F. W., Erős, T., Finn, D. S., Gillanders, B. M., Hadwen, W. L., Harrod, C., Hermoso, V., Jennings, S. *et al.* (2015). Human effects on ecological connectivity in aquatic ecosystems: integrating scientific approaches to support management and mitigation. *Science of The Total Environment* **534:**52-64.

Dohet, A., Hlúbiková, D., Wetzel, C. E., L'Hoste, L., Iffly, J. F., Hoffmann, L. and Ector, L. (2015). Influence of thermal regime and land use on benthic invertebrate communities inhabiting headwater streams exposed to contrasted shading. *Science of The Total Environment* **505:**1112-1126.

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z.-I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A.-H., Soto, D., Stiassny, M. L. J. and Sullivan, C.A. (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews* **81:**163-182.

Elbrecht, V., Feld, C. K., Gies, M., Hering, D., Sondermann, M., Tollrian, R. and Leese, F. (2014). Genetic diversity and dispersal potential of the stonefly *Dinocras cephalotes* in a central European low mountain range. *Freshwater Science* **33:**181-192.

Estoup, A., Beaumont, M., Sennedot, F., Moritz, C. and Cornuet, J.-M. (2004). Genetic analysis of complex demographic scenarios: spatially expanding populations of the Cane Toad, *Bufo marinus*. *Evolution* **58:**2021-2036.

European Commission. (2000). Directive 2000/60/EC of the European parliament and of the council of 23 october 2000 establishing a framework for community action in the field of water policy. *Official Journal of the European Union* **327:**1-72.

Fernandez-Silva, I., Whitney, J., Wainwright, B., Andrews, K. R., Ylitalo-Ward, H., Bowen, B. W., Toonen, R. J., Goetze, E. and Karl, S. A. (2013). Microsatellites for next-generation ecologists: a post-sequencing bioinformatics pipeline. *PLoS ONE* **8:**e55990.

Geismar, J. and Nowak, C. (2013). Isolation and characterisation of new microsatellite markers for the stonefly *Brachyptera braueri* comparing a traditional approach with high throughput 454 sequencing. *Conservation Genetics Resources* **5:**413-416.

Gibbs, H. L., Gibbs, K. E., Siebenmann, M. and and Collins, L. (1998). Genetic differentiation among populations of the rare mayfly *Siphlonisca aerodromia* Needham. *Journal of the North American Benthological Society* **17:**464-474.

Haasl, R. J. and Payseur, B. A. (2011). Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites. *Heredity (Edinb)* **106:**158-171.

Heino, J., Virkkala, R. and Toivonen, H. (2009). Climate change and freshwater biodiversity: detected patterns, future trends and adaptations in northern regions. *Biological reviews of the Cambridge Philosophical Society* **84:**39-54.

Hoffmann, A. A. and Willi, Y. (2008). Detecting genetic responses to environmental change. *Nature Reviews Genetics* **9:**421-432.

Jones, J. W., Neves, R. J. and Hallerman, E. M. (2015). Historical demography of freshwater mussels (Bivalvia: Unionidae): genetic evidence for population expansion and contraction during the late Pleistocene and Holocene. *Biological Journal of the Linnean Society* **114:**376-397.

Junge, C., Museth, J., Hindar, K., Kraabøl, M. and Vøllestad, L. A. (2014). Assessing the consequences of habitat fragmentation for two migratory salmonid fishes. *Aquatic Conservation: Marine and Freshwater Ecosystems* **24:**297-311.

Kalinowski, S. T., Taper, M. L. and Marshall, T. C. (2007). Revising how the computer program cervus accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16:**1099-1106.

Larsen, S. and Ormerod, S. J. (2010). Combined effects of habitat modification on trait composition and species nestedness in river invertebrates. *Biological Conservation* **143:**2638-2646.

Liu, C.-y., Yuan, C., Li, X.-t., Li, Z.-H., Zhao, P., Qian, Z.-Q., Li, J.-f. and Liu, Z.-L. (2015). Development of microsatellites from *Rehmannia glutinosa* transcriptome and evaluation of genetic diversity among *Rehmannia* species. *Biochemical Systematics and Ecology* **59:**177-182.

Liu, Y., Schroeder, J. and Schmidt, B. (2013). Musket: a multistage k-mer spectrum based error corrector for Illumina sequence data. *Bioinformatics* **29:**308-315.

Lopes-Lima, M., Sousa, R., Teixeira, A., Varandas, S., Riccardi, N., Aldridge, D. C. and Froufe, E. (2015). Newly developed microsatellite markers for the pan-European duck mussel, Anodonta anatina: revisiting the main mitochondrial lineages. *Aquatic Conservation: Marine and Freshwater Ecosystems*. Doi:10.1002/aqc.2575.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y. *et al.* (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* **1:**18.

Macdonald, H. C, Cunha, L and Bruford, M, W. (2016) Development of genomic resources for four potential environmental bioindicator species: *Isoperla grammatica*, *Amphinemura sulcicollis*, *Oniscus asellus* and *Baetis rhodani*. bioRxiv Doi: http://dx.doi.org/10.1101/046227

Macdonald, H. C and Macdonald, G. W. (2016) *Primerpipeline*. [ONLINE] available at: http://www.scrufster.com/primerpipeline/ [01th April 2016].

Ormerod, S. J. (2014). Rebalancing the philosophy of river conservation. *Aquatic Conservation: Marine and Freshwater Ecosystems* **24**:147–152.

Östergren, J., Nilsson, J., Lundqvist, H., Dannewitz, J. and Palm, S. (2015). Genetic baseline for conservation and management of sea trout in the northern Baltic Sea. *Conservation Genetics* **17**:177-191.

Peakall, R. and Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics* **28:**2537-2539.

Pérez-Portela, R., Noyer, C. and Becerro, M. A. (2015). Genetic structure and diversity of the endangered bath sponge Spongia lamella. *Aquatic Conservation: Marine and Freshwater Ecosystems* **25:**365-379.

Petersen, I., Masters, Z., Hildrew, A. G. and Ormerod, S. J. (2004). Dispersal of adult aquatic insects in catchments of differing land use. Journal of Applied Ecology **41**:934-950.

Pfrender, M. E., Hawkins, C. P., Bagley, M., Courtney, G. W., Creutzburg, B. R., Epler, J. H., Fend, S., Schindel, D., Ferrington, L. C., Jr., Hartzell, P. L. *et al.* (2010). Assessing macroinvertebrate biodiversity in freshwater ecosystems: advances and challenges in DNA-based approaches. *The Quarterly Review of Biology* **85:**319-340.

Postel, S. and Richter, B. (2003). Rivers for life: managing water for people and nature. Island Press, Washington.

Putman, A. I. and Carbone, I. (2014). Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and Evolution* **4:**4399-4428.

Raeymaekers, J. A. M., Maes, G. E., Audenaert, E. and Volckaert, F. A. M. (2005). Detecting Holocene divergence in the anadromous–freshwater three-spined stickleback (Gasterosteus aculeatus) system. *Molecular Ecology* **14:**1001-1014.

Raymond, M. and Rousset, F. (1995). GENEPOP (Version 1.2): Population genetics software for exact tests and ecumenicism. *Journal of Heredity* **86:**248-249.

Rice, W. R. (1989). Analyzing tabled of statistical tests. *Evolution* **43:**223-225.

Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology* **18:**233-234.

Schwartz, M. K., Luikart, G. and Waples, R. S. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in Ecology & Evolution* **22:**25-33.

Statzner, B. and Bêche, L. A. (2010). Can biological invertebrate traits resolve effects of multiple stressors on running water ecosystems? *Freshwater Biology* **55:**80-119.

Storey, R. (2015). Macroinvertebrate community responses to duration, intensity and timing of annual dry events in intermittent forested and pasture streams. *Aquatic Sciences* 1-20. Doi: 10.1007/s00027-015-0443-2.

Taberlet, P., Zimmermann, N. E., Englisch, T., Tribsch, A., Holderegger, R., Alvarez, N., Niklfeld, H., Coldea, G., Mirek, Z., Moilanen, A. *et al.* (2012). Genetic diversity in widespread species is not congruent with species richness in alpine plant communities. *Ecology Letters* **15:**1439-1448.

Taubmann, J., Theissinger, K., Feldheim, K. A., Laube, I., Graf, W., Haase, P., Johannesen, J. and Pauls, S. U. (2011). Modelling range shifts and assessing genetic diversity distribution of the montane aquatic mayfly *Ameletus inopinatus* in Europe under climate change scenarios. *Conservation Genetics* **12:**503-515.

Thiel, T. (2002) *MISA* - MIcroSAtellite identification tool [ONLINE] available at: http://pgrc.ipk-gatersleben.de/misa/misa.html [01th April 2016].

Theissinger, K., Bálint, M., Feldheim, K. A., Haase, P., Johannesen, J., Laube, I. and Pauls, S. U. (2013). Glacial survival and post-glacial recolonization of an arctic–alpine freshwater insect (*Arcynopteryx dichroa*, Plecoptera, Perlodidae) in Europe. *Journal of Biogeography* **40:**236-248.

Theissinger, K., Bálint, M., Haase, P., Johannesen, J. E. S., Laube, I. and Pauls, S. U. (2011). Molecular data and species distribution models reveal the Pleistocene history of the mayfly *Ameletus inopinatus* (Ephemeroptera: Siphlonuridae). Freshwater Biology **56**:2554-2566.

Theissinger, K., Feldheim, K. A., Seitz, A. and Pauls, S. U. (2009). Isolation and characterization of 11 polymorphic trinucleotide microsatellite markers in the stonefly *Arcynopteryx compacta* (Plecoptera: Perlodidae). *Molecular Ecology Resources* **9:**357-359.

Theissinger, K., Feldheim, K. A., Taubmann, J., Seitz, A. and Pauls, S. U. (2008). Isolation and characterization of 10 highly polymorphic di- and trinucleotide microsatellite markers in the mayfly *Ameletus inopinatus* (Ephemeroptera: Siphlonuridae). *Molecular Ecology Resources* **8:**1285-1287.

Thomsen, P. F. and Willerslev, E. (2015). Environmental DNA – an emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation* **183:**4-18.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B., Remm, M. and Rozen, S. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Research* **40:e115**.

Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. and Shipley, P. (2004). Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4:**535-538.

van Straalen, N. M. and Timmermans, M. J. T. N. (2002). Genetic variation in toxicant-stressed populations: an evaluation of the “genetic erosion” hypothesis. *Human and Ecological Risk Assessment: An International Journal* **8:**983-1002.

Whiteley, A. R., Spruell, P. and Allendorf, F. W. (2006). Can common species provide valuable information for conservation? *Molecular Ecology* **15:**2767-2786.

Williams, H. C., Wilcock, H. R. and Bruford, M. W. (2002). Microsatellite loci for the mayfly *Baetis rhodani* (Baetidae, Ephemeroptera). *Molecular Ecology Notes* **2:**411-412.

Xenopoulos, M. A., Lodge, D. M., Alcamo, J., Märker, M., Schulze, K. and Van Vuuren, D. P. (2005). Scenarios of freshwater fish extinctions from climate change and water withdrawal. *Global Change Biology* **11:**1557-1564.

Yu, J.-N., Won, C., Jun, J., Lim, Y. and Kwak, M. (2011). Fast and cost-effective mining of microsatellite markers using NGS technology: an example of a Korean water deer *Hydropotes inermis argyropus*. *PLoS ONE* **6:**e26933.

Table E1. Characteristics of polymorphic microsatellites in *Isoperla grammatica* (1-18), *Amphinemura sulcicollis* (19-39) and *Baetis rhodani* (40-52) including primer sequences (F = forward primer sequence, R = reverse primer sequence); Repeat (microsatellite repeat motif); N = number of individuals; N_A number of alleles per locus; H_O = observed heterozygosity; H_E = expected heterozygosity; HWE = p value of Hardy-Weinberg Equilibrium test (bold = significant disequilibrium after Bonferroni test (critical P value p ≤0.0028 for *Isoperla grammatica* and p ≤0.0025 for *Amphinemura sulcicollis* and *B. rhodani*)), and PIC = polymorphism information content. GenBank Accession numbers between KR068997-KR069048, respectively.

No.	Locus	Primer Sequence (5'-3')		Repeat	Site 1						Site 2					
					N	N_A	H_O	H_E	HWE	PIC	N	Na	H_O	H_E	HWE	PIC
1	Iso_1	F:	ACTTTCCATCCATGCGCTCA	(CT)22	20	17	1.000	0.942	0.974	0.91	20	13	0.650	0.758	0.368	0.72
		R:	GACTTTGGGTGACACCGTGA													
2	Iso_2	F:	TAGGCCTACTTACCGACCGT	(TC)26	20	25	0.900	0.959	0.162	0.93	20	21	0.850	0.953	0.016	0.92
		R:	ACCACTACCTCCTGTCTCCC													
3	Iso_3	F:	CAGCGACAATAGCCTCGGAT	(CT)24	19	14	0.895	0.925	0.603	0.89	20	15	0.800	0.874	0.551	0.84
		R:	CGAACAAGGCGGTGTGAATG													
4	Iso_4	F:	AACTGATCGCGAACTCCAAG	(TC)23	19	27	0.842	0.979	0.382	0.95	20	27	0.900	0.972	0.622	0.95
		R:	GTTCTTCACAGTCGGCTTCC													
5	Iso_5	F:	AGCTCACAGCCTTGTCAACA	(AG)21	17	21	0.412	0.971	**0.000**	0.94	20	20	0.800	0.908	0.075	0.88
		R:	GTACCCGTATAGGTCGCACG													
6	Iso_6	F:	TGTGAGAGCAACTGCACAGT	(CT)24	20	15	0.450	0.928	**0.000**	0.90	20	19	0.900	0.872	0.765	0.84
		R:	ACACCGCTTCGAACACAAGA													
7	Iso_7	F:	TGTGTTCCGGATGCAAGTAG	(CT)22	19	22	0.737	0.969	0.074	0.94	20	21	0.550	0.954	**0.001**	0.93
		R:	GATGCAAACGCTCGATAAAC													
8	Iso_8	F:	CGTGTTAAACGCGCTCGATA	(GA)23	20	22	0.700	0.965	0.038	0.94	20	14	0.650	0.897	**0.002**	0.87
		R:	CTGCCTGGTCACTGTGTTCA													
9	Iso_9	F:	TACTATCGTGGTTCGCGCTC	(TC)24	18	18	0.556	0.937	**0.000**	0.90	19	16	0.474	0.945	0.003	0.91

No.	Locus	Primer Sequence (5'-3')		Repeat	Site 1						Site 2					
					N	N_A	H_O	H_E	HWE	PIC	N	Na	H_O	H_E	HWE	PIC
		R:	GCAGGCCTCGAAACGAATTG													
10	Iso_10	F:	CGATCTGGGCAGGACAAGAG	(AG)26	20	10	0.500	0.855	**0.000**	0.81	20	11	0.500	0.709	0.005	0.67
		R:	CCAAGAACTTCCAGCCCCTT													
11	Iso_11	F:	CCACTTTGCCTTATGCGTCG	(TC)21	18	15	0.444	0.927	**0.000**	0.89	13	7	0.077	0.809	**0.000**	0.75
		R:	GAGAAAGGGAGAGAGCGCAG													
12	Iso_12	F:	TAACGTGCCATGACGTCACA	(AG)22	19	13	0.421	0.821	**0.000**	0.78	17	15	0.471	0.943	**0.001**	0.91
		R:	TAGGGTCCCTGAGGAGAAGC													
13	Iso_13	F:	AATACACGAGCTGGCCCAAA	(AG)23	20	20	0.400	0.963	**0.000**	0.94	19	23	0.632	0.967	**0.000**	0.94
		R:	AGGCACCCAATCAGTGATCG													
14	Iso_14	F:	GTCGCTTCTGTCGTCAGGAA	(AG)24	17	16	0.353	0.932	**0.000**	0.90	20	19	0.450	0.959	**0.000**	0.93
		R:	AACATAGGTGGCTTCGGCAA													
15	Iso_15	F:	AGACCAGCTCAATCGGGTTG	(GA)22	18	17	0.333	0.951	**0.000**	0.92	20	18	0.400	0.936	**0.000**	0.91
		R:	GTCTAGCTAGCGGGAGGGAA													
16	Iso_16	F:	TTCCCCTAGCCTCCGATCAA	(GA)21	16	15	0.313	0.938	**0.000**	0.90	15	11	0.200	0.892	**0.000**	0.85
		R:	TTTCGGGAAGGTTGTTGGGT													
17	Iso_17	F:	CGCCGGGGTCGTCATAATAA	(AG)25	17	11	0.294	0.854	**0.000**	0.81	18	11	0.389	0.829	**0.000**	0.79
		R:	GTGGCCCCGAATCATCTCTT													
18	Iso_18	F:	TGAATGATGGTGTTCGAAGTG	(GA)21	19	15	0.211	0.899	**0.000**	0.87	20	16	0.350	0.901	**0.000**	0.87
		R:	TATTGACAGCACCGACGTTG													
19	Amp_1	F:	TCAACCTATGGATCTGGAAACGA	(AG)23	15	8	0.867	0.743	0.982	0.69	14	7	0.786	0.720	0.824	0.66
		R:	TGCATTTCCTTTCTCTTTCCAGC													
20	Amp_2	F:	GTGTGAAGTGGAGGCTGTCA	(CA)24	18	14	0.833	0.924	0.898	0.89	13	12	0.769	0.905	0.087	0.86
		R:	ACAGCGCAAACTAAATCGCTC													
21	Amp_3	F:	GCCGTGCAGATCTCCATACT	(TG)23	20	11	0.750	0.690	0.868	0.66	16	6	0.500	0.558	0.929	0.52

No.	Locus	Primer Sequence (5’-3’)		Repeat	Site 1						Site 2					
					N	N_A	H_O	H_E	HWE	PIC	N	Na	H_O	H_E	HWE	PIC
		R:	ACTCACCTGCCTCGAACAAG													
22	Amp_4	F:	GGTCGACGCTGATAGTTTTCC	(TC)20	19	7	0.737	0.799	0.785	0.75	16	6	0.625	0.810	0.329	0.75
		R:	AGAAGGATTATGGGAGGAGGAGA													
23	Amp_5	F:	CCAGGGAGGTGATTGCAAGT	(AG)22	19	14	0.947	0.910	0.663	0.88	14	7	0.714	0.812	0.095	0.75
		R:	TCCGGTGGACTGCACAATTT													
24	Amp_6	F:	TCGGTTTCACGTGCAGTTAG	(AC)21	20	4	0.550	0.619	0.579	0.54	16	5	0.625	0.756	0.021	0.69
		R:	TGTCAATACGTTCCACTGCG													
25	Amp_7	F:	TAACGCACTGTGTACGCACT	(GT)22	20	3	0.400	0.528	0.177	0.42	16	6	0.563	0.758	0.172	0.69
		R:	CCCAAATTGTGATTGGCCGT													
26	Amp_8	F:	CGGATCGGATGCAGTGGTTA	(GA)21	19	10	0.842	0.832	0.102	0.79	16	8	0.625	0.839	0.137	0.79
		R:	GTCGGGCAAGTGACTAGAGG													
27	Amp_9	F:	TCAAGCGGTCTCCTTTGCTC	(GA)22	20	6	0.350	0.686	0.011	0.63	16	5	0.250	0.387	0.004	0.36
		R:	CCCTTTTCACGCATCGCAAT													
28	Amp_10	F:	AGAGATAGCGGCGTATCACG	(GA)21	17	15	0.706	0.911	0.026	0.87	16	10	0.500	0.837	0.004	0.80
		R:	TTTCCTGTCCCTCACACTCC													
29	Amp_11	F:	CATTTTGCTGCCCAGTTGCT	(GT)23	17	16	0.706	0.881	**0.001**	0.85	16	12	0.563	0.883	0.041	0.84
		R:	AAAAGCCGCGAACACACAAA													
30	Amp_12	F:	CAAGTTGGAAGCTCGGGTCT	(CGC)18	11	4	0.091	0.645	**0.001**	0.54	9	7	0.222	0.791	0.019	0.72
		R:	GTTTTTAGTGGGCCCATCGC													
31	Amp_13	F:	TCACACGCTCTATACACCGC	(CT)27	20	9	0.450	0.746	**0.000**	0.70	16	6	0.250	0.391	0.007	0.37
		R:	TTTCTACCCTCTGTGTCGCC													
32	Amp_14	F:	GCAACATTCAGGTCAACGCA	(CA)21	16	8	0.313	0.831	**0.000**	0.78	16	12	0.500	0.857	0.054	0.81
		R:	GGCAGGAGAACACAACACCT													
33	Amp_15	F:	GTGGGATAAGCGCCTAGAGTT	(TG)21	20	6	0.300	0.603	**0.000**	0.53	16	6	0.375	0.720	0.014	0.66

No.	Locus	Primer Sequence (5’-3’)		Repeat	Site 1						Site 2					
					N	N_A	H_O	H_E	HWE	PIC	N	Na	H_O	H_E	HWE	PIC
		R:	GTTAGGGTCGGCATGGAACT													
34	Amp_16	F:	TATCCAGCAGACATCGCCAG	(AC)20	20	14	0.400	0.915	**0.002**	0.88	15	9	0.333	0.871	**0.001**	0.82
		R:	GGGACCCTTGTGAGTCGGTA													
35	Amp_17	F:	TGCCCTTACTGTGTTTGTCG	(AC)25	17	12	0.412	0.845	**0.000**	0.80	13	11	0.385	0.895	**0.000**	0.85
		R:	AGTGCGAATATGGGCCTGTC													
36	Amp_18	F:	CATCCCCATACGGCGGTTAA	(AG)25	20	13	0.350	0.905	**0.000**	0.87	16	10	0.438	0.865	**0.002**	0.82
		R:	TTGCTCTCTCCCTTTCCCCT													
37	Amp_19	F:	ATCCGCTCAATCACACTGCA	(TG)24	20	13	0.350	0.918	**0.000**	0.89	16	11	0.438	0.867	**0.000**	0.82
		R:	CTGAAATTTCGCCAGGTGCC													
38	Amp_20	F:	TCCTTTCCACGTCTCGGTC	(TC)25	18	11	0.333	0.860	**0.000**	0.82	16	13	0.313	0.933	**0.000**	0.90
		R:	TGGCCTGCTCCTAGAAGTTC													
39	Amp_21	F:	AGACGCAGCGAATTTAGAGGT	(TG)24	20	9	0.200	0.768	**0.000**	0.72	16	4	0.313	0.663	**0.000**	0.58
		R:	TGCAAATGGTCACGTTACTGTT													
40	B_1	F:	GCGCGCTCGAATTTATAATT	(CT)26	18	6	0.778	0.695	0.802	0.63	19	8	0.579	0.740	0.347	0.68
		R:	TCGTACATAATAATAATCGGGCA													
41	B_2	F:	GCACACAGCATCAAGCAAAT	(AG)22	19	11	0.684	0.817	0.722	0.78	19	12	0.947	0.878	0.883	0.84
		R:	GCATATGCAGCCAGTCTTCA													
42	B_3	F:	GGCGCCGAATTATTATCATC	(GT)24	19	12	0.632	0.889	0.374	0.85	19	13	0.789	0.916	0.014	0.88
		R:	TTATAAACGGCGCCAAGATC													
43	B_4	F:	CCAATGCCAAATATGTTCCC	(TC)22	20	6	0.400	0.581	0.290	0.53	19	5	0.421	0.617	0.341	0.54
		R:	ACGACACGCACTCTTGGAA													
44	B_5	F:	GATTCGTGTTATGCATTGCG	(CA)26	19	12	0.684	0.842	0.048	0.80	18	12	0.556	0.910	**0.000**	0.87
		R:	ATAGCCGACGAGCATCAAAC													
45	B_6	F:	ACTTGTATTCCACGCTGCG	(CT)20	19	5	0.263	0.450	**0.002**	0.42	18	9	0.722	0.800	0.530	0.75

No.	Locus	Primer Sequence (5’-3’)		Repeat	Site 1						Site 2					
					N	N_A	H_O	H_E	HWE	PIC	N	Na	H_O	H_E	HWE	PIC
		R:	TCGGGTTGCAAGATACACAG													
46	B_7	F:	GCAGAAAGAGCAGCGCTTTA	(GA)23	19	6	0.211	0.724	**0.000**	0.67	19	7	0.632	0.639	0.202	0.59
		R:	GGCACGGTGTTTTAATTGCT													
47	B_8	F:	CGGTGCTCCCTCTTACTCTC	(GA)23	20	8	0.400	0.791	**0.000**	0.75	19	7	0.579	0.781	0.160	0.72
		R:	TTTAATTAACGGACGCCCTC													
48	B_9	F:	TCGCACCAGAACAAGACATC	(GA)24	20	13	0.600	0.818	0.177	0.78	19	6	0.421	0.733	0.328	0.66
		R::	CGCCTCACACTTTGAGTGAA													
49	B_10	F:	GCTTGAGAATTGCATTCACG	(GA)20	20	9	0.300	0.832	**0.000**	0.79	15	13	0.533	0.915	0.015	0.87
		R:	CGGTGAGCAAATGGTGTATG													
50	B_11	F:	GCGTTTTCCTCCATCCCT	(GA)22	20	11	0.350	0.837	**0.001**	0.80	19	7	0.263	0.848	**0.000**	0.80
		R:	TATTATGTCGCCGGCAAAAT													
51	B_12	F:	AATTCACATTTGCGTGATGG	(GA)28	12	6	0.083	0.786	**0.000**	0.72	14	6	0.143	0.741	**0.000**	0.67
		R:	TCATCATTCTTCTTGCTGTCAAA													
52	B_13	F:	CTTGACTTATTGCACGAAATCC	(GA)20	16	5	0.063	0.708	**0.000**	0.64	17	11	0.353	0.806	**0.000**	0.76
		R:	CCACGCCTATATGATCTCCC													

Supporting Information

Among the three species tested, *Isoperla grammatica*, *Amphinemura sulcicollis* and *Baetis rhodani*, there were differences in the number of contigs and primers designed (Table.S1). These differences are in part due to inevitable small differences in NGS but also the relative success of each *de novo* assembly; *A. sulcicollis* had the most successful assembly, represented by the highest N50 value, so the data could be assembled into fewer, larger contigs, arguably better for microsatellite mining. In contrast, *I. grammatica* had the lowest N50 and the least successful de novo assembly; this could be due to the fact that *I. grammatica* also has the largest estimated genome size (1184Mbp compared to 237.6Mbp for *A. sulcicollis*). Multiple assemblies at different kmer sizes were attempted for each species (please see bioRix manuscript Macdonald *et al.* (2016; doi: http://dx.doi.org/10.1101/046227 [1st April 2016]) for full details on the pipeline; raw data are available through SRA, STUDY: PRJNA315680 (SRP072016) and all assemblies have been deposited at DDBJ/ENA/GenBank, accessions LVVV00000000, LVVW00000000 and LVVX00000000).

The criteria for SSR mining and primer design using *PrimerPipeline* (Macdonald and Macdonald 2016; http://www.scrufster.com/primerpipeline/ [1st April 2016]) includes an optimum primer size of 20 bp (range 18-27bp), a primer product size of between 100-300, an optimum GC content of 50% (range 20%-80%), and an optimum annealing temperature of 60°C (range 57-63°C), so that different loci could be multiplexed. The same criteria for primer design was used for all three species. Though many SSRs of different repeat units were designed, di-nucleotide primers were favoured because of their tendency for higher polymorphism rates and ease of scoring. The primers that were tested were chosen based on their robustness to amplification artifacts, null alleles, and overall amplification success rate. Tested primers that are not described here were usually discarded due to poor amplification success, rather than issues with polymorphism. *Baetis rhodani* has the least amount of primers tested and described, however when the seven microsatellite loci that have previously been described are added, it results in 18, 21 and 20 loci which are now available for *Isoperla grammatica*, *Amphinemura sulcicollis* and *Baetis rhodani* respectively.

I. grammatica and *A. sulcicollis* primers were tested from samples at the same locations, however *B. rhodani* was tested using samples from different locations due to cryptic species being present at some sites and because *B. rhodani* is acid sensitive and not always present at the same sites as *I. grammatica* and *A. sulcicollis.*

Table E2. Details of pipeline leading to primer design, including the number of contigs (sequences) the NGS data was assembled into, average size of contig (N50), number of SSRs (microsatellites) and primers found and tested, the location of sites and number of samples tested per site. Coordinates are in EPSG:27700 - OSGB 1936 / British National Grid.

Species	No. contigs assembled into	N50 of assembly	No. SSRs found	No. of primers designed	No. Tested	No. Described	Location of Site 1 (Catchment)	No. samples tested at site 1	Location of Site 2 (Catchment)	No. samples tested at site 2
Isoperla grammatica	6,870,359	568	26,063	1,505	28	18	282215E 249639N (Tywi)	20	257600E 350900N (Glaslyn)	20
Amphinemura sulcicollis	1,632,947	1,543	6,040	512	25	21	282215E 249639N (Tywi)	20	257600E 350900N (Glaslyn)	16
Baetis rhodani	1,418,939	850	717	50	15	13	310400E 268300N (Wye)	20	279110E 319620N (Mawddach)	19

Appendix F - Final Data Set

Complete genotyping data for all three species including: details of which extraction method were used for each sample, how many times each sample was repeated, and tables giving instructions how each allele was binned for each loci.

Table of tables

Table F1....................378
Table F2....................388
Table F3*i-x*....................398
Table F4....................404
Table F5....................412
Table F6*i-x*....................421
Table F7....................435
Table F8....................441
Table F9i-xiii....................448

Appendix F - Final Data Set

Amphinemura sulcicollis – Genotyping and binning

Table F1. Showing the final genotyping data for *A. sulcicollis* at 5 locis (Amp_2- Amp_6). *Site* identifies the 17 different sites in this data set; *N* = each individual's position in the dataset (same order used in all analyses); *Name* is the name given to each sample and *E* represents which extraction method was used for each sample (see note at end of table). For each loci *R* = number of times each sample was repeated successfully, *F* = number of failed samples, and *Allele* are the two alleles found in this sample after binning.

Site	N	Name	E	Amp_2				Amp_3				Amp_4				Amp_5				Amp_6			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
104	1	104A1	1	1		258	262	1		210	210	1		170	176	1		204	222	1		248	248
	2	104A10	2	1		272	280	2		208	220	1		170	174	2		206	208	1		248	254
	3	104A11	2	1		280	284	1		192	192	1		160	170	1		220	234	1		250	254
	4	104A12	2	1		240	258	1		206	210	1		170	174	1		212	228	1		248	254
	5	104A13	2	1		250	258	1		196	210	1		172	174	2		226	228	1		248	254
	6	104A14	2	1		266	268	1		210	212	1		170	170	1		220	228	1		248	254
	7	104A15	2	1		276	292	1		192	212	1		168	172	1		220	228	1		248	254
	8	104A16	2	1		262	276	1		192	192	1		160	172	1		208	228	1		248	254
	9	104A17	2	1		262	282	1		192	210	1		160	170	2		228	230	1		248	254
	10	104A18	2	1		260	280	1		192	210	1		170	172	1		208	212	1		248	252
	11	104A19	2	1		240	256	2		192	206	2		160	160	1		204	212	2		250	254
	12	104A2	1	1		256	264	1		192	204	1		162	172	1		204	212	2		248	254
	13	104A20	2	1		256	276	2		192	218	2		172	174	1		212	216	2		248	254
	14	104A3	2	1		278	284	1		192	212	1		162	170	1		216	228	1		248	254
	15	104A4	2	1		278	280	1		192	192	1		162	170	1		204	210	2		248	252
	16	104A5	2	1		272	282	1		192	210	1		160	176	1		212	234	2		248	254
	17	104A6	2	1		256	258	1		192	198	1		162	174	1		208	230	2		254	256
	18	104A7	2	1		240	260	1		192	212	1		170	172	1		220	228	1		248	250
	19	104A8	2	1		258	278	1		192	192	1		174	174	1		228	232	1		250	250
	20	104A9	2	1		260	264	1		192	208	2		160	176	1		204	220	1		248	250
109	21	109A1	1	1		258	278	1		192	192	1		168	176	1		208	212	1		248	248

Site	N	Name	E	Amp_2				Amp_3				Amp_4				Amp_5				Amp_6			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	22	109A10	2	1		256	258	1		192	192	1		170	174	1		210	228	1		248	250
	23	109A11	2	1		282	320	1		192	210	1		170	174	1		204	212	2		248	254
	24	109A12	2	3		256	260	1		192	192	1		170	172	1		222	224	1		248	256
	25	109A13	2	1		244	282	1		192	192	1		172	176	1		216	228	1		254	256
	26	109A14	2	1		256	278	2		192	210	2		170	176	1		208	214	2		248	254
	27	109A15	2	1		272	296	2		192	196	2		164	170	1		220	228	2		248	250
	28	109A16	2	1		256	280	2		192	220	2		170	174	1		208	212	2		250	254
	29	109A17	2	1		256	280	2		196	212	2		170	174	1		210	210	2		254	258
	30	109A18	2	1		274	278	2		192	192	2		160	172	1		210	220	2		248	250
	31	109A2	1	1		256	280	1		192	196	1		170	172	1		202	228	1		254	260
	32	109A3	1	1		262	278	1		196	216	3		160	162	2	1	206	208	1		248	250
	33	109A4	1	1		252	256	1		192	218	1		160	174	1		212	228	1		248	248
	34	109A5	1	1		272	278	1		210	210	1		162	172	1		208	224	1		248	248
	35	109A6	2	1		260	264	1		192	198	1		172	174	1		204	228	1		252	254
	36	109A7	2	1		256	260	1		192	212	1		174	176	1		212	220	1		252	254
	37	109A8	2	1		256	270	1		210	212	1		160	172	1		206	220	1		250	250
	38	109A9	2	1		256	274	1		192	210	1		170	172	1		210	228	1		256	256
108	39	108A1	1	2		276	296	2		198	210	2		172	172	2		208	234	2		250	254
	40	108A10	2	1		262	272	1		192	196	1		168	174	2		210	220	1		248	252
	41	108A11	2	1		240	258	1		192	196	1		174	176	3		210	230	1		254	254
	42	108A12	2	1		254	260	1		192	192	1		170	170	2		214	226	1		248	250
	43	108A13	2	2		258	286	1		192	192	1		160	170	1		214	236	1		248	254
	44	108A14	2	1		262	276	1		192	228	1		170	172	1		208	222	1		248	248
	45	108A15	2	1		278	284	1		192	210	1		160	174	1		204	210	1		248	248
	46	108A16	2	2		280	298	1		192	216	1	1	160	170	1		210	220	1		248	250
	47	108A17	2	1		278	280	1		192	196	1		174	176	1		210	228	1		248	250
	48	108A18	2	1		256	280	1		190	192	1		172	176	1		204	220	1		254	254
	49	108A19	2	1	1	260	278	1		192	216	1		174	174	2		222	222	1		248	250
	50	108A2	1	2		278	280	2		192	212	2		172	174	2		222	228	2		248	254
	51	108A20	2	1	1	258	280	1		190	192	1		170	174	1		204	210	1		248	254
	52	108A3	2	2		256	258	2		192	192	2		164	170	2		206	222	2		248	252

Site	N	Name	E	Amp_2				Amp_3				Amp_4				Amp_5				Amp_6			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	53	108A4	2	2		260	276	2		192	192	2		160	170	2		206	220	2		248	254
	54	108A5	2	1	1	278	278	1		204	212	1		168	174	2		224	232	1		254	254
	55	108A6	2	1		240	256	1		192	192	1		170	170	1		212	226	1		248	256
	56	108A7	2	2	1	256	290	1		192	210	1		170	170	1		214	222	1		248	248
	57	108A8	2	2		278	296	1		194	212	1		170	174	1		210	232	1		248	248
	58	108A9	2	1		240	262	3		192	200	1		160	174	1		210	220	1		248	248
102	59	102A1	1	1		260	262	1		192	216	1		160	170	1		228	230	1		248	254
	60	102A10	2	1		268	272	1		192	210	1		160	172	2		208	210	1		254	256
	61	102A11	2	1		274	282	1		192	192	1		160	170	1		230	230	1		248	254
	62	102A12	2	1		260	278	1		192	192	1		162	162	1		204	210	1		252	256
	63	102A13	2	1		260	260	1		192	196	1		174	174	1		210	226	1		248	254
	64	102A14	2	1		262	272	1		192	198	1		174	176	1		210	210	1		248	254
	65	102A15	2	1		256	258	1		210	210	1		162	170	1		220	222	1		248	254
	66	102A16	2	1		258	258	1		192	212	1		160	172	1		220	220	1		254	254
	67	102A17	2	1		262	264	1		192	212	1		170	172	1		212	224	1		248	254
	68	102A18	2	1		274	280	1		192	212	1		172	176	1		204	220	1		248	250
	69	102A19	2	1		240	258	1		192	192	1		170	170	1		210	214	1		254	256
	70	102A2	1	1		280	280	1		192	192	1		170	172	1		222	228	1		248	250
	71	102A20	2	1		240	280	1		192	196	1		162	172	1		210	226	1		248	254
	72	102A3	2	1		256	276	1		192	210	1		172	174	1		204	212	1		248	248
	73	102A4	2	1		260	262	1		192	192	1		162	176	1		220	224	1		248	256
	74	102A5	2	1		270	280	1		192	192	1		170	170	1		220	226	1		248	248
	75	102A6	2	1		256	298	1		210	212	1		162	162	1		220	230	1		254	254
	76	102A7	2	1		260	262	1		196	212	1		160	172	1		212	228	1		254	256
	77	102A8	2	1		262	276	1		192	196	1		172	172	1		206	210	1		248	254
	78	102A9	2	1		266	272	1		192	192	1		170	172	1		204	212	1		248	254
112	79	112A1	1	1		268	280	1		192	226	1		174	176	1		204	228	1		248	248
	80	112A10	2	1		256	262	1		192	192	1		160	176	2		208	210	1		248	248
	81	112A11	2	1		236	242	1		192	210	1		160	172	1		208	224	1		248	254
	82	112A12	2	1		256	262	1		192	194	1		160	174	1		212	222	1		248	254
	83	112A13	2	1		256	256	1		192	192	1		162	170	1		204	220	1		248	248

Site	N	Name	E	Amp_2				Amp_3				Amp_4				Amp_5				Amp_6			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	84	112A14	2	1		256	272	1		192	210	1		160	170	1		220	220	1		248	254
	85	112A15	2	1		256	296	1		192	212	1		170	170	1		212	220	1		248	248
	86	112A16	2	1		274	282	1		192	218	1		170	172	1		212	212	1		248	254
	87	112A17	2	1		256	260	1		192	196	1		160	174	1		206	214	1		248	248
	88	112A18	2	1		256	258	1		210	210	1		160	170	1		228	228	1		248	254
	89	112A19	2	1		258	282	1		192	212	1		174	174	1		224	228	1		248	250
	90	112A2	1	3		248	276	1		192	192	1		170	170	1		204	204	2		248	254
	91	112A20	2	2		258	278	1		192	192	2		164	172	4		212	214	2		248	252
	92	112A3	2	2		244	278	2		192	192	2		172	174	2	1	208	220	2	1	248	254
	93	112A4	2	2		240	282	2		210	212	2		160	174	2	1	220	226	2	1	248	254
	94	112A5	2	2		258	294	2		192	206	2		170	178	3		208	234	3		254	256
	95	112A6	2	2		258	264	2		192	196	2		160	176	3		210	212	2	1	248	254
	96	112A7	2	1		260	260	1		192	212	1		162	170	1		220	228	1		248	254
	97	112A8	2	1		260	284	1		192	210	1		160	172	1		206	208	1		248	254
	98	112A9	2	1		256	266	1		192	192	1		170	178	1		210	210	1		248	250
46	99	46A1	3	1		250	274	1		192	210	1		160	170	1		204	204	1		254	254
	100	46A2	3	1		262	276	2	1	192	224	1		170	170	1		204	220	1		248	256
	101	46A3	3	1		260	262	1		192	196	1		160	174	1		208	210	1		250	254
	102	46A4	3	1		260	264	1		192	212	1		172	172	1		210	210	1		254	254
	103	46A5	3	1		264	264	1		192	216	1		162	172	1		204	220	1		248	254
96	104	96A1	1	1		256	286	1		192	220	1		162	176	1		208	228	1		248	250
	105	96A10	2	1		264	278	1		192	210	1		160	170	1		220	220	1		248	254
	106	96A11	2	1		262	278	1		192	192	1		162	172	1		212	212	1		248	248
	107	96A12	2	1		256	276	2		192	192	2		170	170	2		208	220	1		248	254
	108	96A13	2	1		260	302	1		210	212	1		170	174	1		208	220	1		248	250
	109	96A14	2	1		256	262	1		192	210	1		162	170	1		214	220	1		248	254
	110	96A15	2	2		260	272	1	1	212	216	1	1	160	160	2		220	220	1		248	256
	111	96A16	2	1		282	294	1		192	206	1		170	172	1		208	228	1		254	254
	112	96A17	2	1		258	260	1		190	216	1		160	162	1		204	220	1		248	254
	113	96A18	2	1		260	276	1		192	210	1		160	170	1		204	228	1		254	254
	114	96A19	2	1		262	280	1		192	198	2	1	170	174	1		208	228	1		252	254

Site	N	Name	E	Amp_2				Amp_3				Amp_4				Amp_5				Amp_6			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	115	96A2	1	1		260	320	1		192	192	1		160	162	1		202	224	1		248	250
	116	96A3	2	1		258	274	1		192	194	1		170	172	1		210	220	1		248	254
	117	96A4	2	1		274	276	1		192	210	1		160	172	1		208	220	1		248	254
	118	96A5	2	1		258	286	1		192	220	1		160	170	1		220	228	1		254	254
	119	96A6	2	1		256	262	1		194	196	1		172	174	1		208	210	1		248	248
	120	96A7	2	1		256	300	2		192	192	1		172	172	1		220	222	1		248	250
	121	96A8	2	1		256	256	1		192	210	1		172	172	1		210	210	1		248	250
	122	96A9	2	1		258	258	1		190	208	1		170	174	1		208	220	1		250	250
9	123	09A1	4	1		278	278	1		192	212	1		170	172	1	1	206	230	2		248	248
	124	09A10	4	1		256	258	1		192	192	1		170	174	2		204	220	2		248	250
	125	09A11	4	1		262	274	1		192	192	1		160	162	3		220	232	2		248	248
	126	09A12	4	1		262	282	1		192	192	1		160	162	2		204	214	2		248	254
	127	09A13	4	1		256	260	1		192	192	1		162	170	2		228	232	2		250	254
	128	09A14	4	1		262	282	1		192	210	1		162	176	2		220	222	2		256	256
	129	09A15	4	1		258	262	1		192	212	1		170	178	1	1	220	222	2		248	250
	130	09A16	4	2		256	300	1		192	192	1		172	172	1	1	208	210	2		254	254
	131	09A17	4	2		262	270	1		192	192	1		160	170	1	1	220	222	2		254	254
	132	09A18	4	1		256	260	1		192	192	1		170	176	2		204	204	2		248	252
	133	09A19	4	1		272	288	1		192	192	2		162	172	1	1	204	214	2		254	256
	134	09A2	4	2		242	270	2		204	216	1		162	174	2		228	228	2		248	256
	135	09A20	4	1		274	280	1		196	212	1		160	160	2		204	220	2		248	250
	136	09A3	4	1		256	268	1		210	212	1		160	162	1	1	210	226	2		248	248
	137	09A4	4	1		256	258	1		192	196	1		178	178	1	1	220	222	2		254	254
	138	09A5	4	1	1	264	292	1		192	192	1		166	174	1	1	204	230	2		248	248
	139	09A6	4	1		284	288	1		192	192	1		162	174	2		220	228	2		250	250
	140	09A7	4	1		256	264	1		192	212	1		172	174	1	1	204	210	2		248	254
	141	09A8	4	1		256	262	1		192	212	1		160	178	1	1	204	232	2		250	250
	142	09A9	4	2		238	256	1		190	192	1		162	174	1	1	208	214	2		248	248
116	143	116A1	1	1		260	280	1		192	212	2		174	180	1		214	220	1		248	254
	144	116A10	2	1		256	256	1		192	212	1		170	174	2		208	210	1		254	254
	145	116A11	2	2		256	278	2		192	192	2		170	172	2		204	204	1		254	256

Site	N	Name	E	Amp_2				Amp_3				Amp_4				Amp_5				Amp_6			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	146	116A12	2	2		256	272	2		192	196	2		160	172	2		204	212	1		248	254
	147	116A13	2	2		272	274	2		192	210	2		162	174	2		220	232	1		248	254
	148	116A14	2	2		240	262	2		192	192	2		160	170	2		204	220	1		248	250
	149	116A15	2	2		260	266	2		192	192	2		162	174	2		204	208	1		248	254
	150	116A16	2	2		248	260	2		192	192	2		156	174	2		208	212	1		248	252
	151	116A17	2	1		262	262	1		192	192	1		162	170	1		204	208	1		250	250
	152	116A18	2	1		236	260	1		192	192	1		164	176	1		224	232	1		250	256
	153	116A19	2	1		242	282	1		192	192	1		160	170	1		204	204	1		248	254
	154	116A2	1	1		278	280	1		192	192	1		170	174	1		204	204	1		248	250
	155	116A20	2	1		258	282	1		192	192	1		162	170	1		204	220	1		248	250
	156	116A3	2	1		258	260	1		192	212	2		160	176	1		208	230	1		248	254
	157	116A4	2	2		298	308	1		192	192	1		170	174	1		220	224	1		254	254
	158	116A5	2	1		262	282	1		192	192	1		160	162	1		212	226	1		252	254
	159	116A6	2	1		256	280	1		212	212	1		170	174	2		210	220	1		252	254
	160	116A7	2	2		254	282	1		192	194	1		172	176	1		204	222	1		248	254
	161	116A8	2	1		258	262	1		192	192	1		172	174	1		204	204	1		248	254
	162	116A9	2	1		272	286	2		192	192	1		160	162	1		204	204	1		248	252
113	163	113A1	1	1		256	262	1		192	216	1		162	172	1		212	220	1		248	250
	164	113A10	2	2		258	282	2		192	210	2		160	162	2		230	230	1		248	252
	165	113A11	2	2		256	256	2		192	196	2		160	174	2		222	238	1		248	254
	166	113A12	2	2		266	284	2		192	210	2		162	174	2		210	228	1		254	254
	167	113A13	2	2		260	262	2		192	204	2		170	174	2		204	220	1		248	254
	168	113A14	2	2		260	274	2		192	196	2		170	174	2		220	236	1		254	254
	169	113A15	2	1		278	278	1		192	212	1		168	170	1		212	220	1		256	256
	170	113A16	2	1		276	288	1		192	212	1		156	174	1		212	224	1		248	248
	171	113A17	2	2		256	270	2		208	208	1	1	172	176	2	1	204	210	2	1	250	254
	172	113A18	2	2		248	256	2		192	192	2		174	174	3		224	228	3		248	248
	173	113A19	2	2		258	264	3		192	192	2		162	170	2	1	204	210	2	1	248	254
	174	113A2	1	1		264	278	1		192	212	1		172	174	1		212	226	1		248	254
	175	113A20	2	2		256	256	2		210	212	2		160	162	2	1	210	222	2	1	254	258
	176	113A3	2	1		270	286	1		210	210	1		170	172	1		214	230	1		248	250

Site	N	Name	E	Amp_2				Amp_3				Amp_4				Amp_5				Amp_6			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	177	113A4	2	1		258	268	1		192	212	1		160	172	1		220	222	1		248	254
	178	113A5	2	1		256	258	1		192	210	1		160	174	1		206	220	1		248	256
	179	113A6	2	1		260	280	1		192	210	1		170	172	1		220	226	1		248	254
	180	113A7	2	1		256	270	1		192	198	1		170	176	1		212	218	1		248	254
	181	113A8	2	1		240	260	1		192	192	1		170	170	1		212	228	2		246	254
	182	113A9	2	2		260	292	2		192	198	2		170	174	2		220	224	1		254	254
95	183	95A1	1	1		256	260	1		210	212	1		174	176	1		210	220	1		248	270
	184	95A10	2	1		256	274	1		192	210	1		162	176	1		220	220	1		248	252
	185	95A11	2	1		260	260	1		192	216	1		160	170	1		210	220	1		252	256
	186	95A12	2	1		280	282	1		192	192	1		170	174	1		210	220	1		254	270
	187	95A13	2	1		260	272	1		192	226	1		162	170	1		220	220	1		248	270
	188	95A14	2	1		260	278	1		192	192	1		172	178	1		220	220	1		252	270
	189	95A15	2	1		240	240	1		210	226	1		170	172	1		220	220	1		248	270
	190	95A16	2	1		280	284	1		192	212	1		162	170	1		212	230	1		248	254
	191	95A17	2	1		260	266	1		192	212	1		162	174	1		220	220	1		248	252
	192	95A18	2	1		260	280	1		196	210	1		160	172	1		220	220	1		254	254
	193	95A19	2	1		260	274	1		192	210	1		162	172	1		210	220	1		254	270
	194	95A2	1	2		260	278	2		192	216	1		162	176	1		220	224	1		248	254
	195	95A20	2	1		264	264	1		192	216	1		162	172	1		206	220	1		252	270
	196	95A21	2	1		266	280	1		192	192	1		162	172	1		206	220	1		254	270
	197	95A22	2	1		260	260	1		192	210	1		172	174	1		208	220	1		248	248
	198	95A3	1	1		256	286	1		210	212	1		162	170	1		212	220	1		250	252
	199	95A4	1	1		284	288	1		192	194	2		160	174	1		220	226	1		248	270
	200	95A5	1	1		262	278	1		212	226	1		162	170	1		220	226	1		248	270
	201	95A6	2	1		260	266	1		212	212	1		162	174	1		220	220	1		248	250
	202	95A7	2	1		260	260	1		196	212	1		156	172	1		220	220	1		248	254
	203	95A8	2	1		256	260	1		226	226	1		160	170	1		204	220	1		252	270
	204	95A9	2	1		260	278	1		212	212	1		162	162	1		220	228	1		254	254
6	205	06A1	4	2		256	278	2		192	210	3		162	174	2	2	220	220	2	2	248	270
	206	06A10	4	1		274	278	1		210	216	2		162	172	1	1	216	220	2		248	248
	207	06A11	4	2		260	270	2		210	212	2		162	170	1	1	220	220	2		248	248

Site	N	Name	E	Amp_2				Amp_3				Amp_4				Amp_5				Amp_6			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	208	06A12	4	2		260	278	1		212	226	1		156	160	1	1	220	220	2		248	270
	209	06A13	4	2		260	278	1		192	218	1		162	174	1	1	220	220	2		248	270
	210	06A14	4	1		260	260	1		192	192	1		162	172	1	1	220	220	2		248	252
	211	06A15	4	1	1	256	256	2		212	212	2		170	172	1	1	220	222	2	1	248	248
	212	06A16	4	1		260	260	1		192	212	1		162	170	1	1	220	220	2		248	270
	213	06A17	4	1	1	260	280	2		192	192	2		172	174	1	1	220	220	1	1	248	248
	214	06A18	4	1		256	260	1		192	210	2		162	174	2		204	220	2		248	248
	215	06A19	4	1	1	278	278	1		192	192	2		170	172	1	1	220	220	2	1	270	270
	216	06A2	4	1		260	266	2		210	228	3		160	172	1	1	210	216	1	1	248	252
	217	06A20	4	1		260	280	1		192	210	1		162	172	1	1	220	228	2		248	252
	218	06A21	4	2		260	264	1		192	192	1		172	174	1	1	204	220	2		248	248
	219	06A3	4	1		240	262	1		192	192	2		162	174	1	1	220	220	2		248	252
	220	06A4	4	2		258	260	2	1	192	216	2	1	170	172	1	3	220	224	1	3	270	270
	221	06A5	4	1		272	278	1		192	210	2		160	160	1	1	222	222	1	1	248	252
	222	06A6	4	1		260	268	2		192	226	2		170	170	1	1	220	230	2		248	270
	223	06A7	4	1		260	280	1		192	192	2		162	172	1	1	204	220	1	1	256	270
	224	06A8	4	1		260	260	1		192	192	1	1	166	170	1	1	220	224	2		248	270
	225	06A9	4	1		272	274	1		192	192	2		168	176	1	1	220	220	1	1	248	270
94	226	94A2	1	1		276	284	1		192	210	1		168	170	1		220	222	1		254	254
	227	94A21	4	1		278	280	1		192	216	1		170	174	1		212	222	1		248	270
	228	94A22	4	1		256	278	1		204	216	1		170	174	1		220	220	1		256	270
	229	94A23	4	1		260	260	1		192	192	1		170	172	1		208	220	1		248	270
	230	94A3	2	2		228	270	1		192	192	2		170	170	2	1	204	212	1		236	236
4	231	04A1	4	1		258	262	1		192	212	1		164	176	1		220	220	1		248	250
55	232	55A1	3	1		258	260	1		192	212	1	2	164	172	1	2	208	210	1	2	248	254
	233	55A10	3	1		274	278	3		192	212	2		160	170	2		210	230	2		250	254
	234	55A13	3	1		260	264	1		192	210	2		162	164	1	1	210	220	2		248	248
	235	55A2	3	1		258	266	1	1	192	216	2		172	172	2		208	220	2		248	248
	236	55A3	3	1		256	258	1		192	192	1		160	170	1	1	212	220	2		248	252
	237	55A4	3	1		256	260	1		192	192	1		162	170	2		204	224	2		248	254
	238	55A5	3	1		260	286	1		192	216	1		170	170	1	1	206	220	2		252	254

Site	N	Name	E	Amp_2				Amp_3				Amp_4				Amp_5				Amp_6			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	239	55A9	3	1		262	284	1		192	192	1		170	172	2	1	208	210	3		252	260
59	240	59A1	4	1		240	266	1		192	212	1		160	162	1	1	206	220	2		248	254
	241	59A10	4	1		256	260	1		190	210	1		160	174	1	1	210	224	2		248	248
	242	59A11	4	1		254	260	2		192	222	1		170	170	2		210	220	2		250	250
	243	59A12	4	1		256	264	1		192	192	1		160	176	1	1	210	220	2		248	254
	244	59A13	4	1		268	282	1		212	216	1		160	172	2		204	210	2		248	248
	245	59A14	4	1		256	274	2		210	212	2		170	174	2		208	220	2		248	248
	246	59A15	4	1		256	260	1		192	192	2		160	170	2		220	230	2		248	252
	247	59A16	4	1		252	272	1		192	210	1		170	172	1	1	210	220	2		254	270
	248	59A17	4	1		258	260	1		196	210	1		160	172	1	1	204	220	2		254	254
	249	59A18	4	1		256	280	1		192	192	1		162	174	2		228	228	2		248	248
	250	59A19	4	1		272	302	1		192	192	2		164	182	2	1	208	212	2		248	254
	251	59A2	4	2		272	282	1		192	210	1		162	170	1	1	208	228	2		248	248
	252	59A20	4	1		254	284	1		192	216	1		160	160	2		204	222	2		254	254
	253	59A3	4	1		240	266	1		192	194	1		172	174	1	1	208	230	2		250	254
	254	59A4	4	1		258	260	2		192	216	1		162	162	2		208	220	2		254	254
	255	59A5	4	1		256	284	1		192	218	1		160	160	1	1	206	228	2		248	254
	256	59A6	4	1		256	268	1		192	208	2		172	178	2		212	226	2		250	254
	257	59A7	4	1		256	262	1		192	192	1		160	170	1	1	212	220	2		254	254
	258	59A8	4	1		260	272	1		192	192	1		172	172	1	1	220	224	2		254	256
	259	59A9	4	1		258	260	1		192	192	2		160	160	1	1	210	220	1	1	248	254
93	260	93A1	1	2		270	288	2		192	218	2		162	174	2		220	220	2		248	254
	261	93A10	2	2		250	266	2		192	192	3		162	174	2	1	208	218	1		248	254
	262	93A11	2	2		240	260	1		192	192	2		170	170	2		220	220	1		248	248
	263	93A12	2	2		268	276	1		192	192	2		162	162	2		208	230	2		252	254
	264	93A13	2	1		270	288	1		192	220	2		160	170	2		220	220	1		248	254
	265	93A14	2	1		260	292	1		192	212	2		160	170	1	1	204	204	1		248	250
	266	93A15	2	1		260	264	1		194	194	2		160	162	2		204	220	1		250	266
	267	93A16	2	1	1	250	264	1		192	194	2		162	172	2		204	218	1		248	254
	268	93A17	2	2		264	270	1		192	196	2		160	170	2		204	230	1		248	254
	269	93A18	2	2		258	280	2		210	218	2		160	174	2	1	204	208	1		248	266

Site	N	Name	E	Amp_2				Amp_3				Amp_4				Amp_5				Amp_6			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	270	93A19	2	2		228	270	1		192	192	1		170	170	2		204	212	1		248	250
	271	93A2	1	2		258	260	2		192	192	2		162	174	2		208	220	2		248	254
	272	93A3	2	1	1	264	288	2		192	192	1		162	172	2		204	212	2		248	248
	273	93A4	2	2		256	260	2		192	212	1		172	172	2		212	220	2		248	250
	274	93A5	2	2		270	272	2		192	192	1		160	160	2		208	212	2		250	254
	275	93A6	2	2		260	264	2		192	196	1		160	160	2		206	218	2		252	252
	276	93A7	2	2		250	270	2		194	218	1		160	174	1		204	212	2		250	250
	277	93A8	2	2		258	258	2		192	192	1		170	176	2		220	222	2		254	254
	278	93A9	2	2		264	276	2		192	206	1		160	170	2		212	220	1		248	248

NOTE

Extraction method 1 High Pure PCR Template Preparation Kit for blood and tissue following the manufacturer's instructions (Roche Diagnostics GmbH Mannheim, Germany)

Extraction method 2 DNeasy 96, blood and tissue kit following the manufacturer's instructions (Qiagen, Hilden, Germany)

Extraction method 3 Using Chelex® 100 resin, (Bio-Rad, Richmond, CA) (Walsh *et al.* 1991)

Extraction method 4 Gentra Puregene Core Kit A, DNA purification from tissue (Qiagen, Hilden, Germany)

Appendix F - Final Data Set

Table F2. Showing the final genotyping data for *A. sulcicollis* at 5 locis (Amp_8- Amp_11, and Amp_13). *Site* identifies the 17 different sites in this data set; *N* = each individual's position in the dataset (same order used in all analyses) and *Name* is the name given to each sample. For each loci, *R* = number of times each sample was repeated successfully, *F* = Number of failed samples, and *Allele* are the two alleles found in this sample after binning.

Site	N	Name	Amp_8				Amp_9				Amp_10				Amp_11				Amp_13			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
104	1	104A1	1		144	144	1		242	242	1		262	268	1		314	314	1		162	162
	2	104A10	1		136	140	1		228	228	1		260	260	2		314	316	1		162	168
	3	104A11	1		136	148	1		228	228	1		260	260	1		292	292	2		164	164
	4	104A12	1		140	158	1		230	230	1		262	268	1		374	374	1		142	142
	5	104A13	1		136	136	1		228	234	1		260	274	1		314	316	2		162	172
	6	104A14	1		144	148	1		244	244	2		268	292	1		302	302	2		168	168
	7	104A15	1		138	138	1		228	228	1		262	268	1		296	296	1		164	164
	8	104A16	1		140	140	1		228	228	1		260	260	1		284	284	2		168	168
	9	104A17	1		146	150	1		228	258	1		258	268	2		316	334	2		162	166
	10	104A18	1		146	146	1		228	228	1		254	260	1		314	318	1		160	162
	11	104A19	2		144	144	1		228	228	1		260	268	1		310	364	1		170	170
	12	104A2	1		136	140	1		228	228	1		256	260	1		316	316	1		138	138
	13	104A20	2		138	140	1		230	232	1		254	262	1		292	376	2		162	170
	14	104A3	1		148	150	1		226	226	1		260	280	1		314	314	2		170	170
	15	104A4	1		152	152	1		226	228	1		254	254	1		310	324	2		160	164
	16	104A5	1		136	154	1		228	228	1		260	264	1		324	324	1	1	164	164
	17	104A6	1		158	158	1		228	228	1		260	282	1		312	312	2		170	170
	18	104A7	1		138	152	1		228	228		3	-9	-9	1		320	320	1		164	164
	19	104A8	1		138	156	2		270	272	1		258	268	1		304	304	1		160	160
	20	104A9	1		138	138	2	1	226	226	1		260	268	3		366	366	2	1	182	182
109	21	109A1	1		138	140	1		232	232	1		260	282	1		366	366	1		138	138
	22	109A10	1		146	146	1		228	232	1		260	260	2		292	300	2		168	170
	23	109A11	1		136	148	1		224	226	1		262	262	2		382	382	1		162	162
	24	109A12	1		146	146	1		226	226	1		260	260	3		296	296	1		162	162
	25	109A13	1		136	136	1		228	230	1		268	268	1	1	296	366	1		160	168

Site	N	Name	Amp_8				Amp_9				Amp_10				Amp_11				Amp_13			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	26	109A14	2		140	140	1		228	242	1		280	280	1		292	292	1		142	164
	27	109A15	2		148	148	1		226	244	1		258	258	2		268	292	1		160	166
	28	109A16	2		136	148	1		228	228	1		256	256	1		282	294	1		142	160
	29	109A17	2		136	148	2		230	244	1	1	258	258	1		288	294	1		160	162
	30	109A18	1		144	144	1		228	228		3	-9	-9	1		292	292	1		166	166
	31	109A2	1		152	158	1		224	228	1		260	260	2		292	294	1		142	160
	32	109A3	2		140	140	3		244	244	1		264	270	2		294	294	2		160	162
	33	109A4	1		146	154	1		230	230		3	-9	-9	1		300	300	1		138	160
	34	109A5	1		138	146	1		230	230	1		260	260		3	-9	-9	1		162	162
	35	109A6	1		140	148	1		226	244	1		260	260	1		292	292	1		138	138
	36	109A7	1		144	144	1		228	230	1		260	264	1		290	296	1		160	162
	37	109A8	1		138	138	1		228	228	1		258	260	2		358	358	1		162	162
	38	109A9	1		136	150	1		232	232	1		284	284	1		284	314	1		160	160
108	39	108A1	2		146	152	2		228	228	2		264	264	2		290	290	2		162	162
	40	108A10	2		140	144	1		228	228	1		262	268		4	-9	-9	2		178	178
	41	108A11	2		144	158	1		228	230	1		256	284	2		296	368	1		160	162
	42	108A12	2		140	140	1		230	230	2		274	286	2		292	292	1		162	162
	43	108A13	1		140	144	2		226	230	1		258	260	1		292	316	1		160	162
	44	108A14	1		146	154	1		228	228	2		268	280	1		282	298	1		160	162
	45	108A15	1		136	140	2		230	232	1		260	260	1		294	298	1		160	166
	46	108A16	1		136	148	2		228	228	1		254	260	1		370	370	1		138	138
	47	108A17	1		144	148	2		228	228	1		258	258	1		348	348	1		142	142
	48	108A18	1		140	146	1		244	244	1		258	260		4	-9	-9	1		162	162
	49	108A19	1		140	144	1		228	244	1		262	282	2	1	414	414	1		146	166
	50	108A2	2		136	144	2		228	228		4	-9	-9	2		292	294	2		162	170
	51	108A20	1		140	148	2		230	230	1	1	260	260	1		302	302	2		162	162
	52	108A3	2		140	146	2		230	232	2		260	270	2		288	328	2		160	162
	53	108A4	2		144	144	2		228	228		3	-9	-9	2		302	302	2		166	166
	54	108A5	1		140	146	1		230	242	1		258	268	2		288	302	2		162	162
	55	108A6	2		150	162	1		230	230	1		266	272	1		292	292	1		162	162
	56	108A7	1		136	146	1		226	230	1		262	262	1		290	294	1		144	162

Site	N	Name	Amp_8				Amp_9				Amp_10				Amp_11				Amp_13			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	57	108A8	2		148	152	1		228	228	1		258	286	1		318	318	1		160	160
	58	108A9	2		140	140	1		232	232	1		262	262	1		290	290	1		160	160
102	59	102A1	1		136	146	1		228	240	1		260	260	1		290	296	1		160	168
	60	102A10	1		138	150	1		232	232	1		260	260	1		292	292	1		160	160
	61	102A11	1		138	156	1		226	228	1		260	260	2		296	380	1		160	160
	62	102A12	1		148	150	1		226	230	2		256	256	1	1	376	376	1		138	160
	63	102A13	1		136	140	1		226	248	2		260	260	1	1	294	378	2		162	166
	64	102A14	1		136	138	1		226	228	1		260	260	1		294	378	1		162	166
	65	102A15	1		136	136	1		226	228	1		260	264	1		290	376	1		160	162
	66	102A16	1		136	148	1		230	232	1		260	270	1		296	296	1		142	160
	67	102A17	1		146	154	1		228	228	1		260	272	1	2	296	296	2		158	158
	68	102A18	1		138	156	1		228	228	2		270	300	2		340	340	1		138	160
	69	102A19	1		136	146	1		226	226	1		270	270	1		298	298	1		160	160
	70	102A2	1		136	150	1		226	226	1		260	260	2		378	378	1		138	164
	71	102A20	1		134	134	1		226	226	1		264	284	2		376	376	1		162	162
	72	102A3	1		146	152	1		226	228	1		254	262	2		378	378	1		138	160
	73	102A4	1		138	140	1		244	244	1		256	260	1		294	294	1		160	160
	74	102A5	1		136	152	1		228	228	1		272	272	2		296	296	1		162	162
	75	102A6	1		140	146	1		228	228	1		258	258	1		292	292	1		138	160
	76	102A7	1		140	146	1		228	228	1		270	270	1		296	296	1		166	166
	77	102A8	1		136	136	1		226	226	1		270	282	1		372	372	1		166	166
	78	102A9	1		146	152	1		226	240	1		260	260	1		290	362	1		160	160
112	79	112A1	1		140	140	1		228	236	1		254	254	1		288	356	1		160	166
	80	112A10	1		140	144	1		228	228	2		268	268	1		338	338	1		160	160
	81	112A11	1		136	146	1		228	228	1		260	268	1		294	370	1		160	164
	82	112A12	1		146	148	1		228	228	1		264	272	1		298	298	1		168	168
	83	112A13	1		140	140	1		228	228	1		260	262	1		296	296	1		160	160
	84	112A14	1		140	146	1		244	244	1		270	270	1		292	292	1		160	160
	85	112A15	1		136	144	1		228	228	1		258	260	1		290	290	1		160	160
	86	112A16	1		140	140	1		230	230	1		258	268	1		316	316	1		162	162
	87	112A17	1		140	148	1		228	240	1		264	264	1		290	360	1		146	162

Site	N	Name	Amp_8				Amp_9				Amp_10				Amp_11				Amp_13			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	88	112A18	1		140	148	1		234	234		3	-9	-9	1		292	292	1		162	162
	89	112A19	1		140	154	1		228	228	1	1	260	280	1		356	356	1		142	142
	90	112A2	1		140	144	1		230	246	1		260	260	1		296	348	1		160	166
	91	112A20	1		140	146	3		228	246	2		256	270	2		288	294	1		138	160
	92	112A3	2		138	140	2		228	228	2		258	260	2		294	294	2		160	160
	93	112A4	2		144	146	2		224	240	2		260	260	3		294	314	2		138	160
	94	112A5	2		146	146	2		228	228	2		256	268	2		322	322	2		164	164
	95	112A6	2		140	150	2		228	228	1	1	282	282	2		292	294	2		160	162
	96	112A7	1		140	152	3		230	230	1	1	262	280		3	-9	-9	1		166	166
	97	112A8	1		138	152	1		242	242	1		262	262	1		352	352	1		166	166
	98	112A9	1		146	146	1		228	228	1		262	280	1		294	294	1		162	162
46	99	46A1	1		136	136	1		230	230	1	1	268	270	1		298	356	1		162	164
	100	46A2	1		146	158	1		228	228	1		282	282	1		274	274	1		162	162
	101	46A3	1		140	152	1		228	228	1		256	256	1		294	296	2		160	160
	102	46A4	1		140	150	1		228	228	1		260	272	1	1	358	358	1		160	160
	103	46A5	1		152	154	1		234	234	1		260	270	2		308	308	1		160	160
96	104	96A1	1		136	152	1		230	230	1		260	266	3		296	362	1		162	166
	105	96A10	1		136	154	1		228	228	1		260	266	1		294	316	1		162	170
	106	96A11	1		136	150	1		228	228	1		282	282	1		294	294	3		138	170
	107	96A12	1		140	146	2		224	228	1		258	258	3		294	298	2		160	166
	108	96A13	1		140	148	1		228	244	1		260	260	1		294	296	2		160	162
	109	96A14	1		136	140	1		228	232	1		258	258	1		294	296	2		160	162
	110	96A15	1		136	136	2		228	228	1		258	258	2		290	302	2	1	160	160
	111	96A16	1		148	154	1		228	240	2		260	260	1		292	292	2		160	164
	112	96A17	1		136	136	1		228	228	1		268	268	1		294	294	2		160	164
	113	96A18	1		136	146	2		228	228	1		260	260	1		296	296	2		160	160
	114	96A19	1		136	140	1		230	230	1		260	260	1		296	296	1		162	162
	115	96A2	1		144	152	1		230	230	1		260	260	2		386	386	1		160	160
	116	96A3	1		156	156	1		230	230	1		260	264		3	-9	-9	1		142	142
	117	96A4	1		152	152	1		240	242	1		260	260	3		298	384	1		160	160
	118	96A5	1		136	148	1		242	244	1		260	260	3		296	370	1		160	166

Site	N	Name	Amp_8				Amp_9				Amp_10				Amp_11				Amp_13			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	119	96A6	1		154	154	1		228	228	1		268	268	1		298	310	1		160	170
	120	96A7	1		140	152	1		226	228	1	1	260	260	1		292	292	1		160	166
	121	96A8	1		140	140	1		240	240	1		282	282	1		310	310	1		172	172
	122	96A9	1		136	146	1		228	228	1		268	268	1		296	296	2		162	162
9	123	09A1	1		146	146	1		228	228	2		260	260	1		292	294	3		160	160
	124	09A10	1		136	138	1		228	228	1		284	284	2		370	370	2		142	160
	125	09A11	1		140	156	1		228	242	1		284	284	1		292	292	2		142	162
	126	09A12	1		140	144	1		242	242		3	-9	-9	1	2	368	368	2		162	162
	127	09A13	1		136	154	1		228	228	1		268	268	1		288	288	3		160	160
	128	09A14	1		144	148	1		228	228	1		260	268	2		294	404	2		160	162
	129	09A15	1		146	150	1		228	228	1		258	260	1		292	296	2		162	162
	130	09A16	1		140	144	2		232	244	1		260	260	1		298	302	2		162	162
	131	09A17	1		136	140	1		228	228	1		260	284		3	-9	-9	2		142	142
	132	09A18	1		140	146	1		228	228	2		260	260	1		292	292	3		178	178
	133	09A19	1		136	146	1		230	230	1		268	268	1		294	294	2		160	162
	134	09A2	1		144	146	1		228	230	1		260	260	2		292	374	2		138	160
	135	09A20	1		136	136	1		242	244	2		260	276	2		298	370	3		138	138
	136	09A3	1		156	156	1		228	246	2		258	296	2		320	380	2		142	162
	137	09A4	2		136	140	1		230	242	1		260	260	1		294	294	2		138	160
	138	09A5	1		140	146	1		230	246	1		260	260	2		296	392	2		160	168
	139	09A6	1		146	146	1		230	230	1		284	284	1		288	358	2		160	162
	140	09A7	1		136	140	1		228	228	1		268	268		3	-9	-9	2	1	142	142
	141	09A8	1		134	140	1		228	230	1		262	270	1		294	294	2		162	162
	142	09A9	1		136	136	1		228	230	1		256	260	1		294	296	2		162	164
116	143	116A1	1		136	148	1		230	242		3	-9	-9	1		294	294	1		138	160
	144	116A10	1	2	136	150	1		228	228	2		260	288	1		298	298	3		162	162
	145	116A11	2		138	146	2		228	242		4	-9	-9	2		292	320	2		160	166
	146	116A12	2		136	148	2		228	228	2		298	298	2		290	294	2		138	162
	147	116A13	2		136	140	2		228	228	1		262	268	2		290	304	2		160	164
	148	116A14	2		136	136	2		228	228	1		284	284	2		294	294	2		142	142
	149	116A15	2		136	136	2		226	230	1		270	270	2		294	376	2		160	164

Site	N	Name	Amp_8				Amp_9				Amp_10				Amp_11				Amp_13			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	150	116A16	2		136	154	2		228	228	1		256	272	2		298	300	2		160	162
	151	116A17	1		140	148	1		230	230	1		262	262	1		290	290	1		160	160
	152	116A18	1		144	146	1		226	226		4	-9	-9	1		298	298	1		144	144
	153	116A19	1		138	148	1		228	244		4	-9	-9	1		298	312	1		138	160
	154	116A2	1		136	148	1		228	228		3	-9	-9	2		294	294	1		160	160
	155	116A20	1		136	146	1		244	244	1	1	262	262	2		332	332	1		166	166
	156	116A3	1		150	150	1		228	228		4	-9	-9	1		294	294	1		164	164
	157	116A4	1		136	146	1		228	228	1	1	260	260	1		294	304	1		162	162
	158	116A5	1		140	150	1		230	244	1		254	256	1		314	356	2		160	162
	159	116A6	1		138	146	1		228	244		4	-9	-9	1		294	294	2		160	168
	160	116A7	1		134	156	1		228	228	2		286	286	1		294	294	2		160	162
	161	116A8	1		140	140	1		230	244		4	-9	-9	2		272	298	1		142	142
	162	116A9	1		136	150	1		228	244		4	-9	-9	1		298	302	1		160	160
113	163	113A1	1		136	150	1		228	228	1		260	268	1		296	296	1		160	160
	164	113A10	2		144	152	2		228	228	1		260	272	2		294	294	2		164	164
	165	113A11	2		146	146	2		228	230	1		254	270	2		300	318	2		160	170
	166	113A12	2		140	148	2		228	228	1		260	286	2		294	294	2		168	168
	167	113A13	2		150	152	2		228	228	1		268	280	2		290	318	2		160	170
	168	113A14	2		136	138	2		228	228	1		260	262		3	-9	-9	2		166	166
	169	113A15	1		140	146	1		228	240	1		268	270	1		290	304	2		160	166
	170	113A16	1		144	144	1		228	228	1		258	260	1		296	296	1	1	160	160
	171	113A17	2		136	138	2		228	228	2		268	268	2		296	296	2		160	160
	172	113A18	2		140	146	2		234	234	2	1	262	302	2		292	292	2		162	162
	173	113A19	2		152	152	2		228	228	2		286	286	2		294	294	3		172	172
	174	113A2	1		136	140	1		226	230	1		260	260	1		290	322	1		160	162
	175	113A20	2		148	148	2		232	232		4	-9	-9	2		298	298	2		164	164
	176	113A3	1		136	136	1		228	242	1		260	262	1		296	364	1		138	160
	177	113A4	1		138	152	1		242	242	1		288	288	1		314	314	2		174	174
	178	113A5	1		136	148	1		228	230	1		262	262	1		294	296	1		160	160
	179	113A6	1		150	150	1		228	244	1		260	262	1		292	294	1		160	162
	180	113A7	1		144	144	1		228	230	1		260	260	1		292	292	1		162	166

Site	N	Name	Amp_8				Amp_9				Amp_10				Amp_11				Amp_13			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	181	113A8	1		138	140	1		228	228		4	-9	-9	1		298	302	1		160	162
	182	113A9	2		140	140	2		228	228	1		260	260	2		292	298	2		160	166
95	183	95A1	1		136	146	1		228	228	1		256	270	1		276	276	1		164	174
	184	95A10	1		144	152	1		236	236	1		270	286	1		298	298	1		160	160
	185	95A11	1		136	136	1		234	234	1		286	286	1		276	278	1		138	162
	186	95A12	1		144	148	1		230	230	1		282	286	1		294	294	1		138	138
	187	95A13	1		138	146	1		230	240	1		286	286	2		290	384	1		138	160
	188	95A14	1		144	148	1		230	230	1		282	286	1		296	296	1		138	138
	189	95A15	1		136	148	1		226	226	1		270	286	1		278	278	1		166	166
	190	95A16	1		136	144	1		230	230	1		262	266	1		276	276	1		160	160
	191	95A17	1		138	150	2		270	272	1		256	286	2		274	274	2		162	162
	192	95A18	1		140	150	1		226	226	1		258	260	1		296	296	1		138	138
	193	95A19	1		138	144	1		228	228	1		270	270	1		294	294	1		162	162
	194	95A2	1		136	150	2		228	230	1		260	260	1		276	276	1		160	160
	195	95A20	1		144	148	1		234	236		3	-9	-9	1		276	276	1		160	162
	196	95A21	1		150	150	1		230	230	1		262	270	1		276	276	1		162	162
	197	95A22	1		136	140	1		236	236	1		262	270	2		280	298	1		138	162
	198	95A3	1		136	154	1		234	234	1		270	270	1		276	276	1		160	160
	199	95A4	2		138	140	1		230	230	1	1	268	270	2		276	276	2		142	160
	200	95A5	1		146	152	1		230	230	1		270	270	1		276	276	1		160	160
	201	95A6	1		146	150	1		234	234	1		260	286	1		292	292	2		162	162
	202	95A7	1		136	140	1		228	236	1		260	286	1		278	290	1		138	162
	203	95A8	1		140	144	1		228	228	1		286	286	1		318	318	1		138	138
	204	95A9	1		140	140	1		224	234	1		270	282	2		292	292	1		162	162
6	205	06A1	2		144	148	3		246	246	2		270	270	3	1	290	290	3	1	160	160
	206	06A10	1		144	150	1		228	230	1		270	286	1		276	276	2		138	162
	207	06A11	2		136	136	1		226	226	2		286	286		3	-9	-9	2		138	138
	208	06A12	1		136	146	1		230	230	2		262	262	1		276	276	2		160	160
	209	06A13	1		144	146	1		230	230	1		270	270	1		278	278	2		164	164
	210	06A14	1		134	136	1		228	228	1		286	286	1		296	296	1	2	138	160
	211	06A15	2		140	148	1		228	240	1	1	286	286	2		276	276	1	2	138	162

Site	N	Name	Amp_8				Amp_9				Amp_10				Amp_11				Amp_13			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	212	06A16	1		138	146	1	1	232	232	2		270	286	1		298	298	2		138	138
	213	06A17	2		146	158	1		230	230	2		282	282	1	1	278	278	1	2	160	160
	214	06A18	1		138	138	1		228	230	1		270	286	1		276	276	2		142	162
	215	06A19	1	1	136	148	1		228	230	2		282	286	1	1	278	292	1	2	162	162
	216	06A2	1		140	152	2		228	234	1		260	266	2	1	278	374	2		142	162
	217	06A20	1		136	150	1		228	234	1		260	260	1		278	298	2		162	162
	218	06A21	1		140	146	1		230	234	2		286	286	1		276	276	2		160	160
	219	06A3	1		136	136	1	1	226	232	1		256	270	2		276	278	3		162	162
	220	06A4	2		140	150	1	2	230	246	2	1	286	286	2	1	276	278	4		160	160
	221	06A5	2		140	146	2		230	230		3	-9	-9	2		278	292	3		138	138
	222	06A6	1		136	150	2		228	230	2		260	260	2		318	368	2		138	160
	223	06A7	1		140	146	2		228	230	2		270	286	2		276	302	2		138	138
	224	06A8	1	1	140	140	3		228	238	2		286	286	3	1	276	290	1	1	138	162
	225	06A9	1		138	152	1		226	226	1		270	270	1		294	294	3		138	138
94	226	94A2	1		146	148	1		228	228	1		260	280	1	1	294	294	1		142	160
	227	94A21	1		136	146	1		240	240	1		270	286	1		290	290	1		138	138
	228	94A22	1		140	156	1		228	228	1		260	270	1		278	290	2		162	162
	229	94A23	1		146	154	1		228	228	1		286	286	1		278	278	1		142	142
	230	94A3	1		150	150	2		228	228	2		260	260		4	-9	-9	1		158	158
4	231	04A1	1		136	148	1		230	244	1		260	260	2		294	294	2		134	162
55	232	55A1	1	1	140	152	1	1	226	230		3	-9	-9	1	1	292	292		3	-9	-9
	233	55A10	1		144	146	2		240	240	1		264	264	1		316	316	2		160	160
	234	55A13	1		144	152	1		228	244	1		270	270	1		322	330	2		166	166
	235	55A2	1		146	146	1		258	258	2		260	282		3	-9	-9	3		162	162
	236	55A3	1		144	144	1		228	240	1		254	254	2		330	370	2		160	160
	237	55A4	2		146	152	1		228	244	1		260	262	1	1	312	378	2	1	146	172
	238	55A5	1		136	136	1		230	230	1		260	260	1	1	378	378	2		160	160
	239	55A9	1		152	152	1		228	228		3	-9	-9	1		290	290	3		160	160
59	240	59A1	1		146	154	1		226	228	1		256	260	1		318	320	2		164	166
	241	59A10	1	1	136	150	1		228	228	1	1	260	284	1		316	352	2		142	162
	242	59A11	1		144	152	1		242	242	2		260	270	1		330	330	2		162	162

Site	N	Name	Amp_8				Amp_9				Amp_10				Amp_11				Amp_13			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	243	59A12	1		140	148	1		228	228	1		260	266	1		382	382	2		142	142
	244	59A13	1		134	138	1		232	232	1		260	262	1		360	360	2		160	160
	245	59A14	1		136	138	1		228	228	2		270	270	1		358	372	2		160	166
	246	59A15	1		136	136	1		228	242	1		262	262	1		290	380	2		160	172
	247	59A16	1		148	156	1		228	228	1		260	270	1		292	292	3		160	160
	248	59A17	1		144	148	1		228	228	1		262	270	1		362	382	2		160	170
	249	59A18	1		148	152	1		230	250	1		256	256	1		292	322	2		160	160
	250	59A19	2		124	140	2		264	264	1	1	260	268	1		358	358	3		142	142
	251	59A2	1	1	144	150	1		228	248	2		268	284	1		292	318	2		160	176
	252	59A20	1		136	136	1		246	246	1		260	260	1		294	294	2		162	162
	253	59A3	1		138	150	1		228	240	1		260	260	1		374	374	2		162	166
	254	59A4	2		140	148	1		250	250	1		262	262	1		292	292	2		160	160
	255	59A5	1		146	152	1		250	250	1		268	268	1		292	292	3		160	160
	256	59A6	1		146	148	1		240	240	1		260	260	1		296	296	2		162	162
	257	59A7	1		144	148	1		240	246	1		270	288	2		292	388	2		138	142
	258	59A8	1		148	150	2		230	266	1		256	256	1		296	316	3		162	162
	259	59A9	1	1	136	146	1		228	228	1		268	268	1		286	356	2		138	162
93	260	93A1	2		136	146	2		228	228	2		262	262	2		294	318	2		160	160
	261	93A10	2		136	138	2		228	228	1		262	280	2		320	320	2		160	160
	262	93A11	2		122	138	2		228	260	1		268	270	1		296	296	1		160	160
	263	93A12	1		140	146	2		228	228	1		264	264	1		296	362	1		160	160
	264	93A13	1		136	144	1		228	228	1		254	282	1		298	318	1		160	166
	265	93A14	1		136	144	2		226	238	1		260	260	1		288	314	1		160	160
	266	93A15	1		136	146	1		228	228	1		260	260	1		320	320	1		160	160
	267	93A16	1		138	144	1		228	228	1		266	284	2		294	294	1		160	160
	268	93A17	2		138	148	1		228	228	1		266	282	2		318	318	1		160	160
	269	93A18	2		136	144	2		228	228	1		260	282	2		294	314	2		142	162
	270	93A19	2		150	150	1		228	228	1		260	260	1		318	318	1		158	158
	271	93A2	2		136	136	2		228	228	2		256	256	2		296	296	2		160	164
	272	93A3	2		146	146	2		228	228	2		260	260	2		314	314	2		160	160
	273	93A4	2		146	146	2		230	230	2		260	270	2		314	314	3		160	160

Site	N	Name	Amp_8				Amp_9				Amp_10				Amp_11				Amp_13			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	274	93A5	2		146	150	2		228	228	2		260	270	2		270	270	2		160	160
	275	93A6	2		136	150	2		228	242	3		256	260	2		296	318	2		142	160
	276	93A7	2		136	136	2		226	228	3		260	262	2		316	318	2		162	164
	277	93A8	2		140	148	2		226	228	3		254	254	2		286	322	2		160	160
	278	93A9	2		146	146	2		228	228	2		260	282	2		318	318	2		164	164

Table F3*i-x*. Each table represents one microsatellite loci and shows the method of binning for each allele after scoring. The data in these tables are from the complete dataset. Each row *N* represents a different *Allele*. *Min* and *Max* represent the minimum and maximum value (in base pairs) that can be binned into one 'Allele'. *No. alleles* is the total number of times this allele has been scored, not including homozygotes. *Within* is the range (in base pairs) between *Min* and *Max*, *Between* is the difference between the *Max* of one allele and the *Min* of the subsequent allele, and *Between alleles* represents the difference between the *Allele* and the subsequent *Allele*.

Table F3i-x. *i*. Amp_2**.**

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	228.5	229.4	228	4	0.9	6.9	8
2	236.3	236.7	236	2	0.4	2.1	2
3	238.8	239.6	238	2	0.8	0.7	2
4	240.3	241.1	240	19	0.8	1.3	2
5	242.4	243	242	4	0.6	1.5	2
6	244.5	245	244	3	0.5	3.3	4
7	248.3	249.6	248	7	1.3	0.9	2
8	250.5	251	250	7	0.5	1.4	2
9	252.4	252.7	252	2	0.3	1.6	2
10	254.3	254.7	254	5	0.4	1.2	2
11	255.9	257.2	256	88	1.3	0.5	2
12	257.7	258.7	258	55	1	0.9	2
13	259.6	260.5	260	100	0.9	0.9	2
14	261.4	262.2	262	41	0.8	1.1	2
15	263.3	264	264	28	0.7	1.2	2
16	265.2	265.9	266	16	0.7	1.1	2
17	267	269.3	268	10	2.3	1.4	2
18	270.7	271.6	270	25	0.9	0.9	2
19	272.5	273	272	27	0.5	1.3	2
20	274.3	274.9	274	20	0.6	1.2	2
21	276.1	276.9	276	23	0.8	1	2
22	277.9	278.9	278	46	1	0.9	2
23	279.8	280.7	280	36	0.9	1	2
24	281.7	282.1	282	23	0.4	1.4	2
25	283.5	283.9	284	13	0.4	1.5	2
26	285.4	285.9	286	8	0.5	1.3	2
27	287.2	287.5	288	8	0.3	1.7	2
28	289.2	289.7	290	2	0.5	1.3	2
29	291	291.2	292	5	0.2	1.7	2
30	292.9	293	294	3	0.1	1.6	2
31	294.6	295	296	6	0.4	1.4	2
32	296.4	297	298	5	0.6	1.3	2
33	298.3	298.3	300	3	0	3.9	2
34	302.2	302.3	302	2	0.1	5.7	6
35	308	308.4	308	2	0.4	11.5	12
36	319.9	320.1	320	2	0.2		

Table F3i-x. *ii.* Amp_3.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	189.6	189.8	190	6	0.2	1.7	2
2	191.5	192.1	192	277	0.6	1.6	2
3	193.7	194	194	11	0.3	1.7	2
4	195.7	196	196	33	0.3	1.8	2
5	197.8	198.1	198	9	0.3	2.1	2
6	200.2	200.4	200	3	0.2	4	4
7	204.4	204.7	204	7	0.3	2	2
8	206.7	207.1	206	8	0.4	1.8	2
9	208.9	209.3	208	7	0.4	1.6	2
10	210.9	211.6	210	74	0.7	1.5	2
11	213.1	214.1	212	67	1	1.3	4
12	215.4	216.1	216	25	0.7	1.5	2
13	217.6	218.2	218	12	0.6	1.6	2
14	219.8	220	220	7	0.2	2.2	2
15	222.2	222.8	222	2	0.6	1.5	2
16	224.3	224.6	224	2	0.3	1.9	2
17	226.5	226.9	226	8	0.4	1.7	2
18	228.6	228.9	228	3	0.3		

Table F3i-x. *iii*. Amp_4.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	156.7	157	156	5	0.3	3.7	4
2	160.7	161.4	160	109	0.7	1.3	2
3	162.7	163.2	162	98	0.5	1.5	2
4	164.7	165.3	164	13	0.6	1.6	2
5	166.9	167.2	166	2	0.3	1.4	2
6	168.6	168.9	168	8	0.3	1.7	2
7	170.6	171.2	170	148	0.6	1.3	2
8	172.5	173.3	172	108	0.8	1.2	2
9	174.5	175.2	174	113	0.7	1.3	2
10	176.5	177.1	176	36	0.6	1.5	2
11	178.6	179	178	9	0.4	1.7	2
12	180.7	181.2	180	2	0.5	1.5	2
13	182.7	183.1	182	2	0.4		

Table F3i-x. *iv*. Amp_5.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	201.4	201.6	202	2	0.2	1.7	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
2	203.3	204	204	83	0.7	1.5	2
3	205.5	206	206	21	0.5	1.4	2
4	207.4	208.1	208	70	0.7	1.4	2
5	209.5	210.2	210	69	0.7	1.5	2
6	211.7	212.4	212	60	0.7	1.4	2
7	213.8	214.3	214	18	0.5	1.7	2
8	216	216.4	216	5	0.4	1.8	2
9	218.2	218.3	218	7	0.1	1.7	2
10	220	220.7	220	161	0.7	0.8	2
11	221.5	222.7	222	33	1.2	1.5	2
12	224.2	224.7	224	24	0.5	1.6	2
13	226.3	226.6	226	18	0.3	1.7	2
14	228.3	229	228	50	0.7	1.5	2
15	230.5	230.8	230	26	0.3	1.6	2
16	232.4	232.9	232	13	0.5	1.7	2
17	234.6	235.1	234	7	0.5	1.7	2
18	236.8	237	236	3	0.2	2.1	2
19	239.1	239.1	238	2	0		

Table F3i-x. *v*. Amp_6.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	235.6	235.8	236	2	0.2	10.9	11
2	246.7	247.2	246	2	0.5	1.1	2
3	248.3	249.5	248	254	1.2	0.8	2
4	250.3	251.2	250	73	0.9	1.1	2
5	252.3	253.2	252	46	0.9	0.8	2
6	254	255.6	254	188	1.6	0.4	2
7	256	256.7	256	31	0.7	1.3	2
8	258	258.6	258	4	0.6	1.1	2
9	259.7	260.4	260	4	0.7	5.1	6
10	265.5	265.6	266	2	0.1	3.5	4
11	269.1	269.8	270	33	0.7		

Table F3i-x. *vi*. Amp_8.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	121.1	121.1	122	2	0	3.7	2
2	124.8	124.9	124	2	0.1	9.8	10
3	134.7	135.1	134	5	0.4	1.7	2
4	136.8	137.3	136	111	0.5	1.1	2
5	138.4	139.5	138	46	1.1	1.6	2
6	141.1	141.7	140	104	0.6	1.5	4
7	143.2	143.8	144	55	0.6	1.7	2
8	145.5	146.2	146	88	0.7	1.3	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
9	147.5	148.2	148	60	0.7	1.3	2
10	149.5	150.5	150	42	1	1.1	2
11	151.6	152.5	152	37	0.9	1.4	2
12	153.9	154.6	154	15	0.7	1.6	2
13	156.2	156.6	156	9	0.4	1	2
14	157.6	158.6	158	8	1	2.9	4
15	161.5	162.7	162	2	1.2		

Table F3i-x. ***vii***. Amp_9.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	223.9	224.1	224	7	0.2	1.7	2
2	225.8	227.1	226	40	1.3	0.7	2
3	227.8	228.5	228	207	0.7	0.5	2
4	229	231.2	230	82	2.2	0.8	2
5	232	232.5	232	19	0.5	1.7	2
6	234.2	234.6	234	14	0.4	1.9	2
7	236.5	236.7	236	5	0.2	1.7	2
8	238.4	238.8	238	5	0.4	1.4	2
9	240.2	240.8	240	18	0.6	1.4	2
10	242.2	242.7	242	19	0.5	1.5	2
11	244.2	245.1	244	28	0.9	1.5	2
12	246.6	246.9	246	12	0.3	1.7	2
13	248.6	249.1	248	2	0.5	1.5	2
14	250.6	250.8	250	3	0.2	7.4	8
15	258.2	258.4	258	2	0.2	1.8	2
16	260.2	260.6	260	2	0.4	3.2	4
17	263.8	264.1	264	2	0.3	1.5	2
18	265.6	266.1	266	2	0.5	3.3	4
19	269.4	269.9	270	4	0.5	1.4	2
20	271.3	271.7	272	4	0.4		

Table F3i-x. ***viii***. Amp_10.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	255.2	255.9	254	13	0.7	1.4	2
2	257.3	258.1	256	25	0.8	1.3	2
3	259.4	259.9	258	23	0.5	0.3	2
4	260.2	261.8	260	134	1.6	1.2	2
5	263	263.6	262	40	0.6	1.2	2
6	264.8	265.3	264	12	0.5	1.6	2
7	266.9	267.3	266	8	0.4	1.3	2
8	268.6	269.2	268	41	0.6	1.2	2
9	270.4	271.1	270	54	0.7	1.4	2
10	272.5	272.7	272	7	0.2	1.8	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
11	274.5	274.6	274	3	0.1	1.6	2
12	276.2	276.6	276	2	0.4	3.5	4
13	280.1	280.4	280	10	0.3	1.3	2
14	281.7	282.3	282	22	0.6	1.3	2
15	283.6	284.1	284	12	0.5	0.8	2
16	284.9	286	286	38	1.1	1.4	2
17	287.4	287.6	288	4	0.2	3.8	4
18	291.4	291.6	292	2	0.2	4	4
19	295.6	295.9	296	2	0.3	1.8	2
20	297.7	297.9	298	2	0.2	1.6	2
21	299.5	299.9	300	2	0.4	0.7	2
22	300.6	301	302	2	0.4		

Table F3i-x. ***ix***. Amp_11.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	267.1	268.3	268	2	1.2	0.9	2
2	269.2	269.5	270	2	0.3	1.8	2
3	271.3	272.2	272	2	0.9	1.1	2
4	273.3	273.8	274	3	0.5	1.5	2
5	275.3	275.9	276	25	0.6	1.3	2
6	277.2	277.7	278	17	0.5	1.6	2
7	279.3	280.1	280	2	0.8	0.9	2
8	281	281.2	282	2	0.2	1.9	2
9	283.1	283.2	284	2	0.1	3.8	2
10	287	287.3	286	3	0.3	1.6	2
11	288.9	289.3	288	11	0.4	1.6	2
12	290.9	291.4	290	34	0.5	1.4	2
13	292.8	293.5	292	54	0.7	1.3	2
14	294.8	295.5	294	74	0.7	1.3	2
15	296.8	297.5	296	51	0.7	1.3	2
16	298.8	299.5	298	34	0.7	1.4	2
17	300.9	301.4	300	7	0.5	1.5	2
18	302.9	303.4	302	13	0.5	1.7	2
19	305.1	305.3	304	5	0.2	2.1	4
20	307.4	307.7	308	2	0.3	1.7	2
21	309.4	309.6	310	4	0.2	2	2
22	311.6	311.9	312	3	0.3	1.7	2
23	313.6	314.4	314	19	0.8	1.4	2
24	315.8	316.6	316	14	0.8	1.2	2
25	317.8	319	318	26	1.2	1.2	2
26	320	320.5	320	9	0.5	1	2
27	322.3	323.1	322	7	0.8	1.2	2
28	324.3	324.3	324	2	0	4.3	4
29	328.6	328.8	328	2	0.2	1.9	2
30	330.7	331.1	330	4	0.4	1.8	2
31	332.9	334.1	332	2	1.2	0.9	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
32	335	336	334	2	1	1.1	4
33	337.1	337.2	338	1	0.1	2.1	2
34	339.3	340.4	340	2	1.1	7	8
35	347.4	347.7	348	2	0.3	3.8	4
36	351.5	351.5	352	2	0	3.8	4
37	355.3	355.8	356	5	0.5	1.5	2
38	357.3	357.9	358	6	0.6	1.9	2
39	359.8	359.9	360	2	0.1	1.2	2
40	361.1	361.8	362	6	0.7	1.8	2
41	363.6	363.7	364	2	0.1	1.7	2
42	365.4	366.1	366	5	0.7	1.3	2
43	367.4	367.7	368	5	0.3	1.7	2
44	369.4	370	370	11	0.6	1.5	2
45	371.5	371.8	372	2	0.3	1.6	2
46	373.4	374	374	6	0.6	1.3	2
47	375.3	375.8	376	7	0.5	1.6	2
48	377.4	377.8	378	8	0.4	1.5	2
49	379.3	380	380	5	0.7	1.5	2
50	381.5	381.9	382	4	0.4	1.2	2
51	383.1	384.1	384	5	1	1.5	2
52	385.6	385.8	386	2	0.2	1.7	2
53	387.5	387.7	388	2	0.2	4.2	4
54	391.9	392	392	2	0.1	12.2	12
55	404.2	404.4	404	2	0.2	10.1	10
56	414.5	414.9	414	2	0.4		

Table F3i-x. *x*. Amp_13.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	133.4	133.5	134	2	0.1	3.7	4
2	137.2	139	138	60	1.8	2.5	4
3	141.5	143	142	42	1.5	1	2
4	144	144.1	144	2	0.1	1	2
5	145.1	146.3	146	4	1.2	11.3	12
6	157.6	158.1	158	4	0.5	0.9	2
7	159	160.8	160	206	1.8	0.3	2
8	161.1	162.4	162	145	1.3	1.2	2
9	163.6	164.7	164	39	1.1	0.9	2
10	165.6	166.5	166	42	0.9	1.1	2
11	167.6	168	168	16	0.4	1.5	2
12	169.5	170.3	170	22	0.8	1.3	2
13	171.6	172.4	172	10	0.8	1.4	2
14	173.8	174.2	174	3	0.4	1.9	2
15	176.1	176.1	176	2	0	1.3	2
16	177.4	178.1	178	5	0.7	3.7	4
17	181.8	182	182	2	0.2		

Isoperla gramatica – Genotyping and binning

Table F4. Showing the final genotyping data for *I.grammatica* at 5 locis (Iso_1 – Iso_5). *Site* identifies the 13 different sites in this data set; *N* = each individual's position in the dataset (same order used in all analyses); *Name* is the name given to each sample, and *E* = extraction method used (see note at end of table). For each loci *R* = number of times each sample was repeated successfully, *F* = Number of failed samples, and *Allele* are the two alleles found in this sample after binning.

Site	N	Name	E	Iso_1				Iso_2				Iso_3				Iso_4				Iso_5			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
105	1	105I1	1	1		248	304	2		206	222	1		174	178	1		194	194	2		256	262
	2	105I10	2	1		248	250	1		214	224	2		142	180	1		178	182	1		214	214
	3	105I11	2	1		248	250	1		220	234	2		140	140	1		174	180	1		224	252
	4	105I12	2	1		244	262	1		208	208	1		148	148	1		166	196	2	1	214	286
	5	105I13	2	1		250	250	1		242	260	1		140	140	1		194	194	2		218	258
	6	105I14	2	1		248	250	1		214	264	1		140	152	1		192	216	1		224	250
	7	105I15	2	3		242	300	1		202	228	2		144	168	1		194	208	1		226	226
	8	105I2	1	2		258	308	2	1	254	288	1		148	166	1		176	182	2		256	260
	9	105I3	2	2		248	302	1	1	220	308	2		190	224	1	1	214	262	2		254	264
	10	105I4	2	1		266	266	2		222	226	2		164	164	1		166	166	1		218	228
	11	105I5	2	1		242	308	2		222	234	1		146	168	1		178	178	1		218	218
	12	105I6	2	2		242	302	1	1	250	302	1		142	172	1		180	182	2		246	250
	13	105I7	2	1		228	242	1	1	230	260	2		186	188	2		218	218	2		252	252
	14	105I8	2	1		250	308	1	1	222	270	2		140	174	1		176	224	1		216	216
	15	105I9	2	1		260	308	2	1	324	324	2	1	146	190	1		184	184	2		254	254
106	16	106I1	2	1		250	262	2	1	276	290	2	1	174	220	2	1	202	264	2		234	288
	17	106I10	2	1		248	258	1		216	246	3		196	212	1		178	178	2		214	290
	18	106I11	2	1		242	304	1		240	240	2		142	142	2		172	218	2		214	214
	19	106I12	2	1		248	316	1		228	248	2		140	164	1		176	188	2		230	266
	20	106I13	2	2		248	248	2		258	278	2		150	156	1		184	202	2		224	262
	21	106I14	2	2		288	310	2		238	284	3		142	156	1		188	214	2		232	232

Site	N	Name	E	Iso_1				Iso_2				Iso_3				Iso_4				Iso_5			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	22	106I15	2	2		250	304	2		208	238	3		148	178	1		176	178	2		214	214
	23	106I16	2	2		250	250	2		212	212	2		160	160	1		178	198	2		212	224
	24	106I17	2	2		248	248	2		238	248	2		154	166	2		166	242	2		224	226
	25	106I18	2	2	1	242	308	2		236	278	2		146	146	1		196	220	2		232	232
	26	106I19	2	1		302	310	1		246	250	1		140	140	2		178	256	2		216	260
	27	106I2	2	2	1	250	256	1		214	278	2		176	212	1		168	194	2		254	294
	28	106I20	2	2		250	252	1		216	242	2		168	174	3		208	208	2		258	258
	29	106I3	2	1		250	250	1		250	250	3		142	176	1		176	176	2		224	278
	30	106I4	2	1		268	298	1		232	260	2	2	206	238	1		180	198	2		228	228
	31	106I5	2	1		300	302	1		206	230	2	1	146	206	1		176	220	3		288	288
	32	106I6	2	1		250	252	1		214	244	3		216	218	2		222	222	2		328	328
	33	106I7	2	1		248	310	1		214	258	2		132	142	1		198	198	2		296	296
	34	106I8	2	1		248	302	1		276	294	1		160	178	1		188	242		3	-9	-9
	35	106I9	2	1		248	248	1		212	212	1		168	176	2	2	220	270	2	1	206	260
108	36	108I1	2	1		260	302	2	1	216	294	2	1	142	192	1		168	222	1		256	264
	37	108I10	2	1		242	308	1		216	216	1		148	156	1		178	204	1		232	232
	38	108I11	2	1		260	310	2		206	258	2	1	140	192	1		176	186	1		212	212
	39	108I12	2	1		250	264	2		212	234	3		174	188	1		178	194	1		268	268
	40	108I13	2	1		248	260	2		214	234	3		166	168	1		228	252	2		216	294
	41	108I14	2	2		304	308	2	1	214	272	1		140	142	1		272	272	2		214	292
	42	108I15	2	1		302	310	3		216	216	1		140	148	1		178	178	1		280	280
	43	108I16	2	3		242	304	2	1	212	252	2		168	178	2		180	220	3		310	310
	44	108I17	2	3		248	250	2	1	234	276	1	1	140	140	2		176	180	3		262	298
	45	108I18	2	2		300	302	3		220	228	2		142	180	3		218	302	3		320	320
	46	108I19	2	2		248	312	2	1	236	276	2		168	178	2	1	314	356	2	1	324	324
	47	108I2	2	1		266	304	1		216	230	1		174	174	2	1	268	268	2		258	258
	48	108I20	2	2		248	304	2	1	246	280	2		140	174	2		192	242	2		248	248
	49	108I3	2	2		242	270	1	1	204	246	1		140	166	1		192	238	1		230	230
	50	108I4	2	1		248	300	2		278	292	1		140	146	2	1	174	214	1		244	244
	51	108I5	2	2		242	330	1		204	216	2		166	208	1		196	200	2		214	232
	52	108I6	2	1		250	314	3		222	238	2		168	192	2	1	196	264	1	1	282	316

Site	N	Name	E	Iso_1				Iso_2				Iso_3				Iso_4				Iso_5			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	53	108I7	2	1		262	304	1		236	250	1		140	148	1		222	246		3	-9	-9
	54	108I8	2	1		264	318	2	1	286	308	1		166	170	1		224	234		3	-9	-9
	55	108I9	2	1		304	312	1		236	260	1		170	178		3	-9	-9	1		254	254
12	56	12I1	4	1		302	304	1		278	278	1		146	150	1		164	190	2		306	332
	57	12I10	4	2		248	316	1		240	252	1		156	182	1		194	218	1		224	224
	58	12I11	4	2		250	300	1		228	244	1		146	166	1		178	178	2		224	224
	59	12I12	4	1		300	312	1		216	236	1		148	178	1		190	200	1		234	234
	60	12I13	4	2		242	248	1		220	228	1		140	146	1		178	254	1	1	234	234
	61	12I14	4	1		300	302	1		218	234	1		136	136	1		178	272	1		224	252
	62	12I15	4	1		250	304	1		234	238	1		162	168	1		178	190	1		214	252
	63	12I16	4	2		248	262	2		272	284	1	1	170	232	1		166	232	1		266	310
	64	12I17	4	1		244	264	1		204	216	1		154	166	1		182	204	1		242	294
	65	12I18	4	1		248	250	1		252	294	1		148	160	1		176	182	1		232	232
	66	12I19	4	1		248	250	1		240	302	1		142	142	1		180	180	2		260	260
	67	12I2	4	2		250	292	1		204	318	1		140	152	1		178	212	1		212	212
	68	12I20	4	2		242	286	1		220	320	2		176	176	1		166	248	1		228	254
	69	12I3	4	2		248	316	1		214	250	1		148	168	1		192	212	1		232	232
	70	12I4	4	1		304	316	2		232	300	1		174	178	1		164	220	1		224	248
	71	12I5	4	1		244	250	2		240	240	1	1	186	186	1		170	178	1	1	232	232
	72	12I6	4	1		268	276	1		216	272	1		174	176	1		200	222	1		256	276
	73	12I7	4	1		250	308	1		250	258	1		154	176	1		166	166	2	1	260	312
	74	12I8	4	2		248	250	1		230	236	2		140	178	1		178	194	1		206	224
	75	12I9	4	2		248	304	1		226	240	2		140	140	1		182	194	1		252	278
97	76	97I1	2	1		250	260	2	1	294	294	2		168	190	1	1	214	282	2		214	214
	77	97I10	2	1		242	264	1		214	256	2		148	170	1		194	218	1		214	250
	78	97I11	2	1		258	260	1		214	218	2		140	142	1		188	212	2		252	290
	79	97I12	2	1		266	308	1		254	282	1	2	196	204	1		176	192	1		224	224
	80	97I13	2	2	1	248	328	1		228	228	1		140	156	2		234	234	1		214	214
	81	97I14	2	1		266	310	1		206	206	1		168	178	2	1	226	226	1		214	214
	82	97I15	2	1		264	266	2	1	230	234	2		170	206	1		178	178	3		294	294
	83	97I16	2	2		314	386	1		240	276	2		146	168	2		222	242	3		306	306

Site	N	Name	E	Iso_1				Iso_2				Iso_3				Iso_4				Iso_5			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	84	97I17	2	2		264	304	2		230	282	1		148	152	1		182	182	2		238	238
	85	97I2	2	1		248	264	1	1	256	256	1		166	168	2		174	174	2		274	288
	86	97I3	2	1		252	302	1	1	230	266	1		150	156	2		180	218	2		272	272
	87	97I4	2	1		242	314	1	1	244	254	1		144	146	1		178	188	1	2	286	286
	88	97I5	2	1		250	250	1		214	238	1		156	172	1		178	224	2		278	300
	89	97I6	2	1		242	250	2		214	214	2		166	182	1		184	208	1		234	234
	90	97I7	2	1		302	308	1		246	272	1		168	168	1		186	188	3		330	330
	91	97I8	2	1		248	308	1		286	286	1		154	164	1		166	184	1		252	252
	92	97I9	2	1		242	308	1		212	218	2		166	168	1		178	208	3		298	298
98	93	98I1	2	1		228	316	1		240	282	2		142	196	1	1	216	282	1		224	224
	94	98I10	2	1		234	292	2		222	244	1		160	176	1		164	164	2		256	294
	95	98I11	2	1		248	304	1		208	246	1		152	170	2		184	266		3	-9	-9
	96	98I12	2	1		250	310	1		206	274	1		150	154	1		166	248	2		212	286
	97	98I13	2	1		248	294	1		204	224	1		140	142	1		178	234	2		230	274
	98	98I14	2	2		304	308	2		274	274	1		152	174	1		176	276	2		216	216
	99	98I15	2	1		248	264	1		222	222	1		148	178	1		178	194	2		302	302
	100	98I16	2	1		252	304	1		238	238	1		140	164	1		178	224	2		212	212
	101	98I17	2	1		304	314	3		204	288	1	1	142	168	2	2	178	244	1	1	224	244
	102	98I18	2	1		308	310	2		268	276	1		156	170	2		264	264	2		250	292
	103	98I19	2	1		250	264	1		260	286	1	1	156	220	3		236	236	2		274	284
	104	98I2	2	2		242	248	1		252	252	2		140	148	1	1	178	272	2	1	256	256
	105	98I20	2	1		302	318	1		230	230	1		140	166	1	2	232	276	1		214	214
	106	98I3	2	1		302	308	1		214	216	2		166	166	1		190	254	1		212	212
	107	98I4	2	1		242	310	1		222	246	2	1	178	198	1		182	200	1		214	214
	108	98I5	2	1		274	304	1		216	250	1		152	166	1		178	184		3	-9	-9
	109	98I6	2	1		248	268	1		214	280	2		142	142	2		164	164	2		254	268
	110	98I7	2	1		302	308	1		214	226	2		136	140	2		216	216	2		250	288
	111	98I8	2	1		304	308	1		206	240	1		144	170	2		168	196	2	2	320	348
	112	98I9	2	1		308	316	1		220	250	1		140	150	1		184	198	2		298	298
112	113	112I1	2	1		304	304	1	1	214	256	2		168	168	1		178	180	1		216	242
	114	112I10	2	1	1	252	252	2		228	244	2		156	190	1		176	176	1		224	224

Site	N	Name	E	Iso_1				Iso_2				Iso_3				Iso_4				Iso_5			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	115	112I11	2	1		232	248	1	1	212	246	1		150	176	1		164	176	2		296	296
	116	112I12	2	1		252	308	1	1	234	266	1		148	162	1		184	212	1		214	288
	117	112I13	2	2	1	242	268	1	1	216	248	2		142	146	1		196	196	1		214	224
	118	112I14	2	1		250	304	2		214	216	2		140	176	1		166	204	1		226	280
	119	112I15	2	1		310	322	3		206	206	2		164	202	1		174	252	1		214	252
	120	112I16	2	1		252	264	2	1	230	386	2		170	204	1		190	202	1		214	290
	121	112I17	2	2		242	308	2		206	240	2		162	226	1		174	220	1		248	254
	122	112I18	2	1		300	310	2	1	230	230	2		146	166	1		198	246	1		212	224
	123	112I19	2	1		242	274	2		226	228	2		152	176	1		184	198	1	1	316	316
	124	112I2	2	1		248	264	2		216	220	1		146	146	1		214	214	1		214	260
	125	112I20	2	1		258	308	2		216	226	2		142	190	1		182	182	1	1	304	304
	126	112I3	2	1		302	308	1	1	246	256	2		176	188	1		192	218	1		214	214
	127	112I4	2	1		242	308	1	1	222	270	1		142	148	1		180	180	2		262	262
	128	112I5	2	1		302	310	1	1	224	286	1		148	160	1		168	168	1		230	230
	129	112I6	2	1		296	304	1	1	212	260	1		146	170	1		212	236	2		260	288
	130	112I7	2	1		266	302	2		214	214	2		140	170	2		178	250	1		258	258
	131	112I8	2	1		248	264	1	1	260	292	2		146	194	1		178	192	2		260	286
	132	112I9	2	1		300	312	2	1	240	304	2		142	178	1		174	230	2		234	286
46	133	46I1	3	1		242	316	2		212	214	1		140	168	1		178	186	2		232	256
	134	46I2	3	1		242	310	1	1	212	308	1		140	168	1		178	192	2		276	276
	135	46I3	3	1		242	242	2		206	214	1		140	146	1		190	214	1		224	292
	136	46I4	3	1		254	322	2		214	222	1		168	168	1		178	182	2		232	232
	137	46I5	3	1		248	260	1	1	232	276	2		146	184	1		182	190	2		248	276
114	138	114I1	2	2		250	304	1		214	244	1		150	162	1		166	176	1		226	230
	139	114I10	2	2		228	250	3	1	216	288	2		138	152	1		182	190	2		214	262
	140	114I11	2	2		352	352	2		214	230	2		170	178	1		174	178	2		224	224
	141	114I12	2	2		248	300	2		222	242	2		142	166	1		170	178		5	-9	-9
	142	114I13	2	2		232	258	2		250	250	2		140	142	1		196	196	2	2	268	302
	143	114I14	2	1		264	302	1		214	250	1		148	168	2		204	238	1		224	224
	144	114I15	2	1		304	308	1		226	230	1		140	150	1		174	214	1		214	214
	145	114I16	2	1		242	312	1	1	272	272	2		144	204	2		190	196	2		254	254

Site	N	Name	E	Iso_1				Iso_2				Iso_3				Iso_4				Iso_5			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	146	114I17	2	1		248	266	1	1	250	274	1		142	152	1		180	190	2	1	326	326
	147	114I18	2	1		250	260	1	1	216	236	2		166	168	2	1	180	274	2		206	218
	148	114I19	2	2		266	266	2	2	256	362	2		140	168	1		182	198	2		254	254
	149	114I2	2	2		250	260	1		230	230	2		144	168	1		168	168	1		214	224
	150	114I20	2	1		250	308	1	1	242	320	1		142	172	1		182	196	2		292	292
	151	114I3	2	2		248	310	1		212	232	2		140	140	1		160	160	2		232	254
	152	114I4	2	2		250	296	1		270	272	1		140	148	2		206	206	2	2	344	344
	153	114I5	2	2		308	308	1		250	250	1		140	150	2		188	188	2	1	218	218
	154	114I6	2	2		268	300	2	1	216	272	2		162	172	1	1	176	278	1		228	228
	155	114I7	2	2		242	304	2	1	214	286	1		154	168	1		178	204	2		214	256
	156	114I8	2	2		250	308	2		204	208	3		146	194	2		172	228	4		224	270
	157	114I9	2	2		250	308	3	1	214	274	3		140	196	2		178	214	2		232	250
115	158	115I1	2	2		264	308	1	1	264	272	2		174	178	1		170	170	2		262	262
	159	115I10	2	3		252	302	3		204	230	2		154	168	1		178	178	2		240	240
	160	115I11	2	2		232	298	2	1	224	270	2	2	174	230	1		168	168	2		224	260
	161	115I12	2	2		248	264	2	2	240	302	2		168	178	1		218	228	3		290	302
	162	115I13	2	2		242	308	3		216	222	2		154	174	1		178	224	2		224	256
	163	115I14	2	3		242	302	2	1	246	312	3		140	192	1		172	224	2		224	264
	164	115I15	2	1		242	300	1	1	248	308	2		174	200	1		164	190	1		214	214
	165	115I16	2	1		254	264	1	1	272	282	2		186	196	1		174	174	1		216	260
	166	115I17	2	1		292	298	1	1	208	292	2		150	170	2		216	258	2		214	258
	167	115I18	2	1		248	310	1	1	218	298	2		146	156	1		178	180		3	-9	-9
	168	115I19	2	1		302	314	3		214	214	1		156	182	2		208	218	2		242	252
	169	115I2	2	2		248	304	2		214	216	1		142	148	1		178	224	2		216	264
	170	115I20	2	1		234	294	1	1	222	244	1		142	168	1		166	182	2		254	254
	171	115I3	2	2		258	314	1	1	268	312	1		160	172	1		174	182	2		212	254
	172	115I4	2	2		264	276	2		226	230	1		146	146	1		192	192	2		268	268
	173	115I5	2	2		248	304	1	1	258	268	2		162	166	1		190	194	2		214	262
	174	115I6	2	3		250	250	2		214	224	2		168	168	1	1	176	224	2		214	242
	175	115I7	2	2		248	248	2		222	226	1		146	168	1		190	242	2		224	248
	176	115I8	2	2		248	282	1	1	270	308	2		176	176	2		224	230	2		214	214

Site	N	Name	E	Iso_1				Iso_2				Iso_3				Iso_4				Iso_5			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	177	115I9	2	2		310	314	2	1	238	298	2		140	180	1	1	212	276	2		212	212
10	178	10I1	4	1		268	302	1		224	240	2		140	140	2	1	182	322	2		302	302
	179	10I10	4	1	1	248	248	2		232	264	1		140	148	1		168	218	1		250	282
	180	10I11	4	1		270	314	1		220	274	1		148	174	1		214	214	1		246	252
	181	10I12	4	2		248	304	1		222	298	2		142	142	1		214	216	2		256	312
	182	10I13	4	2		248	264	1		230	318	1		142	166	1		176	268	1		212	248
	183	10I14	4	1		242	302	1		216	256	1		150	182	1		236	236	1		214	232
	184	10I15	4	1		310	310	1		270	282	1		142	156	1		200	214	1		224	252
	185	10I16	4	2		250	250	2		216	374	1		140	148	1		190	198	1		214	248
	186	10I17	4	2		242	254	2		310	310	2		148	188	1		188	198	1		232	260
	187	10I18	4	2		244	248	1		274	278	1		170	174	1		164	172	2		214	214
	188	10I19	4	1		250	312	1		232	274	1		140	146	1		178	230	1		214	214
	189	10I2	4	1		248	268	1		216	272	1		168	174	1		174	178	1		232	298
	190	10I20	4	1		274	312	1		206	214	1		170	176	1		176	194	1		256	298
	191	10I3	4	1		248	264	1		216	320	1		148	164	2		196	286	1		216	216
	192	10I4	4	1		242	248	2		232	232	2		170	170	1		180	184	2		224	224
	193	10I5	4	1		302	310	1		230	324	1		150	176	1		218	234		3	-9	-9
	194	10I6	4	1		242	294	1		214	228	1		142	172	1		180	190	2		258	328
	195	10I7	4	1		302	304	2		202	264	1		142	142	1		182	182	2		304	304
	196	10I8	4	1		302	302	1		216	278	1		140	168	1		180	228	2		220	248
	197	10I9	4	2		248	248	1		248	274	1		170	176	1		196	220	1		232	300
118	198	118I1	2	1		248	250	1		212	226	2		146	162	2		196	244	2		214	276
	199	118I10	2	2		302	304	2		230	242	2		144	146	2		222	222	2		214	224
	200	118I11	2	2		242	248	3	1	214	254	3		142	180	2		188	242	2		214	266
	201	118I12	2	2		242	312	2		214	238	3		152	174	2		208	208	3		214	286
	202	118I13	2	2		240	300	2		238	276	3		178	178	2		176	232	1	1	218	218
	203	118I14	2	3		248	300	2	1	206	336	3		150	232	2	1	182	260	2		232	232
	204	118I15	2	1		250	304	2	1	244	296	2		174	228	1		186	196	1		218	258
	205	118I16	2	1		242	300	1		216	220	2		176	176	1		166	168	1		216	216
	206	118I17	2	1		308	318	1		216	274	1	1	142	220	2		182	220	2		252	290
	207	118I18	2	1		304	318	1		206	214	2		168	178	1		178	178	1		206	216

Appendix F - Final Data Set

Site	N	Name	E	Iso_1				Iso_2				Iso_3				Iso_4				Iso_5			
				R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	208	118I19	2	1		242	250	2		270	316	1	1	152	228	1		172	202	2		224	252
	209	118I2	2	2		248	248	1		214	248	2		168	176	1		182	212	1		214	224
	210	118I20	2	1		250	314	1		204	224	2		140	156	2		186	232	2		252	252
	211	118I3	2	1		242	248	2		280	302	3		168	222	2		180	238	2		214	274
	212	118I4	2	2		248	312	1		218	278	2		146	160	2		180	224	2		254	254
	213	118I5	2	2		242	242	1		254	256	2		148	148	2		192	254	2		250	250
	214	118I6	2	2		248	308	3		206	206	2		164	168	1		174	176	1		226	232
	215	118I7	2	2		242	320	1		218	224	2		140	140	1		174	174	1		232	232
	216	118I8	2	1		250	270	1		216	226	2		142	182	1		216	222	1		252	252
	217	118I9	2	2		248	248	2		204	212	2		140	154	1		178	178	2		224	224
93	218	93I1	2	2		248	290	3		206	214	1		178	178	2		176	250	2		224	234
	219	93I10	2	1		302	308	1		214	240	1		150	178	1		170	196	1		206	232
	220	93I11	2	1		248	250	1		254	284	2		166	180	1		178	208	1		214	256
	221	93I12	2	1		242	282	1		212	218	2	1	174	222	2	1	180	256	1		214	264
	222	93I13	2	1		248	248	1		220	220	1		140	178	1		184	218	1		214	286
	223	93I14	2	1		248	248	1		216	218	1		140	146	1		178	190	1		252	252
	224	93I15	2	1		248	248	1		214	216	1		140	162	1		176	182	1		234	264
	225	93I16	2	1		248	320	1		218	218	1		140	166	1		220	250	2	2	316	336
	226	93I17	2	2		248	248	2	1	206	220	2		140	168	1	1	224	228	3		214	214
	227	93I18	2	2		250	302	2	1	212	246	2		140	146	1	1	206	278	3		214	224
	228	93I19	2	2		244	248	3		218	224	2		170	170	2	1	160	312	3		214	214
	229	93I2	2	2		304	304	2	1	204	258	1		140	180	2		190	222	2		260	260
	230	93I20	2	2		248	248	3		206	220	2		148	148	3		190	262	3		228	228
	231	93I3	2	2		248	302	2	1	236	272		3	-9	-9	2		180	180	2		214	226
	232	93I4	2	2		248	248	2	1	268	274	1		140	170	2		168	190	2		224	244
	233	93I5	2	2		252	308	2	1	214	224	3		190	220	2		172	180	2		280	288
	234	93I6	2	2		302	304	2	1	212	278	1		140	174	2		180	192	2		214	214
	235	93I7	2	2		248	318	1	1	224	250	1		140	178	2		218	258	3		226	226
	236	93I8	2	2		304	312	2		220	242	3		172	190	2		220	232	1		214	224
	237	93I9	2	1		248	304	1		276	290	1		140	172	1		186	218	1		224	258

NOTE

Extraction method 1 High Pure PCR Template Preparation Kit for blood and tissue following the manufacturer's instructions (Roche Diagnostics GmbH Mannheim, Germany)

Extraction method 2 DNeasy 96, blood and tissue kit following the manufacturer's instructions (Qiagen, Hilden, Germany)

Extraction method 3 Using Chelex® 100 resin, (Bio-Rad, Richmond, CA) (Walsh *et al.* 1991)

Extraction method 4 Gentra Puregene Core Kit A, DNA purification from tissue (Qiagen, Hilden, Germany)

Table F5. Showing the final genotyping data for *I.grammatica* at 5 locis (Iso_6 – Iso_10). *Site* identifies the 13 different sites in this data set; *N* = each individuals position in the dataset (same order used in all analyses) and *Name* is the name given to each sample. For each loci *R* = number of times each sample was repeated successfully, *F* = Number of failed samples, and *Allele* are the two alleles found in this sample after binning.

Site	N	Name	Iso_6				Iso_7				Iso_8				Iso_9				Iso_10			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
105	1	105I1	1		290	290	1		124	140	1		276	276	1		272	272	1		216	216
	2	105I10	1		292	308	1		150	150	1		270	270	1		286	286	1		216	216
	3	105I11	1		292	294	2		248	260	1		280	284	1		254	274	1		196	196
	4	105I12	1		302	316	2		140	212	1		300	322	1		332	332	2		220	230
	5	105I13	1		308	308	1		136	194	1		272	272	2	1	454	454	1		218	218
	6	105I14	1		274	324	1		124	124	1		268	268	1		264	316	1		214	218
	7	105I15	1		270	270	1		138	140	1		266	282	2		270	318	2		216	288
	8	105I2	1		326	326	2		156	224	1		264	308	2		340	340	1		212	216
	9	105I3	2		270	270	1		136	172	1		276	322	1		272	310	1		212	212
	10	105I4	1		270	270	2		162	180	1		280	294	1		280	280	1		216	230
	11	105I5	1		294	294	1		124	156	1		312	326	1	1	340	340	1		218	218
	12	105I6	2		300	406	1		126	138	3	2	264	372		3	-9	-9	2		214	250
	13	105I7	1		276	306	1		160	162	1		300	304	1	1	268	268	2		214	256
	14	105I8	1		326	326	1		142	150	1		260	284	1	1	272	362	1		230	230

Site	N	Name	Iso_6				Iso_7				Iso_8				Iso_9				Iso_10			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	15	105I9	2		280	414	1		150	154	1		272	278	2		302	346	1		212	212
106	16	106I1	2		268	406	1		126	126	1		256	268	1		294	294	1		214	214
	17	106I10	1		270	280	1		140	140	2		310	312	1		280	280	1		224	224
	18	106I11	1		270	294	1		140	142	1		264	280	2		250	270	1		214	216
	19	106I12	1		270	324	1		156	194	1		278	302	2	1	358	358	1		216	216
	20	106I13	2		294	294	1		148	148	2		284	288	2		316	316	2		230	230
	21	106I14	2		270	292	1		150	150	2		318	324	3		250	250	2		192	218
	22	106I15	2		286	310	1		170	172	2		314	326	2		250	312	2		210	220
	23	106I16	3		294	330	1		128	150	2		292	304		5	-9	-9	2		236	236
	24	106I17	2		292	304	1		122	192	2		266	310	2		288	360	2		218	218
	25	106I18	2		296	312	1		122	142	2		288	310	2		312	312	2	1	292	292
	26	106I19	2		270	292	1		138	146	1		286	292	2	2	370	386	1		232	232
	27	106I2	1		292	310	1		134	168	2		258	266		4	-9	-9	2	1	276	276
	28	106I20	1		270	304	1		126	140	1		266	272	2		318	318	2		214	214
	29	106I3	1		270	270	1		146	146	1		284	284	1		338	338	1		212	212
	30	106I4	1		270	270	1		124	148	2		272	340	1		302	302	1		216	220
	31	106I5	2		476	476	1		138	142	1		304	312	1		272	272	1		214	214
	32	106I6	1		294	294	1		122	148	1		292	292	2		322	322	1		214	214
	33	106I7	1		270	322	1		150	150	1		292	324	1		266	266	1		214	214
	34	106I8	1		274	286	2		152	238	1		280	280	1	1	308	308	1		208	218
	35	106I9	1		296	296	2		150	212	1		284	300	1		250	250	1		216	226
108	36	108I1	2		346	346	1		148	150	1		304	328	1		314	344	1		212	212
	37	108I10	1		270	270	1		174	184	1		286	286	2		274	274	1		232	232
	38	108I11	2		294	294	2		184	298	1		308	316	2		334	390	1		232	232
	39	108I12	1		292	294	1		124	182	2		272	316	2		330	330	1		218	228
	40	108I13	1		276	276	1		142	202	2		280	314	2		330	330	1		212	212
	41	108I14	1		270	270	2		142	304	1		286	286	2		346	346	1		214	218
	42	108I15	1		290	296	2		146	238	1		276	312		3	-9	-9	1		218	218
	43	108I16	2		288	288	3		124	196	2		272	298		3	-9	-9	2		214	216
	44	108I17	2		272	338	2		130	178	2		256	264	2	1	334	334	2		216	232
	45	108I18	2		272	292	2		148	222	2		268	316	2	1	268	358	2		214	218

Site	N	Name	Iso_6				Iso_7				Iso_8				Iso_9				Iso_10			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	46	108I19	2		312	312	2		154	190	2		292	344	2		274	274	2		268	268
	47	108I2	1		296	296		3	-9	-9	1		256	306		3	-9	-9	1		212	216
	48	108I20	2		294	294	2		182	192	2		256	256	2	1	346	346	2		214	218
	49	108I3	1	1	290	296	2		128	222	1		278	294	1		334	334	2		212	274
	50	108I4	1		270	282	2		128	278	2		278	278	2		250	250	1		212	226
	51	108I5	1		290	290	1		124	186	1		270	280	2		292	292	1		228	228
	52	108I6	1		292	296	1		170	196	1		282	282	2		250	250	2		216	216
	53	108I7	1		282	314	2		232	298	2		318	318	3		410	410	1		214	218
	54	108I8	1		292	304	1		140	140	1		280	314	2		296	394	1		218	218
	55	108I9	1		294	306	1		126	128	1		304	312	2	1	340	414	1		212	212
12	56	12I1	2		442	442	1		128	160	1		284	304		3	-9	-9	1		220	220
	57	12I10	1		274	292	1		138	138	1		310	310	2		340	354	2		194	210
	58	12I11	1		294	298	1		148	150	2		316	320	1		314	314	1		212	218
	59	12I12	1		320	320	1		206	206	1		316	316	1		250	250	1		210	210
	60	12I13	1		286	290	1		146	146	1		312	312	2		336	336	1		216	216
	61	12I14	1		274	274	1		134	134	1		272	272	1	1	250	398	1		214	214
	62	12I15	2		290	362	1		154	154	1		264	286	2		380	380	2		216	216
	63	12I16	1		322	322	1		164	164	1		280	308	2	2	268	422	1		226	226
	64	12I17	2		508	508	1		124	158	2		302	432	1		290	334	1		218	218
	65	12I18	1		270	280	1		154	154	1		340	340	1		326	340	1		220	224
	66	12I19	2		366	366	1		134	156	2		256	294	1		332	346	1		216	216
	67	12I2	1		274	290	1		128	146	1		294	294	1		268	330	1		214	218
	68	12I20	1		290	290	1		142	148	1		256	262	1		268	366	2		216	216
	69	12I3	1		292	292	1		138	186	1		282	308	1		288	324	1		214	214
	70	12I4	1		270	274	1		124	124	1		272	290	2		274	376	1		216	216
	71	12I5	1		280	280	1		136	136	1		272	272	1		306	306	1		214	220
	72	12I6	1		324	324	1		150	150	1		300	300	2		432	432	1		214	214
	73	12I7	1		320	320	1		124	182	1		270	270	1	2	438	438	1		220	220
	74	12I8	1		294	360	1		140	150	1		270	270	1	1	276	276	1		216	216
	75	12I9	1		290	296	1		124	166	1		272	286	1		274	274	1		216	216
97	76	97I1	1		296	296	1		130	182	1		272	272	2	1	378	378	1		188	224

Site	N	Name	Iso_6				Iso_7				Iso_8				Iso_9				Iso_10			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	77	97I10	1		290	290	1		124	154	1		270	278	3		338	338	2		224	252
	78	97I11	1	3	274	274	1		152	176	1		264	302	1		274	274	1		218	228
	79	97I12	2		274	428	2		146	258	1		312	318	1		286	286	1		224	262
	80	97I13	2		310	394	1		136	184	1		280	280	1		276	344	1		218	218
	81	97I14	2		274	274	1		138	138	1		320	320	2	2	454	454	1		218	218
	82	97I15	1		296	296	2		134	250	2	1	380	380	2		348	348	1		228	228
	83	97I16	2		270	398	1		138	138	1		278	278	2	1	400	400	1		212	214
	84	97I17	1		302	302	1		148	150	1		264	296	1		288	288	1		268	268
	85	97I2	1		304	324	1		124	142	1		312	312		3	-9	-9	1		210	214
	86	97I3	2		270	364	1		158	166	1	1	354	354	2	1	360	360	1		212	212
	87	97I4	1		270	306	1		134	156	1		314	314		3	-9	-9	1		218	218
	88	97I5	1		270	270	1		160	184	1		274	274	1		278	278	1		224	224
	89	97I6	1		308	308	1		126	134	2		308	308	2	1	272	346	1		216	216
	90	97I7	1		274	274	1		124	150	1		260	294	1		270	270	1		214	216
	91	97I8	2		332	332	1		154	160	1		300	312	2		274	274	1		210	218
	92	97I9	2		488	488	2		132	218	1		256	268	2	1	326	354	1		210	220
98	93	98I1	1		268	304	1		146	146	1		308	308	2		334	348	1		216	216
	94	98I10	1		278	278	1		142	146	1		282	300	1		276	276	1		220	220
	95	98I11	2		272	272	1		140	146	1	1	278	296	2		348	400	1		216	216
	96	98I12	1		314	340	1		138	202	1		262	264	1		270	316	1		216	218
	97	98I13	1		320	320	2		206	258	1		270	278	1		330	330	1		204	224
	98	98I14	1		292	292	2		126	126	1		312	312	1	1	358	400	2	1	224	244
	99	98I15	1		300	308	1		140	140	1		270	318	1		286	286	1		212	230
	100	98I16	1		294	294	1		150	150	2		296	296	1		250	250	1		210	218
	101	98I17	1		294	294	1		162	168	2		256	256		4	-9	-9	1		216	216
	102	98I18	1		280	308	1		140	170	1		264	264	1		276	288	1		212	214
	103	98I19	1		276	276	1		140	150	1		264	312	1		250	250	1		220	230
	104	98I2	2		374	374	1		138	142	1		272	310	1		288	288	1		210	214
	105	98I20	1		280	280	2		174	188	2		276	276		3	-9	-9	1		218	220
	106	98I3	1		268	290	1		166	166	1		256	298	1		274	294	1		216	218
	107	98I4	1	1	268	284	1		146	160	1		268	282	2		374	374	1		216	216

Site	N	Name	Iso_6				Iso_7				Iso_8				Iso_9				Iso_10			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	108	98I5	1		270	270	1		138	150	1		270	278		3	-9	-9	1		212	266
	109	98I6	2		358	358	1		126	146	1		266	330	1		336	336	1		212	218
	110	98I7	1		270	286	1		184	184	1		256	312	2		284	316	1		214	216
	111	98I8	1		294	294	1		134	166	1		264	268	2		250	346	1		218	230
	112	98I9	1		270	296	1		120	120	1		272	314	2		312	334	1		212	212
112	113	112I1	1		354	354	1		122	152	1		264	278	1		274	274	1		214	214
	114	112I10	2		276	392	1		128	146	2		296	318	1	1	384	384	1		216	218
	115	112I11	1		270	290	1		128	164	1		266	290		3	-9	-9	1		216	216
	116	112I12	1		270	286	1		190	194	2		256	256	2		272	272	1		226	226
	117	112I13	1		276	294	1		134	152	1		266	266	1		278	332	2	1	260	270
	118	112I14	1		290	290	1		128	140	1		264	264	2		344	344	1		216	216
	119	112I15	1		280	308	1		134	142	2		328	352	1		280	280	1		218	218
	120	112I16	1		270	296	1		136	146	1		264	324	1		274	274	1		218	218
	121	112I17	2		274	274	1		128	152	1		266	284	1	1	352	352	1		220	220
	122	112I18	1		282	306	2		154	264	1		316	316	2		330	330	1		212	212
	123	112I19	1		274	274	2		252	252	1		256	256	1		282	332	1		216	216
	124	112I2	1		270	320	1		138	138	2		354	354	1		254	254	1		208	228
	125	112I20	1		282	290	1		140	164	2		308	320	2	2	416	416	1		220	220
	126	112I3	1		306	306	1		128	154	1		256	274	1	1	366	366	1		216	218
	127	112I4	1		270	270	2		136	136	1		290	290	2	2	386	438	1		214	214
	128	112I5	1		274	296	1		140	150	1		268	330	2	1	342	342	1		228	236
	129	112I6	1		294	294	1		124	148	1		288	298	1	1	390	390	1		214	226
	130	112I7	1		274	290	1		166	166	1		292	308	1		274	306	1		212	214
	131	112I8	1		294	314	1		146	146	1		264	304	2		256	314	1		216	218
	132	112I9	1		278	290	1		156	168	1		304	314	1	1	378	378	1		224	224
46	133	46I1	1		302	302	1		150	150	1		300	320	1		250	250	1		218	220
	134	46I2	2		270	410	2		140	140	2		282	328	1	1	312	312	1		214	214
	135	46I3	1		270	270	1		148	176	1		300	322	1	2	422	422	2		368	368
	136	46I4	1		322	322	1		138	250	1		264	292	2		312	312	1		208	218
	137	46I5	2		268	344	1		142	182	1		268	270	1	1	284	284	1		214	218
114	138	114I1		4	-9	-9	3		136	226	1		264	264	2		314	356	2		214	220

Site	N	Name	Iso_6				Iso_7				Iso_8				Iso_9				Iso_10			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	139	114I10	2		294	300	1		150	164	2		272	330	2		250	290	2		214	216
	140	114I11	2		288	288	1		128	128	2		266	302	2		338	338	2		208	218
	141	114I12	2		276	276	2		128	252	2		284	284	3		270	348	2		212	236
	142	114I13	3		322	374	2		152	270	2		296	312	3		306	306	2		216	220
	143	114I14	1		310	310	1		128	128	1		316	316	2	1	336	366	1		216	216
	144	114I15	1		270	300	1		136	148	1		262	278	2	2	398	398	1		190	190
	145	114I16	1		274	292	1		128	140	2		278	278		3	-9	-9	1		226	232
	146	114I17	1		284	288	2		128	216	1		284	284	1	1	348	348	1		216	218
	147	114I18	1		290	310	1		136	154	1		332	332	1	1	332	370	1		220	220
	148	114I19	1		270	304	1		146	158	1		266	286	2		352	352	1		210	210
	149	114I2	1		270	310	1		128	128	1		286	306	2		250	316	2		212	212
	150	114I20		3	-9	-9	2		260	260	1		264	266	1	1	268	338	1		216	216
	151	114I3	1		296	296	1		148	156	1		268	286	1		270	270	2		216	216
	152	114I4	2		342	342	1		128	174	1		288	296	2	2	384	384	2		212	214
	153	114I5	1		302	302	1		124	160	1		270	290	1	1	276	366	2		214	214
	154	114I6	1		270	274	1		148	154	1		306	306	1		318	340	2		210	210
	155	114I7	1		272	272		4	-9	-9	1		274	278	2		280	300	2		220	254
	156	114I8	2		274	290	1		138	138	2		290	290	2		250	250	2		212	216
	157	114I9	2		278	278	1		126	150	2	1	276	306	4		344	344	2		212	212
115	158	115I1	2		296	316	1		138	142	1		264	322	2		276	276	1		204	214
	159	115I10	2		278	324	1		148	148	2		288	296	2		264	270	2		212	232
	160	115I11	2		300	352	1		128	188	2	1	276	390		4	-9	-9	2		212	228
	161	115I12	2		308	324	1		138	142	2		264	268	2	1	316	358	2		218	218
	162	115I13	2	1	474	508	1		138	154	2		260	296	2		250	250	2		214	218
	163	115I14	2		274	298	1		150	156	2		298	298	3		312	318	2		220	224
	164	115I15		3	-9	-9	3		122	122	1		260	292	1		250	250	1		212	214
	165	115I16	1		282	288	2		128	128	1		290	290	2		302	334	1		232	232
	166	115I17	1		276	354	1		158	186	1		282	286	2	1	394	394	1		220	230
	167	115I18	1		286	292	2		164	260	1		266	316	2		278	278	1		216	218
	168	115I19	1		278	278	2		240	240	1		306	322	2		336	348	1		218	218
	169	115I2	2		270	270	1		122	148	1		268	282	1		250	250	1		214	214

Site	N	Name	Iso_6				Iso_7				Iso_8				Iso_9				Iso_10			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	170	115I20	1		294	302	1		124	168	1		266	266	1		250	250	1		214	214
	171	115I3	1		274	294	1		146	146	2		308	346	1		270	270	1		226	226
	172	115I4	1		290	290	1		138	164	1		304	304		3	-9	-9	1		214	220
	173	115I5	1		286	286	1		132	132	1		294	294	1		274	274	1		220	220
	174	115I6	1		304	304	1		148	148	1		270	304	2		250	306	1		224	232
	175	115I7	1		268	270	2		158	252	2		288	288	1		278	278	1		214	220
	176	115I8	2		350	350	2		158	170	2		280	298	1	1	380	380	1		224	224
	177	115I9	2		270	314	1		148	172	3		274	282	3		328	328	2		220	220
10	178	10I1	1		312	330	1		148	148	1		264	304		3	-9	-9	1		212	212
	179	10I10	1		290	310	1		190	190	1		302	314	2		316	418	1		188	218
	180	10I11	1		270	292	1		134	156	1		264	278	1		316	316	1		188	188
	181	10I12	1		292	300	1		186	186	2		264	446	1		272	356	1		220	220
	182	10I13	2		468	468	1		128	128	1		278	278	1		268	268	2		224	224
	183	10I14	1		280	296	1		138	140	1		256	262	1		318	358	1		216	216
	184	10I15	1		272	314	1		126	154	1		284	284	1		290	324	1		216	216
	185	10I16	1		268	300	2		146	146	1		294	308	1	1	334	410	2		214	214
	186	10I17		3	-9	-9	1		122	206	1		318	318	1		332	332	1		218	218
	187	10I18	1		290	290	1		138	140	1		256	284	1		310	362	1		218	218
	188	10I19	2		316	462	1		128	146	1		302	316	1		276	284	2		220	220
	189	10I2	1	2	292	300	1		148	148	1		280	288	1		314	358	1		188	218
	190	10I20	1		274	278	1		128	148	1		256	262	2		312	312	1		212	218
	191	10I3	1		302	308	1		140	150	1		262	296	1		250	250	2		266	266
	192	10I4	1		270	320	2		124	236	1	1	282	414	2		294	294	1		218	226
	193	10I5	1		292	292	1		150	150	1		314	332	1		250	336	1		188	216
	194	10I6	1		300	300	1		128	156	2		300	334	1		266	332	2		218	218
	195	10I7	1		360	360	1		140	164	1		278	280	1		284	284	2		196	242
	196	10I8	1		270	322	1		156	186	1		266	292	1		336	348	2		198	198
	197	10I9	1		296	300	1		128	138	2		264	264	1	1	254	254	1		218	218
118	198	118I1	1		270	330	1		140	140	1		268	268	2		274	360	1		214	214
	199	118I10	2		270	270	1		148	152	2		266	266	2		410	410	2		218	218
	200	118I11	2		270	294	1		130	130	2		308	308	2		378	406	2		210	210

Site	N	Name	Iso_6				Iso_7				Iso_8				Iso_9				Iso_10			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	201	118I12	2		270	270	1		186	250	2		266	304	2		268	268	2		216	216
	202	118I13	2		292	310	1		138	170	2	1	276	276		3	-9	-9	2		188	240
	203	118I14	3		270	360	1		134	134	3		264	286	2		308	372	3		268	268
	204	118I15	1		272	284	1		138	140	2	1	364	414	2	1	418	418	1		218	218
	205	118I16	1		286	298	1		190	190	2		264	320	1		308	316	1		216	240
	206	118I17	2		330	330	2		150	198	1		282	284	1		250	274	1		210	226
	207	118I18	1		300	300	2		150	196	1		276	286	2		370	370	1		214	216
	208	118I19	1		270	270	2		148	194	1		286	286	1		250	250	1		226	226
	209	118I2	1		270	270	2		126	200	1		302	302	2		268	348	1		216	216
	210	118I20	1		298	312	2		138	262	2		300	344		3	-9	-9	1		210	214
	211	118I3	1		278	278	1		126	126	1		268	318	2		362	370	1		212	212
	212	118I4	1		274	274	1		148	182	1		276	376	1		270	270	1		214	218
	213	118I5	1		288	288	1		182	182	1		296	296	1		250	250	1		218	218
	214	118I6	1		290	290	1		134	134	1		264	306	1		272	312	1		216	218
	215	118I7	1		290	316	1		134	168	1		256	280	1		250	250	1		220	220
	216	118I8	1		278	278	1		146	152	2		278	278	2		250	336	1		220	220
	217	118I9	2		294	294	2		150	260	2		270	270	2	1	336	380	2		216	216
93	218	93I1	3		270	270	3	1	212	248	3		298	316		5	-9	-9	2		188	214
	219	93I10	1		270	270	1		124	142	1		316	316	1		292	292	1		188	216
	220	93I11	1		302	316	1		126	146	1		276	276	1		250	282	1		188	188
	221	93I12	1		270	340	1		134	134	1		268	318	1		348	348	1		240	240
	222	93I13	1		270	312	1		134	148	1		264	300	1		250	268	1		190	214
	223	93I14	1		284	298	1		120	122	1		264	322	1		330	346	1		188	188
	224	93I15	1		270	292	1		134	134	1		256	264	1		266	316	1		214	228
	225	93I16	1		270	274	1		124	124	1		264	320	2		364	364	1		188	228
	226	93I17	2		274	312	2		128	128	2		264	316	2	1	330	358	2		218	226
	227	93I18	2		270	310	2		124	124	3		320	376		4	-9	-9	2		220	220
	228	93I19	2		282	302	2		128	194	2		282	302	2	1	250	330	2		188	188
	229	93I2	3		294	304	4		130	130	3		256	322	4	1	348	348	1	1	188	188
	230	93I20	2		324	340	2		140	198	2		256	298	2	1	316	374	2		188	272
	231	93I3	3		284	306	4		122	122	2	1	264	384	3	1	374	374	2		188	188

Site	N	Name	Iso_6				Iso_7				Iso_8				Iso_9				Iso_10			
			R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele	R	F	Allele	Allele
	232	93I4	3		270	300	4		178	252	3		298	318	3	1	348	348	2		188	188
	233	93I5	3		308	312	4		182	182	3		266	266	3	1	360	360	2		188	262
	234	93I6	3		270	336	4		142	186	3		264	264	3	1	360	360	2		228	228
	235	93I7	2		270	310	3		186	186	1	1	264	266	2	1	346	346	2		188	188
	236	93I8	2		270	308	3		178	256	2		292	292		5	-9	-9	2		188	228
	237	93I9	1		270	346	1		128	150	1		256	264	1		250	312	1		188	188

Table F6*i-x*. Each table represents one microsatellite loci and shows the method of binning for each allele after scoring. The data in these tables are from the complete dataset. Each row *N* represents a different *Allele*. *Min* and *Max* represent the minimum and maximum value (in base pairs) that can be binned into one *Allele*. *No. alleles* is the total number of times this allele has been scored, not including homozygotes. *Within* is the range (in base pairs) between *Min* and *Max*, *Between* is the difference between the *Max* of one allele and the *Min* of the subsequent allele, and *Between alleles* represents the difference between the *Allele* and the subsequent *Allele.*

Table F6i-x. *i*. Iso_1.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	227.9	228.1	228	4	0.2	3.9	4
2	232	232.3	232	5	0.3	1.8	2
3	234.1	234.4	234	2	0.3	5	6
4	239.4	239.7	240	2	0.3	1.4	2
5	241.1	242.3	242	64	1.2	1.5	2
6	243.8	244.3	244	7	0.5	2.7	4
7	247	249.1	248	116	2.1	0.6	2
8	249.7	251.1	250	65	1.4	0.9	2
9	252	252.8	252	13	0.8	1.1	2
10	253.9	254.5	254	4	0.6	0.9	2
11	255.4	256.1	256	2	0.7	1.5	2
12	257.6	258.6	258	9	1	1.3	2
13	259.9	260.8	260	10	0.9	1.2	2
14	262	262.3	262	5	0.3	0.6	2
15	262.9	264.4	264	24	1.5	1.1	2
16	265.5	266.2	266	9	0.7	1.2	2
17	267.4	268.2	268	9	0.8	1.1	2
18	269.3	269.8	270	4	0.5	3.3	4
19	273.1	273.7	274	3	0.6	1.4	2
20	275.1	275.7	276	3	0.6	6.5	6
21	282.2	283	282	3	0.8	2.2	4
22	285.2	285.4	286	2	0.2	1.9	2
23	287.3	287.4	288	2	0.1	1.9	2
24	289.3	289.4	290	2	0.1	3.6	2
25	293	293.3	292	4	0.3	1.6	2
26	294.9	295.1	294	3	0.2	1.7	2
27	296.8	296.9	296	3	0.1	1.8	2
28	298.7	298.9	298	4	0.2	1.7	2
29	300.6	301	300	24	0.4	1.7	2
30	302.7	303.1	302	43	0.4	1.7	2
31	304.8	305.3	304	54	0.5	1.6	4
32	306.9	307.4	308	46	0.5	1.5	2
33	308.9	309.3	310	21	0.4	1.8	2
34	311.1	311.6	312	14	0.5	1.5	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
35	313.1	313.6	314	12	0.5	1.7	2
36	315.3	315.8	316	9	0.5	1.7	2
37	317.5	317.6	318	6	0.1	2	2
38	319.6	319.9	320	3	0.3	1.5	2
39	321.4	321.4	322	2	0	7	6
40	328.4	328.5	328	2	0.1	1.3	2
41	329.8	330	330	2	0.2	22.5	22
42	352.5	352.6	352	2	0.1	33.1	34
43	385.7	385.7	386	2	0		

Table F6i-x. ii. Iso_2.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	202.5	202.9	202	3	0.4	1.4	2
2	204.3	205.2	204	18	0.9	1.1	2
3	206.3	208	206	30	1.7	0.6	2
4	208.6	209.6	208	7	1	1	4
5	210.6	211.7	212	22	1.1	0.9	2
6	212.6	213.7	214	61	1.1	1.2	2
7	214.9	215.9	216	44	1	1.2	2
8	217.1	218.1	218	12	1	1	2
9	219.1	220	220	20	0.9	1.2	2
10	221.2	222.5	222	26	1.3	0.9	2
11	223.4	224.5	224	16	1.1	0.8	2
12	225.3	226.3	226	15	1	1.3	2
13	227.6	228.4	228	13	0.8	1.3	2
14	229.7	231.4	230	27	1.7	0.2	2
15	231.6	233	232	10	1.4	1	2
16	234	235.1	234	14	1.1	0.9	2
17	236	237.1	236	11	1.1	1.5	2
18	238.6	239.5	238	18	0.9	0.7	2
19	240.2	242.1	240	17	1.9	0.4	2
20	242.5	243.7	242	9	1.2	0.7	2
21	244.4	245.9	244	11	1.5	0.6	2
22	246.5	247.8	246	14	1.3	1	2
23	248.8	249.8	248	7	1	0.6	2
24	250.4	252.2	250	14	1.8	0.7	2
25	252.9	253.6	252	5	0.7	1.4	2
26	255	255.7	254	9	0.7	0.9	2
27	256.6	257.2	256	8	0.6	0.9	2
28	258.1	259.6	258	9	1.5	0.7	2
29	260.3	261.6	260	7	1.3	2.9	4
30	264.5	265.8	264	6	1.3	0.6	2
31	266.4	266.4	266	2	0	1.7	2
32	268.1	269.1	268	6	1	0.8	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
33	269.9	271.3	270	9	1.4	0.6	2
34	271.9	273.2	272	15	1.3	0.6	2
35	273.8	275.1	274	14	1.3	0.7	2
36	275.8	276.5	276	14	0.7	1.5	2
37	278	278.6	278	13	0.6	0.9	2
38	279.5	280.5	280	5	1	1.4	2
39	281.9	282.5	282	6	0.6	1.4	2
40	283.9	284.8	284	5	0.9	0.7	2
41	285.5	286.2	286	7	0.7	1.2	2
42	287.4	288.2	288	8	0.8	1.2	2
43	289.4	289.8	290	3	0.4	1.4	2
44	291.2	291.6	292	4	0.4	1.7	2
45	293.3	293.9	294	6	0.6	1.4	2
46	295.3	295.6	296	2	0.3	1.6	2
47	297.2	297.6	298	4	0.4	1.7	2
48	299.3	299.7	300	2	0.4	1.2	2
49	300.9	301.9	302	6	1	1.1	2
50	303	304	304	2	1	3.1	4
51	307.1	307.7	308	6	0.6	2	2
52	309.7	309.9	310	2	0.2	1.3	2
53	311.2	312.2	312	3	1	3.8	4
54	316	316.6	316	2	0.6	1.3	2
55	317.9	318.2	318	2	0.3	1.5	2
56	319.7	320.5	320	3	0.8	3.3	4
57	323.8	324.4	324	3	0.6	12	12
58	336.4	336.6	336	2	0.2	24.7	26
59	361.3	361.7	362	2	0.4	12.4	12
60	374.1	374.1	374	2	0	11.3	12
61	385.4	385.7	386	2	0.3		

Table F6i-x. iii. Iso_3.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	131.2	131.2	132	2	0	4.1	4
2	135.3	135.4	136	3	0.1	2	2
3	137.4	137.6	138	2	0.2	1.7	2
4	139.3	139.9	140	85	0.6	1.7	2
5	141.6	142	142	54	0.4	1.9	2
6	143.9	144.2	144	10	0.3	1	2
7	145.2	146.6	146	42	1.4	1.6	2
8	148.2	148.7	148	33	0.5	1.6	2
9	150.3	150.8	150	18	0.5	1.5	2
10	152.3	152.8	152	15	0.5	1.7	2
11	154.5	154.9	154	13	0.4	1.6	2
12	156.5	157	156	20	0.5	3.6	4

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
13	160.6	161.2	160	9	0.6	1.1	2
14	162.3	163.2	162	12	0.9	1.3	2
15	164.5	165	164	11	0.5	1.2	2
16	166.2	167	166	33	0.8	1.1	2
17	168.1	169.2	168	61	1.1	1	2
18	170.2	171	170	27	0.8	1.1	2
19	172.1	172.8	172	11	0.7	1.2	2
20	174	175	174	35	1	1.3	2
21	176.3	177.1	176	27	0.8	1.1	2
22	178.2	178.9	178	35	0.7	1.2	2
23	180.1	180.7	180	12	0.6	1.8	2
24	182.5	182.8	182	7	0.3	1.5	2
25	184.3	185	184	2	0.7	1.5	2
26	186.5	187	186	5	0.5	1.1	2
27	188.1	188.9	188	9	0.8	1	2
28	189.9	190.8	190	16	0.9	1.2	2
29	192	192.6	192	9	0.6	1.9	2
30	194.5	195.4	194	5	0.9	0.9	2
31	196.3	196.7	196	11	0.4	1.8	2
32	198.5	198.7	198	2	0.2	1.8	2
33	200.5	200.7	200	2	0.2	1.8	2
34	202.5	202.5	202	2	0	2.1	2
35	204.6	205	204	5	0.4	1.8	2
36	206.8	207.3	206	6	0.5	1.4	2
37	208.7	208.7	208	2	0	2.3	4
38	211	211.4	212	5	0.4	3.9	4
39	215.3	215.6	216	3	0.3	1.8	2
40	217.4	217.7	218	3	0.3	1.3	2
41	219	219.6	220	7	0.6	1.7	2
42	221.3	222.5	222	5	1.2	1	2
43	223.5	223.6	224	2	0.1	2.1	2
44	225.7	226.6	226	2	0.9	1.5	2
45	228.1	228.2	228	3	0.1	1.7	2
46	229.9	230.6	230	2	0.7	1.4	2
47	232	232.6	232	4	0.6	5.9	6
48	238.5	239.4	238	2	0.9		

Table F6i-x. iv. Iso_4.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	160.9	161	160	3	0.1	3.9	4
2	164.9	165.1	164	8	0.2	1.4	2
3	166.5	167	166	13	0.5	0.8	2
4	167.8	169	168	11	1.2	1.6	2
5	170.6	171	170	4	0.4	1.7	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
6	172.7	173.1	172	9	0.4	1.4	2
7	174.5	174.9	174	15	0.4	1.1	2
8	176	177	176	25	1	0.6	2
9	177.6	179.1	178	50	1.5	0.6	2
10	179.7	181	180	31	1.3	0.6	2
11	181.6	182.3	182	24	0.7	0.6	2
12	182.9	185	184	12	2.1	1.6	2
13	186.6	187.1	186	7	0.5	1.5	2
14	188.6	188.9	188	11	0.3	1.6	2
15	190.5	190.9	190	24	0.4	1.6	2
16	192.5	193	192	14	0.5	1.6	2
17	194.6	195.7	194	12	1.1	0.9	2
18	196.6	197	196	19	0.4	1.6	2
19	198.6	198.8	198	9	0.2	1.9	2
20	200.7	200.9	200	5	0.2	1.8	2
21	202.7	202.8	202	5	0.1	2.1	2
22	204.9	205.4	204	6	0.5	1.6	2
23	207	207.3	206	3	0.3	1.5	2
24	208.8	209.5	208	11	0.7	1.7	4
25	211.2	211.7	212	7	0.5	1.5	2
26	213.2	213.8	214	13	0.6	1.6	2
27	215.4	216	216	8	0.6	1.5	2
28	217.5	218.2	218	21	0.7	1.5	2
29	219.7	220.4	220	14	0.7	1.4	2
30	221.8	222.5	222	12	0.7	1.2	2
31	223.7	224.8	224	13	1.1	1.3	2
32	226.1	226.2	226	2	0.1	1.9	2
33	228.1	228.8	228	6	0.7	1.3	2
34	230.1	230.9	230	4	0.8	1.6	2
35	232.5	233.9	232	8	1.4	0.6	2
36	234.5	235.2	234	5	0.7	1.4	2
37	236.6	237.1	236	5	0.5	0.8	2
38	237.9	239.4	238	5	1.5	1	4
39	240.4	242.7	242	10	2.3	0.8	2
40	243.5	244.2	244	4	0.7	1	2
42	245.2	245.6	246	2	0.4	1.1	2
43	246.7	246.9	248	2	0.2	2.4	2
44	249.3	250.2	250	5	0.9	0.9	2
45	251.1	252	252	2	0.9	1.6	2
46	253.6	254.9	254	4	1.3	0.5	2
47	255.4	255.9	256	4	0.5	1.4	2
48	257.3	257.9	258	4	0.6	1.6	2
49	259.5	259.6	260	2	0.1	1.2	2
50	260.8	261.5	262	4	0.7	0.7	2
51	262.2	263.6	264	6	1.4	1.6	2
52	265.2	265.3	266	2	0.1	1.7	2
53	267	267.3	268	3	0.3	1.5	2
54	268.8	269.1	270	2	0.3	1	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
55	270.1	271	272	3	0.9	1.6	2
56	272.6	272.8	274	2	0.2	1.9	2
57	274.7	275.6	276	3	0.9	1.2	2
58	276.8	277.9	278	2	1.1	5	4
59	282.9	283.4	282	2	0.5	2.6	4
60	286	286.1	286	2	0.1	15	16
61	301.1	301.8	302	2	0.7	10.2	10
62	312	312.6	312	2	0.6	1.9	2
63	314.5	314.6	314	2	0.1	7	8
64	321.6	322.4	322	2	0.8	32.3	34
65	354.7	355.6	356	2	0.9		

Table F6i-x. v. Iso_5.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	205.4	206.1	206	7	0.7	5.7	6
2	211.8	212.4	212	15	0.6	0.8	2
3	213.2	214.6	214	76	1.4	0.4	2
4	215	216.6	216	14	1.6	1.7	2
5	218.3	218.6	218	10	0.3	2	2
6	220.6	220.7	220	2	0.1	2.9	4
7	223.6	224.8	224	58	1.2	0.9	2
8	225.7	227	226	11	1.3	1.2	2
9	228.2	228.9	228	8	0.7	1	2
10	229.9	231.2	230	7	1.3	1.2	2
11	232.4	233.3	232	27	0.9	1.4	2
12	234.7	235.2	234	10	0.5	1.8	4
13	237	237.3	238	2	0.3	1.8	2
14	239.1	239.2	240	2	0.1	1.9	2
15	241.1	243	242	6	1.9	1.3	2
16	244.3	245.2	244	4	0.9	1.4	2
17	246.6	247.4	246	3	0.8	1.8	2
18	249.2	249.8	248	12	0.6	0.9	2
19	250.7	251.7	250	13	1	0.8	2
20	252.5	253.5	252	22	1	1.5	2
21	255	255.5	254	23	0.5	0.7	2
22	256.2	257.5	256	20	1.3	1.2	2
23	258.7	259.3	258	13	0.6	1.3	2
24	260.6	261.1	260	21	0.5	1.3	2
25	262.4	263.7	262	15	1.3	0.7	2
26	264.4	264.8	264	9	0.4	1.5	2
27	266.3	266.6	266	5	0.3	1.6	2
28	268.2	268.5	268	7	0.3	1.7	2
29	270.2	270.5	270	4	0.3	1.7	2
30	272.2	272.4	272	2	0.2	1.4	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
31	273.8	274.3	274	8	0.5	1.4	2
32	275.7	276.1	276	7	0.4	1.4	2
33	277.5	278	278	5	0.5	1.4	2
34	279.4	279.7	280	4	0.3	1.4	2
35	281.1	281.5	282	2	0.4	1.8	2
36	283.3	283.4	284	2	0.1	1.6	2
37	285	285.6	286	12	0.6	1.3	2
38	286.9	287.3	288	14	0.4	1.1	2
39	288.4	289.2	290	10	0.8	1.5	2
40	290.7	291.2	292	7	0.5	0.9	2
41	292.1	293	294	10	0.9	1.5	2
42	294.5	294.8	296	4	0.3	1.2	2
43	296	296.7	298	10	0.7	1.5	2
44	298.2	299.7	300	3	1.5	2	2
45	301.7	302.4	302	9	0.7	1.3	2
46	303.7	304.4	304	3	0.7	1.4	2
47	305.8	306.5	306	5	0.7	3.2	4
48	309.7	310.6	310	4	0.9	1.5	2
49	312.1	313.8	312	4	1.7	2.3	4
50	316.1	316.9	316	4	0.8	3.5	4
51	320.4	320.7	320	5	0.3	3.7	4
52	324.4	324.6	324	2	0.2	0.9	2
53	325.5	326.6	326	3	1.1	1.8	2
54	328.4	328.5	328	4	0.1	2	2
55	330.5	330.9	330	3	0.4	0.9	2
56	331.8	332.4	332	2	0.6	2.2	4
57	334.6	336.4	336	2	1.8	7.8	8
58	344.2	344.3	344	2	0.1	3.7	4
59	348	348.4	348	2	0.4		

Table F6i-x. vi. Iso_6.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	268.8	269.4	268	9	0.6	0.4	2
2	269.8	271.4	270	82	1.6	1.4	2
3	272.8	273.3	272	9	0.5	1	2
4	274.3	275.3	274	29	1	1.2	2
5	276.5	277.1	276	9	0.6	1.3	2
6	278.4	279	278	10	0.6	1.4	2
7	280.4	280.8	280	9	0.4	1.5	2
8	282.3	282.8	282	7	0.5	1.5	2
9	284.3	284.6	284	7	0.3	1.5	2
10	286.1	286.6	286	9	0.5	1.5	2
11	288.1	289.1	288	7	1	0.8	2
12	289.9	291.2	290	25	1.3	0.7	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
13	291.9	292.2	292	26	0.3	0.6	2
14	292.8	294.2	294	34	1.4	1.3	2
15	295.5	296.1	296	18	0.6	1.5	2
16	297.6	297.9	298	6	0.3	1.5	2
17	299.4	299.8	300	17	0.4	1.8	2
18	301.6	302.7	302	9	1.1	0.9	2
19	303.6	304.8	304	11	1.2	0.9	2
20	305.7	306.9	306	8	1.2	0.7	2
21	307.6	308.2	308	14	0.6	0.7	2
22	308.9	310.3	310	15	1.4	0.6	2
23	310.9	312.4	312	12	1.5	1.6	2
24	314	315.3	314	6	1.3	1	2
25	316.3	318.4	316	7	2.1	0.9	4
26	319.3	320.7	320	5	1.4	0.7	2
27	321.4	323.5	322	7	2.1	0.6	2
28	324.1	325.6	324	10	1.5	0.7	2
29	326.3	326.5	326	2	0.2	3.2	4
30	329.7	330.9	330	7	1.2	2	2
31	332.9	333	332	2	0.1	3.9	4
32	336.9	337.2	336	3	0.3	0.6	2
33	337.8	337.9	338	2	0.1	0.9	2
34	338.8	340.2	340	4	1.4	0.7	2
35	340.9	341	342	2	0.1	2.7	2
36	343.7	344.2	344	2	0.5	1.5	2
37	345.7	346	346	3	0.3	3.7	4
38	349.7	349.9	350	2	0.2	1.7	2
39	351.6	352.1	352	2	0.5	1.5	2
40	353.6	353.7	354	2	0.1	3.8	4
41	357.5	357.6	358	2	0.1	1	2
42	358.6	359.6	360	5	1	1.8	2
43	361.4	361.4	362	2	0	1.9	2
44	363.3	363.5	364	2	0.2	1.6	2
45	365.1	365.3	366	2	0.2	7.7	8
46	373	373.6	374	5	0.6	17.8	18
47	391.4	391.5	392	2	0.1	2.1	2
48	393.6	395.3	394	2	1.7	3.5	4
49	398.8	398.9	398	2	0.1	6.8	8
50	405.7	405.9	406	4	0.2	4.7	4
51	410.6	410.6	410	2	0	3.9	4
52	414.5	414.7	414	2	0.2	13.9	14
53	428.6	429	428	2	0.4	12.4	14
54	441.4	441.7	442	2	0.3	19.7	20
55	461.4	461.4	462	2	0	5.7	6
56	467.1	467.4	468	2	0.3	7.3	6
57	474.7	474.7	474	2	0	2	2
58	476.7	477	476	2	0.3	10	12
59	487	487.1	488	2	0.1	20.5	20
60	507.6	510.1	508	4	2.5		

Table F6i-x. vii. Iso_7.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	119.8	120.2	120	2	0.4	1.3	2
2	121.5	122.4	122	14	0.9	1.2	2
3	123.6	124.6	124	25	1	1.2	2
4	125.8	126.5	126	15	0.7	1.1	2
5	127.6	128.5	128	36	0.9	1.3	2
6	129.8	130.5	130	8	0.7	1.5	2
7	132	132.3	132	3	0.3	1.8	2
8	134.1	134.8	134	17	0.7	1.1	2
9	135.9	136.4	136	12	0.5	1.7	2
10	138.1	139.1	138	25	1	1.1	2
11	140.2	141.2	140	27	1	1.4	2
12	142.6	143.5	142	20	0.9	1.3	4
13	144.8	146.7	146	23	1.9	0.4	2
14	147.1	148	148	26	0.9	1.4	2
15	149.4	150.1	150	31	0.7	1.4	2
16	151.5	152.2	152	10	0.7	1.3	2
17	153.5	154.5	154	14	1	1.2	2
18	155.7	156.1	156	12	0.4	1.4	2
19	157.5	158.1	158	8	0.6	1.8	2
20	159.9	160.3	160	6	0.4	1.5	2
21	161.8	162.3	162	4	0.5	1.5	2
22	163.8	164.2	164	8	0.4	1.7	2
23	165.9	166.2	166	5	0.3	1.5	2
24	167.7	168.3	168	5	0.6	1.4	2
25	169.7	170.5	170	6	0.8	1.4	2
26	171.9	172.2	172	3	0.3	2	2
27	174.2	174.5	174	4	0.3	1.3	2
28	175.8	176.6	176	2	0.8	1.4	2
29	178	178.6	178	9	0.6	1.3	2
30	179.9	180.3	180	2	0.4	1.6	2
31	181.9	182.7	182	12	0.8	1.6	2
32	184.3	184.6	184	6	0.3	1.4	2
33	186	186.7	186	13	0.7	1.3	2
34	188	188.7	188	3	0.7	0.9	2
35	189.6	191	190	5	1.4	1.1	2
36	192.1	192.6	192	3	0.5	1.5	2
37	194.1	194.7	194	7	0.6	1.3	2
38	196	196.6	196	6	0.6	1.5	2
39	198.1	198.9	198	4	0.8	1.2	2
40	200.1	200.2	200	2	0.1	2.3	2
41	202.5	202.8	202	2	0.3	4	4
42	206.8	207.2	206	4	0.4	3.5	6
43	210.7	211.6	212	7	0.9	3.2	4
44	214.8	215.1	216	2	0.3	2.6	2
45	217.7	217.8	218	2	0.1	3.9	4

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
46	221.7	222.8	222	4	1.1	0.7	2
47	223.5	224.1	224	2	0.6	1.4	2
48	225.5	226.1	226	3	0.6	6.6	6
49	232.7	233.1	232	2	0.4	3.8	4
50	236.9	237.1	236	2	0.2	1.4	2
51	238.5	239.4	238	4	0.9	1	2
52	240.4	240.5	240	2	0.1	6.2	8
53	246.7	248.1	248	5	1.4	1.4	2
54	249.5	250.8	250	4	1.3	0.5	2
55	251.3	252.9	252	10	1.6	2.5	4
56	255.4	255.9	256	3	0.5	1.7	2
57	257.6	257.8	258	4	0.2	0.8	2
58	258.6	260.8	260	8	2.2	2	2
59	262.8	263	262	2	0.2	1.3	2
60	264.3	264.3	264	2	0	6.2	6
61	270.5	271.2	270	2	0.7	7.4	8
62	278.6	278.7	278	2	0.1	19	20
63	297.7	298	298	4	0.3	5.7	6
64	303.7	303.9	304	2	0.2		

Table F6i-x. viii. Iso_8.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	256.3	257.5	256	29	1.2	1.2	2
2	258.7	259	258	2	0.3	1.4	2
3	260.4	261.6	260	5	1.2	0.8	2
4	262.4	262.6	262	6	0.2	0.4	2
5	263	265.1	264	49	2.1	0.6	2
6	265.7	267	266	25	1.3	0.6	2
7	267.6	268.1	268	16	0.5	0.5	2
8	268.6	270.2	270	14	1.6	1.3	2
9	271.5	271.8	272	18	0.3	1.5	2
10	273.3	273.8	274	5	0.5	1.4	2
11	275.2	275.8	276	14	0.6	1.3	2
12	277.1	279.6	278	20	2.5	0.4	2
13	280	281.2	280	15	1.2	1.4	2
14	282.6	283.2	282	16	0.6	1.5	2
15	284.7	285.1	284	14	0.4	1.4	2
16	286.5	287.1	286	14	0.6	1.2	2
17	288.3	289.2	288	11	0.9	0.9	2
18	290.1	291.3	290	7	1.2	0.8	2
19	292.1	292.4	292	13	0.3	1.5	2
20	293.9	294.8	294	8	0.9	1.1	2
21	295.9	296.6	296	15	0.7	1	2
22	297.6	299	298	16	1.4	0.4	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
23	299.4	300	300	13	0.6	1.5	2
24	301.5	301.9	302	11	0.4	0.9	2
25	302.8	303.9	304	14	1.1	1.8	2
26	305.7	305.9	306	7	0.2	1.6	2
27	307.5	308.2	308	15	0.7	1.5	2
28	309.7	309.9	310	8	0.2	1.1	2
29	311	312.1	312	15	1.1	1.6	2
30	313.7	315	314	10	1.3	0.7	2
31	315.7	316.8	316	18	1.1	0.8	2
32	317.6	318.3	318	14	0.7	1.3	2
33	319.6	320.3	320	12	0.7	1.5	2
34	321.8	322.3	322	9	0.5	1.7	2
35	324	324.1	324	4	0.1	1.8	2
36	325.9	326.3	326	3	0.4	1.5	2
37	327.8	328.2	328	5	0.4	1.8	2
38	330	330.3	330	4	0.3	1.7	2
39	332	332.3	332	2	0.3	2	2
40	334.3	334.3	334	2	0	5.7	6
41	340	340.4	340	3	0.4	3.4	4
42	343.8	344.3	344	4	0.5	1.8	2
43	346.1	346.4	346	2	0.3	5.4	6
44	351.8	352	352	2	0.2	2.2	2
45	354.2	354.7	354	3	0.5	8.6	10
46	363.3	363.3	364	2	0	8	8
47	371.3	372	372	3	0.7	3	4
48	375	376.6	376	4	1.6	2.8	4
49	379.4	379.5	380	2	0.1	4.1	4
50	383.6	383.7	384	2	0.1	6.4	6
51	390.1	392	390	2	1.9	21.5	24
52	413.5	414.5	414	3	1	16.5	18
53	431	431.4	432	2	0.4	14.8	14
54	446.2	446.8	446	2	0.6		

Table F6i-x. ix. Iso_9.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	250.1	251.4	250	45	1.3	2.6	4
2	254	254.9	254	3	0.9	1.4	2
3	256.3	256.9	256	2	0.6	6.8	8
4	263.7	264	264	3	0.3	1.6	2
5	265.6	266.3	266	3	0.7	1	2
6	267.3	268.5	268	14	1.2	0.8	2
7	269.3	269.9	270	14	0.6	1.4	2
8	271.3	272.1	272	10	0.8	1.1	2
9	273.2	274.1	274	19	0.9	0.9	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
10	275	276	276	8	1	0.9	2
11	276.9	277.5	278	5	0.6	1.6	2
12	279.1	279.9	280	5	0.8	1.1	2
13	281	281.2	282	2	0.2	1.6	2
14	282.8	283.4	284	5	0.6	1.7	2
15	285.1	285.2	286	3	0.1	1.6	2
16	286.8	287.4	288	6	0.6	1.6	2
17	289	289.3	290	4	0.3	1.3	2
18	290.6	291.5	292	3	0.9	1.3	2
19	292.8	293.3	294	4	0.5	3	2
20	296.3	296.5	296	2	0.2	4.2	4
21	300.7	300.8	300	2	0.1	0.7	2
22	301.5	302.5	302	5	1	2.9	4
23	305.4	306.7	306	7	1.3	1.9	2
24	308.6	308.7	308	4	0.1	2	2
25	310.7	311.1	310	2	0.4	1.5	2
26	312.6	313.5	312	16	0.9	1.2	2
27	314.7	315.6	314	7	0.9	0.2	2
28	315.8	317.8	316	17	2	0.7	2
29	318.5	319.6	318	9	1.1	1.5	4
30	321.1	321.7	322	2	0.6	1.9	2
31	323.6	323.9	324	2	0.3	1.5	2
32	325.4	325.8	326	3	0.4	1.4	2
33	327.2	327.5	328	3	0.3	1.3	2
34	328.8	330.1	330	13	1.3	1	2
35	331.1	332.4	332	7	1.3	0.8	2
36	333.2	335.1	334	13	1.9	0.4	2
37	335.5	336.5	336	13	1	0.5	2
38	337	338.1	338	7	1.1	1.4	2
39	339.5	341.1	340	9	1.6	1.6	2
40	342.7	343.1	342	2	0.4	0.5	2
41	343.6	344.1	344	8	0.5	0.5	2
42	344.6	346.6	346	14	2	0.8	2
43	347.4	350.7	348	23	3.3	0.4	4
44	351.1	351.6	352	3	0.5	2.9	2
45	354.5	355	354	4	0.5	1	2
46	356	356.6	356	3	0.6	0.8	2
47	357.4	359	358	11	1.6	0.3	2
48	359.3	361.5	360	12	2.2	1.4	2
49	362.9	363.7	362	4	0.8	0.5	2
50	364.2	364.4	364	2	0.2	2.6	2
51	367	367.7	366	5	0.7	2.4	4
52	370.1	371.3	370	7	1.2	1.1	2
53	372.4	373	372	2	0.6	0.9	2
54	373.9	374.9	374	7	1	1.1	2
55	376	376.6	376	2	0.6	0.4	2
56	377	378.9	378	5	1.9	0.6	2
57	379.5	380.9	380	5	1.4	1.6	4

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
58	382.5	384.5	384	3	2	1.4	2
59	385.9	387.2	386	4	1.3	2.3	4
60	389.5	392.1	390	3	2.6	1	4
61	393.1	396.4	394	4	3.3	0.7	4
62	397.1	398.6	398	3	1.5	1.7	2
63	400.3	401.8	400	5	1.5	4.9	6
64	406.7	407	406	2	0.3	1.3	4
65	408.3	410.5	410	6	2.2	2.3	4
66	412.8	413	414	2	0.2	2.3	2
67	415.3	418.2	416	2	2.9	0.7	2
68	418.9	419.5	418	4	0.6	1.8	4
69	421.3	423.4	422	3	2.1	8.3	10
70	431.7	431.9	432	2	0.2	5.5	6
71	437.4	439.2	438	3	1.8	14.2	16
72	453.4	453.8	454	4	0.4		

Table F6i-x. x. Iso_10.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	188.3	188.6	188	29	0.3	1.7	2
2	190.3	190.5	190	2	0.2	1.9	2
3	192.4	192.4	192	2	0	1.1	2
4	193.5	194.2	194	2	0.7	2	2
5	196.2	196.3	196	3	0.1	1.8	2
6	198.1	198.2	198	2	0.1	6.3	6
7	204.5	204.6	204	2	0.1	3.4	4
8	208	208.7	208	5	0.7	1.2	2
9	209.9	211.2	210	17	1.3	1.3	2
10	212.5	213.7	212	39	1.2	0.9	2
11	214.6	216	214	63	1.4	0.7	2
12	216.7	217.9	216	65	1.2	0.9	2
13	218.8	219.8	218	65	1	1.1	2
14	220.9	222.3	220	38	1.4	0.8	4
15	223.1	223.7	224	17	0.6	1.3	2
16	225	225.6	226	12	0.6	1.4	2
17	227	227.7	228	14	0.7	0.3	2
18	228	229.8	230	10	1.8	1.4	2
19	231.2	233.7	232	10	2.5	1.7	4
20	235.4	235.9	236	5	0.5	3.6	4
21	239.5	240.2	240	4	0.7	1.6	2
22	241.8	242.3	242	2	0.5	1.6	2
23	243.9	244.3	244	2	0.4	5.7	6
24	250	250.4	250	2	0.4	1.9	2
25	252.3	252.3	252	2	0	0.9	2
26	253.2	253.3	254	2	0.1	2.5	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
27	255.8	255.9	256	2	0.1	3.7	4
28	259.6	259.8	260	2	0.2	1.4	2
29	261.2	261.7	262	3	0.5	3.3	4
30	265	265.5	266	3	0.5	1.1	2
31	266.6	267.2	268	6	0.6	1.8	2
32	269	269.2	270	2	0.2	1.3	2
33	270.5	271	272	2	0.5	3.2	2
34	274.2	274.7	274	2	0.5	1.8	2
35	276.5	276.6	276	2	0.1	11.2	12
36	287.8	287.8	288	2	0	3.2	4
37	291	291.5	292	2	0.5	76.3	76
38	367.8	368.3	368	2	0.5		

Baetis rhodani – Genotyping and Binning

Table F7. Showing the final genotyping data for *B.rhodani* at 6 locis (B_1 to B_5, B_7). *Site* identifies the 10 different sites in this data set; *N* = each individuals position in the dataset (same order used in all analyses); *Name* is the name given to each sample and *E* refers to the extraction method used for each sample (see note at end of table). For each loci *R* = number of times each sample was repeated successfully, *F* = Number of failed samples, and *Allele* are the two alleles found in this sample after binning.

Site	N	Name	E	B_1				B_2				B_3				B_4				B_5				B_7			
				R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
112	1	112B1	1	1		128	130	1		226	232	1		189	205	1		186	186	1		154	154	1		242	278
	2	112B11	1	1		128	132	1		228	234	1		189	191	1		188	188	1		148	148	1		242	248
	3	112B12	1	1		128	156	1		236	238	1		191	199	1		186	186	1		146	154	2		244	248
	4	112B13	1	1		128	132	1		228	232	1		195	195	1		186	188	1		152	152	2		242	250
	5	112B15	1	1		144	156	1		224	228	1		167	199	2		180	202	1		142	142	2		242	274
	6	112B16	1	1		128	128	1		228	238	1		197	197	1		186	186	1		148	148	1		248	248
	7	112B17	1	1		128	132	1		228	230	1		189	197	1		190	190	1		142	142	1		230	248
	8	112B18	1	1		128	140	2		228	240	1		191	199	1		186	188	1		142	142	1		278	286
	9	112B19	1	1		132	134	1		228	230	1		189	203	1		188	188	1		148	160	1		230	242
	10	112B2	1	1		128	128	1		224	226	1		195	195	1		186	186	1		148	152	1		246	246
	11	112B20	1	1		134	140	1		228	238	1		191	199	1		186	186	1		142	142	1		246	248
	12	112B21	1	1		128	156	1		226	234	1		189	191	1		188	188	3		144	152	2		242	244
	13	112B22	1	1		128	128	1		230	232	1		193	197	1		186	190	3		270	270	2		244	246
	14	112B3	1	1		132	132	1		228	228	1		195	195	1		186	186	1		144	144	1		242	242
	15	112B4	1	1		132	132	1		230	230	1		187	189	1		186	188	1		146	146	1		242	268
	16	112B5	1	1		128	134	1		228	252	1		197	197	1		186	188	1		142	148	2		230	262
	17	112B7	1	1		128	146	1		232	236	1		193	201	1		186	186	1		142	148	1		242	242
	18	112B8	1	1		128	132	1		228	232	1		193	195	1		180	188	1		146	146	1		246	246
	19	112B9	1	1		132	132	1		228	234	1		189	197	2		204	204	1		152	152	1		242	248
102	20	102B1	1	1		128	134	1		232	238	1		197	217	1		190	190	1		150	152	1		248	278
	21	102B10	1	1		128	128	1		226	248	1		195	199	1		186	186	1		142	152	2		246	280

Site	N	Name	E	B_1				B_2				B_3				B_4				B_5				B_7			
				R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
	22	102B11	1	3		132	200	1		224	238	1		201	201	1		186	204	1		148	152	1		242	248
	23	102B12	1	1		128	130	1		224	228	1		199	199	1		186	188	1		142	144	1		242	242
	24	102B13	1	1		132	134	1		228	232	1		193	195	1		188	188	1		142	166	1		242	246
	25	102B14	1	1		128	132	1		222	234	1		189	199	2		186	192	1		182	182	1		276	276
	26	102B15	1	1		128	132	1		232	238	1		189	191	2		204	204	1		148	148	1		242	242
	27	102B16	1	1		134	134	1		228	232	1		197	201	2		186	186	1		142	142	2		278	284
	28	102B17	1	1		128	160	1		224	228	1		199	215	2		182	186	1		142	152	1		242	246
	29	102B19	1	1		132	156	1		234	240	1		195	199	2	1	184	184		3	-9	-9	1		242	250
	30	102B2	1	2		134	134	2		232	280	1		193	195	1		186	186	1		152	152	2		242	266
	31	102B20	1	1		132	132	1		222	228	1		189	189	2		186	186	1		142	164	1		242	244
	32	102B21	1	1		130	132	1		228	254	1		185	189	1		186	186	1		148	148	1		248	248
	33	102B22	1	1		132	136	1		224	228	1		195	195	1		188	188	1		148	154	1		242	246
	34	102B3	1	2		128	170	1		228	238	1		185	197	1		186	188	1		148	170	1		242	242
	35	102B4	1	1		128	132	1		238	238	1		167	167	1		182	186	1		142	150	1		242	244
	36	102B5	1	1		130	132	1		228	238	1		193	205	1		180	182	1		168	168	2		250	302
	37	102B6	1	1		128	128	1		228	254	1		167	205	1		186	186	1		142	148	1		242	242
	38	102B7	1	1		128	156	1		228	230	1		189	197	1		180	186	1		148	152	1		242	246
	39	102B8	1	1		128	132	1		228	248	1		167	197	1		186	186	1		142	144	1		242	242
	40	102B9	1	1		128	132	2		240	240	1		167	197	1		186	186	1		142	142	1		242	248
97	41	97B10	1	1		128	132	1		222	232	1		215	231	1		186	204	1		142	150	1		242	242
	42	97B11	1	1		128	132	1		246	252	1		185	199	2		172	186	1		152	152	1		242	242
	43	97B12	1	1		128	132	1		228	230	2		197	205	1		180	186	1		142	154	1		242	242
	44	97B13	1	1		128	132	1		230	230	1		189	231	2	1	186	186	1		142	174	1		242	242
	45	97B14	1	1		132	158	1		228	230	1		191	211	2		186	186	1		142	148	2		244	278
	46	97B15	1	1		132	132	1		224	236	1		195	195	1		188	188	1		144	164	1		246	248
	47	97B16	1	1		128	128	1		228	254	1		191	195	1		186	186	1		148	148	1		242	248
	48	97B17	1	1	1	128	132	1	1	226	232	2		189	209	1	1	188	188	1		148	148	2		242	242
	49	97B18	1	1		134	134	1		232	246	1		203	203	1		180	186	1		142	152	1		242	286
	50	97B19	1	1		134	134	1		228	232	1		203	203	1		186	186	1		142	152	1		242	286
	51	97B20	1	1		132	168	1		226	236	1		189	191	1		186	204	1		140	152	1		242	244
	52	97B21	1	2		128	128	2		234	238	3		195	195	2		188	188	2		152	152	2		242	242

Site	N	Name	E	B_1				B_2				B_3				B_4				B_5				B_7			
				R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
	53	97B3	1	2		132	134	1		224	252	1		201	211	1		186	188	1		152	152	1		242	242
	54	97B4	1	2		132	156	2		228	234	2		191	193	2		186	186	2		142	152	2		242	242
	55	97B5	1	1		156	168	1		232	234	1		189	191	1		188	188	1		152	174	1		242	250
	56	97B6	1	2		128	156	2		228	228	2		189	199	2		186	186	2		142	142	2		242	244
	57	97B7	1	1		132	136	1		224	240	1		193	207	1		184	188	1		142	144	1		242	276
	58	97B8	1	1		134	134	1		228	232	1		189	203	1		180	186	1		142	152	1		242	246
	59	97B9	1	1		130	132	1		228	230	2		171	191	1		186	186	1		142	142	1		242	252
96	60	96B1	1	2		128	160	1		222	228	2	1	189	240	1		186	186	1		148	150	2		242	248
	61	96B10	1	1		128	128	1		224	234	1		191	191	1		180	186	2		152	262	2		244	288
	62	96B11	1	2		132	166	1		232	234	1		197	205	2		186	188	1		142	148	1		242	244
	63	96B12	1	1		128	128	1		222	238	1		195	211	1		184	184	1		140	152	1		242	242
	64	96B13	1	1		128	132	1		230	240	1		193	195	1		182	182	2		152	152	1		242	244
	65	96B14	1	1		132	134	1		228	236	1		189	191	1		184	184	1		142	142	1		244	250
	66	96B16	1	1		128	128	1		230	234	1		187	191	1		180	186	1		148	152	2		242	318
	67	96B17	1	1		132	132	1		228	230	1		191	201	1		186	186	2		142	144	1		244	244
	68	96B18	1	1		128	134	1		228	238	1		193	197	1		186	186	2		142	196	1		242	244
	69	96B19	1	1		132	132	1		230	234	1		193	197	1		188	190	1		142	142	1		242	282
	70	96B2	1	1		128	158	1		224	226	1		187	201	2		186	188	1		142	142	1		250	274
	71	96B21	1	1		130	134	1		228	230	2	1	193	215	1		186	186	2		148	202	1		242	242
	72	96B22	1	1		128	136	1		226	230	1		195	197	1		186	186	2		144	146	1		242	246
	73	96B4	1	1		128	132	1		226	232	1		167	189	1		186	186	1		144	144	1		242	242
	74	96B5	1	1		132	132	1		228	230	1		197	217	1		180	188	1		150	150	1		248	282
	75	96B6	1	1		128	134	1		224	224	1		177	191	2		186	188	1		142	142	1		242	244
	76	96B7	1	1		128	156	1		228	228	1		199	211	1		186	186	1		152	152	1		242	242
	77	96B8	1	1		128	130	1		226	228	1		189	231	1		186	186	2		146	146	1		250	268
9	78	09B1	4	1		132	134	1		234	248	1	1	193	207	1		186	186	1		152	166	1		242	248
	79	09B11	4	1		130	130	2		224	238	1	1	189	189	1		186	186	2		256	256	1		242	242
	80	09B13	4	1		128	132	1		236	238	1	1	189	189	1		188	188	1		142	142	1		242	250
	81	09B14	4	1		128	144	1		230	240	1	1	197	197	1		186	188	1		142	148	1		240	242
	82	09B15	4	1		128	128	1		228	234	1	1	189	201	1		188	188	1		144	144	1		242	244
	83	09B2	4	1		128	134	1		224	228	1	1	167	185	2		186	186	2		142	142	1	1	250	250

Site	N	Name	E	B_1				B_2				B_3				B_4				B_5				B_7			
				R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
	84	09B20	4	1		128	132	2		228	228	1	1	191	199	1		188	188	1		142	150	1		242	242
	85	09B21	4	1		128	128	2		222	228	1	1	187	199	1		180	186	1		142	142	1		242	246
	86	09B22	4	1		132	138	1		222	230	1	1	191	217	1		188	194	1		142	152	1		242	248
	87	09B4	4	1		130	144	2		230	240	1	1	191	193	1		184	188	1		142	182	1		242	242
	88	09B5	4	1		128	128	2		230	230	1	1	191	197	1		186	188	1		144	150	1		242	250
106	89	106B1	1	2		128	130	2		228	234	2		189	195	2		186	186	2		144	206		3	-9	-9
	90	106B10	1	1		128	130	1		228	228	1		191	201	1		186	186	2		150	192	1		242	242
	91	106B11	1	1		128	130	3		228	228	1		195	195	1		186	186	2		142	142	1		242	250
	92	106B12	1	1		130	130	1		228	228	1		187	189	1		186	188	1		152	170	1		250	250
	93	106B14	1	1		130	130	1		228	274	1		189	201	1		180	186	1		150	150	2		242	250
	94	106B15	1	1		130	132	2		228	232	1		199	205	2		188	188	1		142	148	1		242	244
	95	106B16	1	1		128	134	1		224	228	1	1	185	189	1		186	186	1		148	152	1		246	246
	96	106B17	1	1		128	130	1		230	246	1		191	191	1		182	188	1		148	148	1		242	242
	97	106B18	1	1	1	128	132	2		228	240	1		191	191	1		186	188	3		314	314	1		244	244
	98	106B19	1	2	1	128	130	1		224	224	1		197	197	1		186	186	2		142	178	1		242	242
	99	106B2	1	2		130	130	2		228	228	2		205	205	2		186	186	2		142	142	2		250	250
	100	106B20	1	1		130	138	1		228	232	1		191	199	2		186	186	1		148	152	1		250	250
	101	106B21	1	1		128	140	1		224	230	1		195	195	1		186	194	2		162	198		3	-9	-9
	102	106B22	1	1		132	132	1		222	222	1		185	197	1		186	186	1		142	154	1		242	246
	103	106B3	1	2		130	132	2		222	224	2		191	199	2		188	188	2		142	150	2		244	244
	104	106B4	1	1		130	132	1		228	228	1		177	201	1		182	188	1		142	152	1		242	242
	105	106B5	1	1		130	132	1		230	238	1		193	195	1		190	190	1		144	152	1		242	242
	106	106B7	1	2		128	130	1		232	234	2		201	201	1		182	186	1		148	148	1		268	268
	107	106B8	1	2		130	132	1		226	232	2		191	195	1		186	186	1		142	142	1		242	244
	108	106B9	1	2		128	132	1		228	228	2		195	199	1		186	186	1		142	148	1		242	242
113	109	113B1	1	2		128	130	2		228	232	1	1	189	193	2		186	188	2		186	186	2		246	250
	110	113B10	1	1		130	132	1		228	264	1		189	199	1		186	186	2		142	250	1		246	250
	111	113B11	1	2		130	168	2		228	262	1		195	201	1		182	186	1		146	152	1		242	242
	112	113B12	1	1		128	132	1		234	234	1		189	193	1		180	186	1		150	150	1		240	242
	113	113B13	1	1		128	130	1		224	240	1		189	199	1		180	186	1		144	198	1		242	242
	114	113B14	1	1		132	156	1		228	230	1		167	191	1		182	188	2		142	206	1		242	246

Site	N	Name	E	B_1				B_2				B_3				B_4				B_5				B_7			
				R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
	115	113B15	1	1		130	152	2		222	228	1		193	221	1		186	186	2		144	240	1		242	250
	116	113B16	1	1		130	132	1		222	226	1		191	193	1		186	188	1		142	142	1		242	250
	117	113B18	1	1		130	130	1		226	234	1		191	203	1		182	188	1		154	154	2	1	250	250
	118	113B19	1	1		128	130	1		230	232	1		167	187	1		186	186	1		142	166	1		242	242
	119	113B2	1	2		130	142	1	1	226	228	1	1	191	191	2		186	186	2		142	142	2		246	246
	120	113B21	1	1		128	136	1		228	230	1		193	193	1		186	188	1		148	148	1		242	244
	121	113B22	1	1		128	132	1		228	248	2		187	195	1		186	188	1		142	142	1		246	246
	122	113B3	1	2		134	136	2		228	230	1	1	189	205	2		190	190	2		142	328	2		242	250
	123	113B4	1	2		128	130	2		228	228	1	1	197	197	2		186	186	2		142	200	2		244	250
	124	113B7	1	2		128	130	2		224	230	1	1	193	199	2		186	186	2		148	150	3		244	250
	125	113B8	1	3		128	168	2		228	274	1	1	187	191	2		182	186	3		144	152	2		242	258
	126	113B9	1	2		130	134	1	1	234	246	1	1	191	195	1	1	186	188	1	1	148	152	1	1	242	250
115	127	115B1	1	1		134	134	1		228	238	1		167	189	1		188	188	1		148	148	1		242	248
	128	115B10	1	1		128	128	1		222	252	1		193	193	1		186	188	1		146	152	1		230	242
	129	115B11	1	1		132	146	1		224	238	1		189	211	1		180	184	2		204	204		3	-9	-9
	130	115B12	1	1		128	132	1		228	248	1		191	197	1		186	188	1		142	142	1		242	242
	131	115B13	1	2		128	136	2		230	232	2		191	195	2		186	186	4		146	238	2		242	242
	132	115B14	1	1		128	128	1		222	232	1		195	195	1		186	186	1		146	152	1		242	246
	133	115B15	1	1		132	134	2		224	270	1		187	193	1		186	186	2		182	182	1		242	242
	134	115B17	1	2		128	152	2		228	234	2		193	197	2		186	188	4		144	194	2		242	246
	135	115B18	1	1		128	128	1		224	230	1		175	201	1		182	186	1		142	152	1		242	242
	136	115B19	1	1		132	132	1		226	232	2		195	197	1		180	188	1		142	144	1		242	248
	137	115B2	1	2		128	134	1		228	236	1		201	205	1		186	186	1		144	152	1		242	242
	138	115B21	1	1		128	134	1		228	232	2		167	185	1		186	188	1		152	166	1		242	246
	139	115B22	1	1		132	140	1		224	232	2		191	195	1		186	194	1		164	164	1		244	248
	140	115B3	1	1		128	134	1		224	230	1		189	189	2		184	184	2		246	246	1		242	242
	141	115B4	1	1		128	132	1		224	232	1		195	195	1		188	188	1		144	166	1		242	246
	142	115B5	1	1		128	134	1		228	238	1		191	207	1		186	186	2		142	152	1		246	250
	143	115B6	1	1		128	128	1		228	228	1		167	185	1		188	188	1		148	164	1		242	242
	144	115B7	1	1		134	158	1		228	230	1		191	197	1		186	186	2		152	152	1		244	248
	145	115B9	1	1		128	146	1		228	234	1		167	185	1		186	188	1		150	150	1		242	250

Site	N	Name	E	B_1				B_2				B_3				B_4				B_5				B_7			
				R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
94	146	94B1	1	1		128	132	1		224	232	1		195	197	1		186	186	1		148	164	1		248	248
	147	94B10	1	1		128	130	1		230	244	1		189	197	1		186	186	1		152	152	1		250	250
	148	94B11	1	1		130	130	1		226	244	1		185	191	1		186	186	2		142	144	1		250	250
	149	94B12	1	1		138	138	1		228	236	1		199	199	1		186	186	1		144	150		3	-9	-9
	150	94B15	1	1		130	138	1		230	230	1		193	193	1		188	188	1		168	168	1		250	250
	151	94B16	1	2		128	130	2		228	228	2		201	201	2		186	188	2		148	152	2		250	250
	152	94B17	1	2		128	130	1		222	228	1		191	197	2	1	188	188	2		142	144	2		250	250
	153	94B18	1	1		130	132	2		228	260	1		193	193	1		186	186	1		142	148	1		246	246
	154	94B19	1	1		128	128	1		228	228	1		207	207	1		186	186	1		144	144	1		242	242
	155	94B2	1	1		132	138	1		228	230	1		193	201	1		186	186	1		146	146	1		242	242
	156	94B20	1	1		130	132	1		228	228	1		191	191	1		182	190	1		148	148	1		246	250
	157	94B21	1	1		130	132	1		224	234	1		177	197	1		186	186	2		184	184	1		242	250
	158	94B22	1	1		130	130	1		224	224	1		193	221	1		186	188	1		142	148	1		250	250
	159	94B3	1	1		128	130	1		228	238	1		185	191	2		186	186	1		142	142	1		242	242
	160	94B4	1	1		130	130	1		228	248	1		189	193	1		186	186	2		148	158	1		250	250
	161	94B5	1	1		132	138	1		228	252	1		175	193	1		182	186	1		142	142	1		244	250
	162	94B6	1	1		128	130	1		234	240	1		189	201	1		188	188	1		142	142	1		242	248
	163	94B7	1	1		128	130	1		232	234	1		189	205	1		186	188	1		150	150	1		242	242
	164	94B8	1	1		132	136	1		228	232	1		189	191	1		186	186	1		142	142		3	-9	-9
	165	94B9	1	1		130	132	1		228	238	1		193	199	1		186	188	1		148	148	1		242	242
118	166	118B1	1	1		128	154	1		226	230	1		199	201	1		186	188	1		218	218	1		244	268
	167	118B10	1	1		128	128	1		226	228	1		189	197	1		188	188	1		142	200	1		242	242
	168	118B11	1	2		132	132	1		228	228	1		189	193	1		186	186	1		148	148	2		242	242
	169	118B12	1	1		132	158	1		228	264	1		193	195	1		186	186	3		144	276	1		242	246
	170	118B13	1	1		128	132	1		228	232	1		191	197	1		186	188	2		160	160	1		242	246
	171	118B14	1	1		128	130	1		230	230	1		193	193	1		186	186	1		142	152	1		242	246
	172	118B15	1	1		128	134	1		224	228	1		189	199	1		186	186	1		152	152	1		244	244
	173	118B16	1	1		128	132	1		228	232	1		187	187	1		186	190	1		144	150	1		244	246
	174	118B17	1	2		132	134	1		228	232	1		167	199	1		182	186	2		144	150	1		242	244
	175	118B18	1	1		128	132	1		222	228	1		191	207	1		182	186	1		142	150	1		242	244
	176	118B19	1	1		132	134	1		238	238	1		193	201	1		186	186	1		152	218	1		242	242

Site	N	Name	E	B_1				B_2				B_3				B_4				B_5				B_7			
				R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
	177	118B2	1	1		128	134	1		226	230	1		189	189	1		182	186	1		144	144	1		242	244
	178	118B21	1	1		128	132	1		226	234	1		197	197	1		184	188	1		144	144	2	1	246	268
	179	118B22	1	2		132	132	1		222	238	1		193	193	1		186	186	1		148	148	2		260	312
	180	118B3	1	1		132	132	1		222	236	1		197	197	1		188	190	1		144	152	1		244	244
	181	118B4	1	1		128	132	1		224	248	1		191	193	1		186	186	1		142	152	1		242	244
	182	118B5	1	1		128	128	1		228	230	1		193	197	1		184	184	1		150	204	1		242	244
	183	118B6	1	1		132	132	1		228	234	2		199	199	1		186	186	1		148	148	1		248	252
	184	118B7	1	1		132	136	1		228	228	1		195	199	2		186	186	3	1	152	302	1		242	244
	185	118B8	1	1		128	154	1		226	230	1		201	207	1		186	186	1		142	150	1		244	276
	186	118B9	1	1		128	128	1		224	230	1		185	191	2		180	188	1		142	152	1		248	250

NOTE

Extraction method 1 High Pure PCR Template Preparation Kit for blood and tissue following the manufacturer's instructions (Roche Diagnostics GmbH Mannheim, Germany)

Extraction method 4 Gentra Puregene Core Kit A, DNA purification from tissue (Qiagen, Hilden, Germany)

Table F8. Showing the final genotyping data for *B.rhodani* at 7 locis (Brh-1 to Brh-7). *Site* identifies the 10 different sites in this data set; *N* = each individuals position in the dataset (same order used in all analyses) and *Name* is the name given to each sample. For each loci *R* = number of times each sample was repeated successfully, *F* = Number of failed samples, and *Allele* are the two alleles found in this sample after binning.

Site	N	Name	Brh-1				Brh-2				Brh-3				Brh-4				Brh-5				Brh-6				Brh-7			
			R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
112	1	112B1	2		223	231	1		146	152	1	1	213	213	1	1	183	185	1		210	210	2		111	125	2		162	162
	2	112B11	2	1	227	229	1		154	154	2		213	213	2		185	185	1		208	212	2		109	113	2		162	166
	3	112B12	2		221	227	1		150	150	1	1	213	213	1	1	183	189	1		212	222	2		103	109	1	1	162	184
	4	112B13	1	3	225	225	3		146	150	4		213	213	2	2	185	213	3		212	212	4		111	113	4		160	172
	5	112B15	3	1	221	221	3		146	148	4		225	225	3	1	187	191	3		208	208	3	1	109	113	2	2	176	222

Site	N	Name	Brh-1				Brh-2				Brh-3				Brh-4				Brh-5				Brh-6				Brh-7			
			R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
	6	112B16	2		223	223	1		148	148	2		215	215	1	1	175	189	1		220	220	2		111	111	3		158	182
	7	112B17	3		229	229	1		150	150	2		213	217	1	1	187	191	1		214	214	2		109	109	2		164	174
	8	112B18	2		219	225	1		150	150	2		227	227	1	1	179	191	1		214	214	2		109	111	1	1	174	188
	9	112B19	2		219	223	1		148	148	1	2	213	215	1	1	185	189	1		234	248	2		111	111	2		152	172
	10	112B2	1	1	219	221	1		150	150	2	1	213	213	1	1	187	191	1		208	214	2		113	113	2		156	172
	11	112B20	2		225	225	1		150	150	2		213	213	1	1	183	187	1		214	216	2		105	109	1	3	236	240
	12	112B21	3	1	219	219	3		148	150	3	1	197	213	3	1	179	187	3		208	212	3	1	103	113	2	2	154	212
	13	112B22	2		219	227	2		154	154	2		213	213	2		203	245	1		214	236	2		111	111	2	1	166	196
	14	112B3	2		225	237	1		150	150	2		213	213	1	1	185	187	2		212	264	2		109	115	2		154	154
	15	112B4	2		233	233	1		148	148	2		213	215	2		175	183	1		210	218	2		111	111	2		154	172
	16	112B5	4		221	221	1		148	150	2		213	213	1	1	187	187	1		212	212	2		111	115	2		172	172
	17	112B7	2		221	221	1		146	154	2		233	233	1	1	183	185	1		210	210	2	1	111	111	1	1	184	186
	18	112B8	2		221	231	1		146	150	2		197	197	2		187	191	2		212	212	2		103	113	2		182	186
	19	112B9	3		223	223	1		150	150	2		213	213	1	1	187	209	1		210	212	2		103	163	2		156	176
102	20	102B1	1	1	221	233	1		146	154	2		213	213	1	1	185	191	1		212	216	2		111	111	2	1	196	196
	21	102B10	2		221	221	1		148	148	2		219	219	1	1	185	245	1		214	236	2		103	109	2	1	190	216
	22	102B11	2		221	225	1		148	156		4	-9	-9	2		175	189	1		212	216	2		109	111	3		160	160
	23	102B12	3	2	231	231	4		150	166	3	2	215	215	4	1	187	247	4		216	224	4	1	111	111	4	1	164	174
	24	102B13	3	2	221	221	4		150	150	5		225	237	5		187	187	4		208	230	5		111	111	4	1	162	202
	25	102B14	2	1	221	221		4	-9	-9	2	1	211	211	1		221	221	2		206	206	2	1	109	109	1		212	212
	26	102B15	1	1	221	239		4	-9	-9	2	1	213	215	3		183	187	2		208	214	1	1	111	115	1		214	216
	27	102B16	1		221	229	1		146	154	1		213	213	1		183	183	1		216	216	2		103	115	1		152	170
	28	102B17		3	-9	-9	1		150	150	1	1	211	215	1		179	187	1		212	218	1		103	109	1		154	162
	29	102B19	1	1	221	221	1		150	154	1		231	233	1	1	183	187	1		212	238	1		103	103	1		152	218
	30	102B2	1	1	221	221	1		150	150	1	2	213	213	1	1	179	185	1		212	212	2		103	105	2	1	174	176
	31	102B20	1		223	223	1		146	146	1		195	213	2	1	185	187	1		214	220	1		111	111	1	1	194	218
	32	102B21	1	2	223	223	2	1	150	168	1		213	213	1	1	175	175	1	1	212	226	1		103	111	1		154	190
	33	102B22	1	1	225	225	1		154	154	1		219	219	2	1	185	185	1		208	212	1		109	111	1		178	228
	34	102B3	1	1	225	225	1		148	148	2		233	233	1	1	179	187	1		210	210	2		103	109	1	1	164	164
	35	102B4	1	1	221	221	1		150	150	2		219	219	1	1	175	191	1		212	222	2		111	117	1	1	152	178
	36	102B5	1	1	223	223	1		146	156	2		213	213	1	1	183	185	1		214	222	2		107	111	1	1	166	180
	37	102B6	2		225	231	2		150	150	2	1	213	225	2		185	185	1		210	214	2		109	111	2		162	184
	38	102B7	2		221	239	1		154	154	2		213	213	1	1	185	199	2		212	232	2		109	115	2		166	184
	39	102B8	2		229	229	2		148	182	2		215	215	1	1	187	207	1		212	212	2		111	113	2		170	170

Site	N	Name	Brh-1				Brh-2				Brh-3				Brh-4				Brh-5				Brh-6				Brh-7			
			R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
	40	102B9	1	1	221	225	1		150	150	1	1	215	215	1	1	175	183	1		208	208	2		111	111	1	1	172	180
97	41	97B10	1		221	221	1		154	154	1	2	213	213	1		187	187	1		208	226	1		103	111	1		152	172
	42	97B11	1		221	221	1		150	150	1		215	215	1		183	199	1		212	214	1		103	113	1		156	212
	43	97B12	1		221	221	1		146	150	1		213	213	1		185	189	1		208	210	1		109	109	1		174	208
	44	97B13	1		223	223	1		146	154	1		213	213	1		187	187	1		208	210	1		103	111	1		152	172
	45	97B14	1		219	219	1		146	150	3	1	215	215	1		181	189	1		208	212	1		109	113	1		176	236
	46	97B15	1		221	225	1		154	154	1		215	215	1		183	189	1		210	222	1		109	111	1		170	188
	47	97B16	1		221	221	1		154	154	1	1	213	213	1		183	185	1		210	212	1		111	111	1		162	194
	48	97B17	1		221	221	1		150	150	1		215	215	1		187	191	1		210	220	1		109	111	1		160	170
	49	97B18	1		229	229	1		150	154	1		225	225	1		179	187	1		208	210	1		111	113	1		152	158
	50	97B19		3	-9	-9	1		150	154	1		213	215	1		179	185	1		208	214	1		111	111	1		158	160
	51	97B20	1		219	219	1		148	148	1		213	213	2		179	189	1		204	216	1		111	113	2		152	234
	52	97B21	1		219	225	1		154	154	2		197	255	1		185	197	1		208	216	1		111	111	1		176	188
	53	97B3	1		221	221	1		150	154	1		213	213	1		183	187	1		208	218	1		111	113	2		216	232
	54	97B4	1		219	219	1		150	150	1		227	227	1		183	185	1		212	218	1		109	125	1		164	180
	55	97B5		3	-9	-9	2	2	154	154	1		213	213	1		179	183	1		208	214	1		109	113	1		152	206
	56	97B6	1		221	221	1		150	150	1		213	213	1		183	187	2		214	240	2	1	109	109	1		158	212
	57	97B7	1		221	221	1		150	150	1		197	225	1		179	185	1		210	224	1		115	115	1		162	210
	58	97B8	1		223	223	1		150	154	1		225	225	1		179	187	1		208	214	1		111	113	1		152	158
	59	97B9		3	-9	-9	1		146	146	1		225	225	1		183	183	1		210	210	1		113	113	2		186	246
96	60	96B1	2		221	221	3		152	152	2		213	233	2		179	185	1		216	222	2		111	143	2	1	166	194
	61	96B10	2		221	221	1		148	148	2		197	225	1	1	179	187	1		210	210	2		109	109	2	2	226	300
	62	96B11	1		239	239	1		148	150	1		197	197	1		179	185	2		214	302	1		111	115	1		182	186
	63	96B12	2		229	229	1		146	148	1	2	213	213	2		183	187	1		208	212	2		111	111	2		162	170
	64	96B13	1	1	219	223	1		146	148	2		213	213	1	1	179	187	1		210	242	2		103	109	3		196	196
	65	96B14	1		219	227		3	-9	-9	2		197	227	1		179	195	1		210	232	2		103	131	1		158	170
	66	96B16	2		221	229	1		148	148	1	2	213	213	2		185	185	1		208	210	2	1	111	111	2		164	174
	67	96B17	2		219	219	1		148	154	2	2	213	213	1	1	191	199	1		218	218	1	1	105	111	3		166	166
	68	96B18	2		221	221	1		148	148	2		213	225	2	1	177	183	1		208	208	2		109	123	2		162	178
	69	96B19	2		223	223	1		154	154	2		213	213	2		179	185	1		212	212	2		103	109	1	2	188	208
	70	96B2	1	1	219	219	1		150	150	2	1	215	215	2	1	183	231	2		210	214	2		107	107	2		164	164
	71	96B21	2		221	221	1		146	148	2		225	225	1	1	183	191	1		210	238	2	1	127	127	2		186	216
	72	96B22	3		219	219	1		148	150	2		215	215	1	1	179	181	1		212	222	2		109	145	2		156	206
	73	96B4	3		221	221	1		146	146	3		213	213	1	1	221	221	1		210	212	2	1	113	123	1	2	214	214

Site	N	Name	Brh-1				Brh-2				Brh-3				Brh-4				Brh-5				Brh-6				Brh-7			
			R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
	74	96B5	2	2	219	247	1		150	154	3		215	215	3	1	183	185	1		212	214	1	1	111	113	1	1	216	216
	75	96B6	1	1	219	219	2		148	148	2		227	227	2	1	179	205	2		208	208	2		111	127	2		182	236
	76	96B7	1	1	221	221	1		154	154	2		213	213	2		187	187	2		210	236	2		109	109	1	1	180	180
	77	96B8	2		223	231	1		150	156	2		213	213	2		185	193	1		212	214	2		111	115	1	1	188	210
9	78	09B1	1		221	221	1		148	150	1		213	215	1		185	201	1		212	212		3	-9	-9	2	1	190	196
	79	09B11	1		225	231	2		150	150	2		227	227	1		181	187	1		210	216	1		105	123	2	1	206	216
	80	09B13	1		231	231	1		154	156	2	1	211	213	1		185	185	1		212	242	1		111	137	2	1	172	238
	81	09B14	1		221	221	1		150	152	1		211	215	1		175	197	1		212	248	1		103	149	2	1	204	210
	82	09B15		3	-9	-9	1		148	148	1		213	213	1		179	179	1		212	212		3	-9	-9	2	1	170	346
	83	09B2	2		217	217	2	1	154	156	2	1	231	231		3	-9	-9	1		206	216		3	-9	-9	2	1	200	212
	84	09B20	1		221	229	2		142	148	2		197	213	1		183	187	1		210	210	1		103	111	2	1	178	228
	85	09B21	1		221	221	1		148	154	2		213	213	1		185	185	1		212	220	1		111	113	2	1	170	176
	86	09B22	2		231	231	2		148	148	1		195	195	1		185	199	1		210	214	1		109	109	2	1	244	244
	87	09B4	2		233	233	1		150	152	1		213	225	1		183	187	2		212	228	1		115	149	2	1	172	204
	88	09B5	1		221	229	2		150	150	3		213	213	1		179	179	1		212	216	1		111	113	2	1	152	184
106	89	106B1	1	1	221	223	1		150	150	2		195	195	1	1	183	185	1		208	212	2		107	127	1	1	162	194
	90	106B10	1	3	223	231	3		146	150	4		211	211	3	1	185	195	3		216	216	4		109	111	4		160	160
	91	106B11	2		221	221	1		150	154	2		215	215	1	1	179	185	1		210	210	2		111	111	2		174	188
	92	106B12	2		219	219	1		148	150	2		197	213	2		183	183	1		210	214	2		113	115	2	3	262	342
	93	106B14	2		229	229	1		146	150	1	1	211	213	1	1	181	185	2		210	220	2		127	153	2		156	156
	94	106B15	1	1	221	239	1		146	152	2		213	213	2		183	183	1		208	210	2		113	127	1	1	172	176
	95	106B16	1	1	221	229	1		150	150	2		213	227	1	1	183	185	2		212	214	2		105	109	2		164	164
	96	106B17	2	2	221	221	3		166	166	4		213	213	4		179	183	3		214	216	4		105	115	3	1	164	188
	97	106B18	3	1	221	237	4		150	150	4		211	211	4		183	183	3		212	244	4		111	113	3	1	152	190
	98	106B19	2		221	233	2		148	186	1	2	213	213	2		181	185	1		206	210	2		105	113	2		150	172
	99	106B2	2		225	225	1		150	154	2		213	213	1	1	185	187	1		214	214	2		115	119	1	1	162	174
	100	106B20	2	1	219	219	1		154	154	2		195	195	1	1	183	183	1		214	214	2		111	137	2		174	174
	101	106B21	4		221	221	3		146	150	4		219	219	3	1	183	185	3		216	216	4		115	115	4		188	188
	102	106B22	2		221	221	1		148	150	2		195	215	2		187	187	1		210	210	2		111	111	1	2	172	198
	103	106B3	2		221	233	1		154	154	2	1	213	213	2		179	189	1		208	210	2		107	111	2		166	174
	104	106B4	2		229	237	1		148	148	2		197	197	2		183	183	1		208	232	2		111	111	1	2	170	210
	105	106B5	2		219	239	1		148	154	2		213	213	2		175	179	1		210	214	2		103	115	2		162	162
	106	106B7	1	1	219	219	1		148	154	2		225	225	1	1	185	187	1		210	218	2		111	111	2		154	156
	107	106B8	2		221	221		4	-9	-9	1	2	213	213	2		185	187	1		208	220	2		105	111	2	2	238	266

Site	N	Name	Brh-1				Brh-2				Brh-3				Brh-4				Brh-5				Brh-6				Brh-7			
			R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
	108	106B9	1	1	225	225	1		148	154	2		213	213	1	1	179	187	2		212	214	2		103	103	2		160	170
113	109	113B1		4	-9	-9	1	2	152	152	2		211	213	1	1	203	203	2		210	210	2		105	111	2		170	188
	110	113B10	4	1	223	223	4		148	150	4	1	213	215	4	1	185	187	4		210	222	4	1	111	149	4	1	172	220
	111	113B11	1	1	219	219	1		150	150	2		199	213	1	1	183	185	2	1	214	274	2		115	115	2	2	196	226
	112	113B12	1	1	221	221	1		154	154	2		213	213	2	1	179	215	2		210	214	2		111	119	2	1	180	182
	113	113B13	1	1	219	221	1		154	164	1	1	213	213	1	1	179	185	1		208	210	2		113	113	2	1	210	210
	114	113B14	3		219	219	2		148	154	2		213	213	2		179	183	1		212	216	2		111	111	1	1	180	208
	115	113B15	2		225	225	1		148	150	1	1	213	213	2		183	183	1		208	214	2		115	143	2		156	186
	116	113B16	2	2	219	219	3		150	154	4		213	213	3	1	179	185	3		212	212	4		111	111	4		188	194
	117	113B18	2		221	221	1		154	156	2		213	213	1	1	183	185	2		212	272	2		111	111	1	1	186	198
	118	113B19	2		219	219	1		150	154	2		213	221	2	1	187	201	1		210	210	2		105	111	2		174	174
	119	113B2		4	-9	-9	1		154	154	2		225	225	1	1	185	185	2		206	212	2		111	125	2		158	158
	120	113B21	2		243	243	1		148	150	2		213	213	2		183	185	1		208	208	2		105	113	1	1	166	202
	121	113B22	2		221	221	1		146	154	2		213	213	1	1	183	183	2		234	234	2		111	111	1	1	176	184
	122	113B3	1	1	221	233	1		148	150	2		215	215	1	1	183	187	1		206	210	2		103	103	1	1	170	180
	123	113B4	2		233	233	1		150	150	2		213	213	2	1	185	187	1		216	218	2		111	115	2	1	206	266
	124	113B7	3	2	221	221	4		154	154	4	1	213	213	5	1	183	185	4		210	212	4	1	109	109	3	2	188	208
	125	113B8	2		219	219	1		150	150	2		213	213	1	1	185	185	1		208	210	2		113	113	2		174	174
	126	113B9	2		219	219	1		146	146	2		213	213	2		175	185	1		208	220	2		127	127	2	1	172	212
115	127	115B1	3		223	223	1		154	154	2		197	213	2	1	183	235	1		210	232	2		109	113	1	2	200	200
	128	115B10	1		231	231	1		150	182	1		197	211	1		185	185	1		210	216	1		111	113	1		174	196
	129	115B11	1		223	223	1		150	152	1		213	213	1		183	185	1		210	222	1		103	105	1		160	166
	130	115B12	1		219	219	1		150	150	2		215	215	1		181	183	1		210	230	1		111	111	1		156	166
	131	115B13	1		221	221	1		150	150	2		213	213	1		183	187	1		212	222	1		113	113	1		164	216
	132	115B14	1		219	225	1		148	156	1		225	225	1		185	185	1		214	216	1		113	113	1		202	218
	133	115B15	1		217	237	1		148	154	1		213	213	1		179	187	1		208	214	1		111	125	1		154	204
	134	115B17	1		219	223	2		144	144	1		213	213	1		185	191	1		210	212	1		111	111	1		160	190
	135	115B18	1		243	243	1		150	150	1		197	197	1		183	207	1		214	214	1		111	113	1		166	194
	136	115B19	1		221	221	2		148	202	1		229	229	1		197	199	1		216	234	1		105	111	1		154	170
	137	115B2	2		221	225		4	-9	-9	1	1	213	213	2		185	185	1		210	226	2		113	113	1	1	164	186
	138	115B21	1		219	219	1		154	168	1		227	227	1		175	181	1		212	212	1		103	105	1		162	186
	139	115B22	1		233	233	1		150	150	1		225	225	1		179	191	1		210	212	2		111	129	1		154	156
	140	115B3	1		223	229	1		150	150	2		213	213	2		185	243		3	-9	-9	1		103	111	2	1	178	280
	141	115B4	1		219	219	1		150	154	1		215	215	1		179	185		3	-9	-9	2		111	161	1		176	210

Site	N	Name	Brh-1				Brh-2				Brh-3				Brh-4				Brh-5				Brh-6				Brh-7			
			R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
	142	115B5	1		223	223	1		148	148	1		195	195	1		183	189	1		208	214	2		103	143	1		212	222
	143	115B6	1		219	219	1		150	150	1		213	213	1		185	185	1		210	214	1		113	145	1		170	182
	144	115B7	1		221	221	1		150	150	1	1	213	213	2		179	201	1		210	212	1		111	111	1		160	170
	145	115B9	1		219	221	1		150	168	1		213	215	1		179	187	1		210	212	1		119	119	1		182	200
94	146	94B1	1	1	221	223	3		154	154	2		195	195	1	1	183	189	1		212	214	2	1	115	157	2	1	184	194
	147	94B10	1		225	237	1		150	150	1		213	213	1		185	185	1		212	222	1		107	115	1		194	216
	148	94B11	1		221	221	1		164	164	2		213	215	1		179	187	1		210	230	1		111	111	1		152	174
	149	94B12	1		221	237	1		150	168	2		229	245	1		175	189	1		208	214	1		113	117	1		180	216
	150	94B15	1		237	239	1		150	150	1	1	211	215	1		185	209	1		210	222	1		111	111	1		154	186
	151	94B16	1		225	233	1	1	150	150	2		213	213	1		185	187	1		212	222	1		111	111	1		164	206
	152	94B17	1		239	239	1	1	146	148	1		215	215	1		183	185	1		210	232	1		111	113	1		166	174
	153	94B18	1		221	221	1		154	154	1		213	213	1		175	179	2		212	256	1		105	111	1		160	160
	154	94B19	1		219	233	1		150	154	2		213	213	1		179	185	1		212	214	2		111	147	1		160	240
	155	94B2	1	1	219	221	1		150	150	2		195	211	1	1	183	195	1		210	214	2		113	129	1	1	186	222
	156	94B20	1		219	233	1		154	154	1		219	219	1		183	189	1		206	210	1		111	125	1		160	170
	157	94B21	1		221	221	2		152	152	2		215	215	2		185	185	1		214	222	1		107	109	1		162	206
	158	94B22	1		221	221	2		152	152	1		213	213	1		179	183	1		206	214	1		105	109	1		178	178
	159	94B3	1		223	223	1		150	150	1		215	215	1		183	185	1		208	208	1		109	115	1		158	166
	160	94B4	1		221	221	1	1	150	154	1		213	213	1		185	189	1	1	212	214	1		111	111	2		282	282
	161	94B5	1		219	221	2	1	148	190	2	1	213	213	1		187	197	1	1	210	212	1		105	117	1		160	166
	162	94B6	1		223	223	1	1	154	154	1	1	213	221	1		175	185	1	1	214	214	1		109	119	1		160	172
	163	94B7	1		219	239	1	1	146	148	2		207	207	1		183	185	1	1	214	214	1		111	119	1		152	174
	164	94B8	1		221	221	1	1	148	150	1		213	213	1		189	189	1	1	210	214	1		111	111	1		162	170
	165	94B9	1		221	221	1	1	148	150	1		197	227	1		175	179	1	1	204	208	1		111	111	1		152	174
118	166	118B1	3		221	221	1	1	152	152	3		195	219	1	1	183	185	2		216	224	2	2	105	105		4	-9	-9
	167	118B10		3	-9	-9	1		148	148	1		213	213	1		183	183	1		206	214	2		123	155	2		178	180
	168	118B11	1		225	225	1		146	146	1		213	213	1		181	183	1		212	212	1		103	103	1		178	184
	169	118B12	1		223	223	1		148	154	1		227	227	1		183	189	2		210	236	1		109	111	1		152	190
	170	118B13	1		225	225	1		148	150	1		213	219	1		201	201	1		210	218	1		109	109	1		188	212
	171	118B14	1		219	219	1		154	154	1		213	213	1		183	185	1		208	216	1		111	115	2		188	190
	172	118B15	1		221	221	1		148	148	1		215	215	1		185	191	1		208	216	2		125	165	1		152	162
	173	118B16	1		219	223	1		148	154	1		213	213	1		183	185	1		206	214	1		109	111	1		166	198
	174	118B17	1		221	223	1		146	150	1		213	213	1		181	221	1		210	224	1		111	111	1		152	174
	175	118B18	1		217	221	1		150	150	1		213	213	1		179	189	1		214	216	1		105	115	1		156	180

Site	N	Name	Brh-1				Brh-2				Brh-3				Brh-4				Brh-5				Brh-6				Brh-7			
			R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles		R	F	Alleles	
	176	118B19	1		221	225	1		150	150	1	1	213	213	1		191	193	1		208	210	2	1	109	109	1	1	160	190
	177	118B2	3		219	219	1		146	150	2		219	219	2		185	199	1		210	214	2	1	103	111	1	1	152	190
	178	118B21	1		223	225	1		150	150	1		213	213	1		187	189	1		210	212	1		109	111	2		198	256
	179	118B22	2		225	225	1		150	150	1		211	211	1		179	193	1		208	212	2		119	181	1		194	204
	180	118B3	1		221	233	1		146	154	1		219	219	1		179	187	1		208	212	1		117	117	2		164	252
	181	118B4	1		221	221	1		150	156	1		213	213	1		183	189	2		212	218	1		111	111	1		166	184
	182	118B5	1		229	229	1		150	152	1		227	227	1		185	187	1		210	212	1		111	111	1		154	158
	183	118B6	1		221	239	1		148	148	1		219	219	1		183	191	1		214	214	1		111	111	1		162	170
	184	118B7	1		221	221	1		154	154	2		213	241	1		185	193	1		210	220	1		105	107	1		156	162
	185	118B8	1		223	233	1		150	150	1		213	213	1		181	197	1		218	218	1		111	111	1		170	212
	186	118B9	1		237	239	1		148	150	1		195	195	1		185	185	1		212	216	1		109	111	1		172	178

Table F9i-xiii. Each table represents one microsatellite loci and shows the method of binning for each allele after scoring. The data in these tables are from the complete dataset. Each row *N* represents a different *Allele*. *Min* and *Max* represent the minimum and maximum value (in base pairs) that can be binned into one *Allele*. *No. alleles* is the total number of times this allele has been scored, not including homozygotes. *Within* is the range (in base pairs) between *Min* and *Max*, *Between* is the difference between the *Max* of one allele and the *Min* of the subsequent allele, and *Between alleles* represents the difference between the *Allele* and the subsequent *Allele*.

Table F9i. Brh-1.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	216.4	217.8	217	4	1.4	0.3	2
2	218.1	219.3	219	64	1.2	0.8	2
3	220.1	221.5	221	123	1.4	0.3	2
4	221.8	223.2	223	41	1.4	0.3	2
5	223.5	226.3	225	35	2.8	0.2	2
6	226.5	227.1	227	7	0.6	1.2	2
7	228.3	230.5	229	22	2.2	0.2	2
8	230.7	233.1	231	18	2.4	0.8	2
9	233.9	235	233	21	1.1	1	4
10	236	237.1	237	12	1.1	1.3	2
11	238.4	239	239	12	0.6	2.9	4
12	241.9	243.2	243	3	1.3	2.9	4
13	246.1	247.7	247	4	1.6		

Table F9ii. Brh-2.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	108.6	108.6	108	2	0	11.5	12
2	120.1	120.1	120	2	0	22.3	22
3	142.4	142.6	142	2	0.2	1.5	2
4	144.1	144.2	144	2	0.1	1.9	2
5	146.1	147.2	146	38	1.1	1.3	2
6	148.5	150.4	148	76	1.9	0.3	2
7	150.7	151.4	150	123	0.7	0.8	2
8	152.2	153.2	152	16	1	0.4	2
9	153.6	154.9	154	71	1.3	1.1	2
10	156	157.1	156	9	1.1	6	8
11	163.1	163.2	164	2	0.1	3.7	2
12	166.9	167.4	166	7	0.5	1.5	2
13	168.9	169.1	168	5	0.2	12	14
14	181.1	181.8	182	3	0.7	3.6	4
15	185.4	185.7	186	2	0.3	3.5	4

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
16	189.2	189.4	190	2	0.2	12	12
17	201.4	201.5	202	2	0.1		

Table F9iii. Brh-3.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	195.8	196.5	195	16	0.7	0.3	2
2	196.8	197.3	197	25	0.5	1.6	2
3	198.9	199	199	2	0.1	8.8	8
4	207.8	208.1	207	2	0.3	3.2	4
5	211.3	212.7	211	22	1.4	0.8	2
6	213.5	215.1	213	184	1.6	0.5	2
7	215.6	217.1	215	52	1.5	0.5	2
8	217.6	217.8	217	2	0.2	0.6	2
9	218.4	218.9	219	18	0.5	1	2
10	219.9	220.1	221	3	0.2	4.6	4
11	224.7	225.2	225	30	0.5	1.8	2
12	227	227.7	227	15	0.7	1.3	2
13	229	229.7	229	3	0.7	1.7	2
14	231.4	231.8	231	4	0.4	2.1	2
15	233.9	235.9	233	7	2	1.9	4
16	237.8	238.1	237	5	0.3	2	4
17	240.1	241	241	2	0.9	3.3	4
18	244.3	244.4	245	2	0.1	10.4	10
19	254.8	255	255	2	0.2		

Table F9iv. Brh-4.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	174.9	175.4	175	18	0.5	1.7	2
2	177.1	177.1	177	2	0	1.9	2
3	179	180.1	179	60	1.1	0.9	2
4	181	181.5	181	11	0.5	1.5	2
5	183	183.8	183	90	0.8	1.3	2
6	185.1	186.7	185	113	1.6	0.5	2
7	187.2	187.7	187	77	0.5	1.5	2
8	189.2	189.7	189	22	0.5	1.6	2
9	191.3	191.5	191	20	0.2	1.9	2
10	193.4	193.7	193	5	0.3	1.7	2
11	195.4	195.7	195	5	0.3	1.7	2
12	197.4	197.6	197	5	0.2	1.9	2
13	199.5	199.8	199	7	0.3	1.6	2

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
14	201.4	201.5	201	6	0.1	2	2
15	203.5	203.7	203	3	0.2	1.7	2
16	205.4	205.5	205	2	0.1	2	2
17	207.5	207.5	207	2	0	1.9	2
18	209.4	209.5	209	2	0.1	4	4
19	213.5	213.5	213	2	0	1	2
20	214.5	215.4	215	4	0.9	4.7	6
21	220.1	221.6	221	4	1.5	1.8	2
22	223.4	223.5	223	2	0.1	8.1	8
23	231.6	231.8	231	2	0.2	4	4
24	235.8	235.9	235	2	0.1	6.2	8
25	242.1	242.1	243	2	0	1.9	2
26	244	244.3	245	3	0.3	1.8	2
27	246.1	246.2	247	4	0.1		

Table F9v. Brh-5.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	203.7	204.4	204	2	0.7	0.9	2
2	205.3	205.9	206	11	0.6	1.5	2
3	207.4	208.2	208	52	0.8	1.2	2
4	209.4	210.2	210	81	0.8	1.1	2
5	211.3	212.2	212	88	0.9	1.2	2
6	213.4	214.1	214	61	0.7	1	2
7	215.1	216.2	216	35	1.1	1.3	2
8	217.5	218.3	218	12	0.8	1.4	2
9	219.7	220.2	220	9	0.5	1.5	2
10	221.7	222.2	222	16	0.5	1.6	2
11	223.8	224.4	224	8	0.6	1.2	2
12	225.6	226.2	226	3	0.6	1.7	2
13	227.9	227.9	228	2	0	2.1	2
14	230	230.6	230	8	0.6	1.2	2
15	231.8	232.5	232	6	0.7	1.6	2
16	234.1	234.7	234	4	0.6	1.3	2
17	236	236.8	236	6	0.8	1.7	2
18	238.5	238.7	238	2	0.2	1.8	2
19	240.5	242	240	2	1.5	0.5	2
20	242.5	242.9	242	2	0.4	1.7	2
21	244.6	244.8	244	3	0.2	3.8	4
22	248.6	248.9	248	2	0.3	7.8	8
23	256.7	256.7	256	2	0	7.9	8
24	264.6	264.9	264	2	0.3	7.9	8
25	272.8	273.1	272	2	0.3	1.6	2
26	274.7	274.8	274	2	0.1	26.3	28
27	301.1	301.4	302	2	0.3		

Table F9*vi* Brh-6

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	103.2	103.9	103	48	0.7	1.5	2
2	105.4	106	105	33	0.6	1.6	2
3	107.6	107.9	107	11	0.3	1.7	2
4	109.6	110.2	109	73	0.6	1.1	2
5	111.3	112.4	111	157	1.1	1.3	2
6	113.7	114.4	113	61	0.7	1.3	2
7	115.7	116.6	115	44	0.9	1.7	2
8	118.3	118.4	117	6	0.1	1.2	2
9	119.6	120.6	119	9	1	1.5	4
10	122.1	123	123	7	0.9	1.6	2
11	124.6	124.9	125	10	0.3	1.8	2
12	126.7	127	127	12	0.3	1.8	2
13	128.8	129	129	5	0.2	1.8	2
14	130.8	132.1	131	2	1.3	4.9	6
15	137	137.3	137	3	0.3	5.8	6
16	143.1	143.2	143	6	0.1	2	2
17	145.2	145.3	145	3	0.1	1.8	2
18	147.1	147.5	147	2	0.4	1.7	2
19	149.2	149.5	149	6	0.3	3.8	4
20	153.3	153.4	153	2	0.1	2.1	2
21	155.5	156.9	155	2	1.4	0.6	2
22	157.5	157.7	157	2	0.2	3.9	4
23	161.6	161.7	161	2	0.1	2	2
24	163.7	163.7	163	2	0	2	2
25	165.7	165.7	165	2	0	14.7	16
26	180.4	180.6	181	2	0.2		

Table F9vii. Brh-7.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	136.8	137.6	136	4	0.8	5.2	6
2	142.8	142.9	142	2	0.1	6.7	8
3	149.6	150.6	150	3	1	1.1	2
4	151.7	152.8	152	24	1.1	1	2
5	153.8	154.8	154	15	1	1.4	2
6	156.2	157.2	156	17	1	1.2	2
7	158.4	159.6	158	12	1.2	0.6	2
8	160.2	161.5	160	24	1.3	0.9	2
9	162.4	163.4	162	28	1	1.1	2
10	164.5	165.5	164	23	1	1.4	2
11	166.9	167.9	166	24	1	0.6	4
12	168.5	170.6	170	26	2.1	0.5	2

Appendix F - Final Data Set

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
13	171.1	172.4	172	32	1.3	0.8	2
14	173.2	174.4	174	29	1.2	1	2
15	175.4	176.6	176	13	1.2	1	2
16	177.6	178.7	178	14	1.1	1.3	2
17	180	180.6	180	13	0.6	1.5	2
18	182.1	183.4	182	12	1.3	0.7	2
19	184.1	185.2	184	14	1.1	1.2	2
20	186.4	187.4	186	16	1	1	2
21	188.4	189.7	188	26	1.3	0.9	2
22	190.6	192	190	14	1.4	0.7	4
23	192.7	193.7	194	15	1	1.3	2
24	195	195.9	196	12	0.9	1.2	2
25	197.1	198.2	198	5	1.1	1.2	2
26	199.4	199.8	200	4	0.4	1.5	2
27	201.3	201.6	202	6	0.3	2.1	2
28	203.7	204.1	204	6	0.4	1.7	2
29	205.8	206.5	206	9	0.7	1.1	2
30	207.6	207.9	208	6	0.3	1.9	2
31	209.8	210.5	210	8	0.7	1.1	2
32	211.6	212.9	212	12	1.3	1	2
33	213.9	214.6	214	4	0.7	1.4	2
34	216	217.1	216	13	1.1	1.1	2
35	218.2	218.4	218	3	0.2	1.6	2
36	220	220.4	220	4	0.4	1.7	2
37	222.1	222.4	222	4	0.3	2.3	4
38	224.7	225.6	226	4	0.9	1.7	2
39	227.3	229.9	228	5	2.6	1.4	4
40	231.3	231.7	232	2	0.4	1.8	2
41	233.5	233.9	234	2	0.4	1.6	2
42	235.5	236.5	236	4	1	1.4	2
43	237.9	238.2	238	4	0.3	1.5	2
44	239.7	240.7	240	4	1	3.8	4
45	244.5	244.5	244	2	0	1.8	2
46	246.3	246.7	246	2	0.4	5.9	6
47	252.6	252.8	252	2	0.2	3.2	4
48	256	256.4	256	2	0.4	4.8	6
49	261.2	261.7	262	2	0.5	3.8	4
50	265.5	265.6	266	4	0.1	14.1	14
51	279.7	280	280	2	0.3	1.8	2
52	281.8	281.9	282	2	0.1	17	18
53	298.9	299.1	300	2	0.2	41.7	42
54	340.8	341.1	342	2	0.3	4	4
55	345.1	345.1	346	2	0		

Table F9viii. B_1.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	127.3	128.3	128	0	1	1	2
2	129.3	130.3	130	0	1	1.4	2
3	131.7	133	132	0	1.3	0.9	2
4	133.9	134.8	134	0	0.9	1.1	2
5	135.9	136.4	136	0	0.5	2.4	2
6	138.8	139.2	138	0	0.4	1	2
7	140.2	141.1	140	0	0.9	1.6	2
8	142.7	142.7	142	0	0	2.2	2
9	144.9	144.9	144	0	0	1.9	2
10	146.8	147.1	146	0	0.3	3.9	6
11	151	151.5	152	0	0.5	2	2
12	153.5	153.5	154	0	0	1.9	2
13	155.4	155.7	156	0	0.3	1.5	2
14	157.2	157.7	158	0	0.5	1.8	2
15	159.5	159.6	160	0	0.1	5.9	6
16	165.5	165.5	166	0	0	1.1	2
17	166.6	167.7	168	0	1.1	1.7	2
18	169.4	169.6	170	0	0.2	29.8	30
19	199.4	199.6	200	0	0.2		

Table F9ix. B_2.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	222	223	222	20	1	1	2
2	224	225.1	224	36	1.1	1.3	2
3	226.4	227.8	226	23	1.4	0.5	2
4	228.3	229.8	228	120	1.5	0.8	2
5	230.6	231.6	230	46	1	1.1	2
6	232.7	233.8	232	41	1.1	1	2
7	234.8	235.8	234	31	1	1.3	2
8	237.1	237.8	236	9	0.7	0.9	2
9	238.7	240	238	25	1.3	0.9	2
10	240.9	242.4	240	14	1.5	1.8	4
11	244.2	245.2	244	2	1	1.1	2
12	246.3	247.4	246	4	1.1	0.9	2
13	248.3	249.5	248	7	1.2	1.6	4
14	251.1	252.8	252	5	1.7	1.9	2
15	254.7	254.9	254	3	0.2	4.5	6
16	259.4	259.9	260	2	0.5	1.1	2
17	261	261.1	262	2	0.1	3.7	2
18	264.8	265.6	264	2	0.8	4.6	6
19	270.2	270.4	270	2	0.2	3.6	4
20	274	274.4	274	3	0.4	6.3	6

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
21	280.7	280.8	280	2	0.1		

Table F9x. B_3.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	167	167.4	167	15	0.4	3.9	4
2	171.3	171.4	171	2	0.1	3.9	4
3	175.3	175.6	175	2	0.3	1.7	2
4	177.3	177.5	177	3	0.2	7.5	8
5	185	186.2	185	13	1.2	0.8	2
6	187	187.5	187	11	0.5	0.5	2
7	188	189.6	189	51	1.6	1.4	2
8	191	191.8	191	57	0.8	1.4	2
9	193.2	194.3	193	39	1.1	0.9	2
10	195.2	196.6	195	41	1.4	0.7	2
11	197.3	197.9	197	38	0.6	1.5	2
12	199.4	200.8	199	33	1.4	0.8	2
13	201.6	202.9	201	24	1.3	1	2
14	203.9	204.5	203	5	0.6	0.5	2
15	205	205.5	205	12	0.5	0.7	2
16	206.2	207.3	207	6	1.1	1.3	2
17	208.6	208.6	209	2	0	2.1	2
18	210.7	211.1	211	5	0.4	2.9	4
19	214	215.2	215	4	1.2	2.3	2
20	217.5	218	217	3	0.5	2.6	4
21	220.6	220.6	221	2	0	11	10
22	231.6	231.8	231	3	0.2	8.7	9
23	240.5	241.3	241	2	0.8		

Table F9*xi*. B_4.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	171.1	171.2	172	2	0.1	9.4	8
2	180.6	181.1	180	19	0.5	1.5	2
3	182.6	182.9	182	19	0.3	0.8	2
4	183.7	184.1	184	12	0.4	0.5	2
5	184.6	185.9	186	161	1.3	0.6	2
6	186.5	187.8	188	78	1.3	0.8	2
7	188.6	188.9	190	10	0.3	0.9	2
8	189.8	189.9	192	2	0.1	1.8	2
9	191.7	191.9	194	3	0.2	9.7	8
10	201.6	202.1	202	2	0.5	1.6	2
11	203.7	204.1	204	7	0.4		

Table F9*xii*. B_5.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	140	140.2	140	2	0.2	1.4	2
2	141.6	142.3	142	93	0.7	1.3	2
3	143.6	144.6	144	46	1	0.9	2
4	145.5	146.6	146	15	1.1	1.7	2
5	148.3	148.9	148	47	0.6	1.6	2
6	150.5	151.4	150	25	0.9	0.9	2
7	152.3	153.4	152	66	1.1	0.9	2
8	154.3	155.2	154	7	0.9	3.7	4
9	158.9	159.3	158	2	0.4	1.2	2
10	160.5	160.7	160	3	0.2	1.8	2
11	162.5	162.7	162	2	0.2	1.6	2
12	164.3	164.5	164	5	0.2	1.6	2
13	166.1	166.4	166	5	0.3	1.7	2
14	168.1	168.2	168	2	0.1	1.9	2
15	170.1	170.2	170	2	0.1	3.4	4
16	173.6	173.7	174	2	0.1	3.5	4
17	177.2	177.3	178	2	0.1	4.1	4
18	181.4	181.6	182	4	0.2	1.7	2
19	183.3	183.6	184	2	0.3	1.7	2
20	185.3	185.6	186	2	0.3	7.5	6
21	193.1	193.3	192	2	0.2	1.2	2
22	194.5	194.8	194	4	0.3	1.8	2
23	196.6	197	196	2	0.4	1.6	2
24	198.6	199.1	198	3	0.5	1.4	2
25	200.5	200.7	200	3	0.2	1.9	2
26	202.6	203	202	2	0.4	1.5	2
27	204.5	204.7	204	3	0.2	1.2	2
28	205.9	206.9	206	4	1	11.9	12
29	218.8	219	218	2	0.2	18	20
30	237	237.6	238	4	0.6	1.8	2
31	239.4	240.2	240	2	0.8	4.9	6
32	245.1	245.7	246	2	0.6	3.5	4
33	249.2	249.5	250	2	0.3	7.4	6
34	256.9	257.6	256	2	0.7	4.6	6
35	262.2	263.3	262	2	1.1	6	8
36	269.3	269.6	270	3	0.3	7.2	6
37	276.8	277	276	3	0.2	25.4	26
38	302.4	302.6	302	3	0.2	12.1	12
39	314.7	315	314	3	0.3	13.9	14
40	328.9	330.7	328	2	1.8		

Table F9*xiii*. B_7.

N	Min (bp)	Max (bp)	Allele	No. alleles	Within (bp)	Between (bp)	Between alleles (bp)
1	230.6	231.1	230	5	0.5	9.3	10
2	240.4	240.7	240	2	0.3	1.2	2
3	241.9	243.8	242	140	1.9	0.4	2
4	244.2	245.5	244	49	1.3	0.7	2
5	246.2	247.3	246	38	1.1	0.8	2
6	248.1	249.1	248	26	1	0.4	2
7	249.5	251.7	250	48	2.2	1	2
8	252.7	252.8	252	2	0.1	5.7	6
9	258.5	258.6	258	2	0.1	0.8	2
10	259.4	260.4	260	2	1	1.4	2
11	261.8	262	262	2	0.2	3.6	4
12	265.6	265.9	266	2	0.3	1.7	2
13	267.6	268.1	268	4	0.5	5	6
14	273.1	273.4	274	3	0.3	1.9	2
15	275.3	275.4	276	4	0.1	1.3	2
16	276.7	277.4	278	7	0.7	1.1	2
17	278.5	279.1	280	2	0.6	1.8	2
18	280.9	281	282	2	0.1	1.3	2
19	282.3	282.9	284	2	0.6	1.9	2
20	284.8	284.8	286	3	0	1.2	2
21	286	286.6	288	2	0.6	14.5	14
22	301.1	301.7	302	2	0.6	9.6	10
23	311.3	312	312	2	0.7	5.5	6
24	317.5	318	318	2			

References

Walsh, P. S., Metzger, D. A. and Higuchi, R. (1991). Chelex-100 as a medium for simple extraction of DNA for PCR-basedtyping from forensic material. *BioTechniques* **10:**506-513.

Appendix G - Descriptive Statistics per Loci

Appendix G contains descriptive statistics per site, per loci for all species, using data where half sibs and closer related individuals were removed, calculated as described within the chapter.

Table of tables

Table G1 458

Table G2 460

Table G3 463

Appendix G - Descriptive Statistics per Loci

Table G1. Summary statistics for *Amphinemura sulcicollis* per loci, per site (10 loci, 13 sites). Sites with less than 10 samples were excluded (Site codes 46, 94, 4, 55). Including *N* = number of individuals; N_A = number of alleles per locus; H_O = observed heterozygosity; H_E = expected heterozygosity; *HWE* = p value of Hardy-Weinberg Equilibrium test (**bold** = significant disequilibrium after Bonferroni test (critical P value p ≤0.005)), and *PIC* = polymorphism information content.

Site	Locus	N	N_a	H_o	H_e	HWE	PIC
104	Amp_2	17	16	1.000	0.947	1.000	0.91
	Amp_3	17	10	0.765	0.766	0.524	0.72
	Amp_4	17	7	0.824	0.852	0.454	0.80
	Amp_5	17	13	1.000	0.900	1.000	0.86
	Amp_6	17	5	0.882	0.663	0.756	0.58
	Amp_8	17	11	0.529	0.900	**0.000**	0.86
	Amp_9	17	10	0.294	0.731	**0.000**	0.69
	Amp_10	17	11	0.824	0.834	0.254	0.79
	Amp_11	17	15	0.353	0.936	**0.000**	0.90
	Amp_13	17	10	0.294	0.889	**0.000**	0.85
109	Amp_2	16	14	1.000	0.897	1.000	0.86
	Amp_3	16	7	0.688	0.661	0.755	0.61
	Amp_4	16	8	1.000	0.837	1.000	0.79
	Amp_5	16	12	0.938	0.909	0.645	0.87
	Amp_6	16	7	0.813	0.815	0.256	0.76
	Amp_8	16	9	0.438	0.879	**0.000**	0.83
	Amp_9	16	7	0.563	0.821	0.010	0.77
	Amp_10	15	10	0.267	0.837	**0.000**	0.79
	Amp_11	16	13	0.500	0.885	**0.000**	0.84
	Amp_13	16	8	0.563	0.829	**0.004**	0.78
108	Amp_2	16	13	0.938	0.915	0.812	0.88
	Amp_3	16	10	0.750	0.679	0.975	0.64
	Amp_4	16	7	0.813	0.774	0.756	0.71
	Amp_5	16	14	0.938	0.911	0.663	0.87
	Amp_6	16	5	0.563	0.593	0.289	0.52
	Amp_8	16	10	0.875	0.865	0.662	0.82
	Amp_9	16	6	0.438	0.712	0.022	0.64
	Amp_10	14	13	0.643	0.889	0.026	0.84
	Amp_11	16	15	0.500	0.944	**0.000**	0.91
	Amp_13	16	8	0.500	0.778	**0.001**	0.72
102	Amp_2	17	14	0.824	0.934	0.184	0.90
	Amp_3	17	5	0.588	0.622	0.452	0.57
	Amp_4	17	6	0.706	0.825	0.089	0.77
	Amp_5	17	12	0.824	0.900	0.210	0.86
	Amp_6	17	4	0.765	0.663	0.892	0.57
	Amp_8	17	9	0.824	0.863	0.145	0.82
	Amp_9	17	7	0.412	0.713	**0.000**	0.64
	Amp_10	17	11	0.412	0.770	**0.000**	0.72
	Amp_11	17	11	0.294	0.902	**0.000**	0.86
	Amp_13	17	6	0.471	0.738	0.014	0.68
112	Amp_2	17	21	1.000	0.952	1.000	0.92
	Amp_3	17	8	0.706	0.608	0.976	0.57
	Amp_4	17	7	0.824	0.838	0.500	0.79
	Amp_5	17	12	0.765	0.914	0.043	0.88
	Amp_6	17	5	0.765	0.579	0.994	0.50
	Amp_8	17	9	0.765	0.802	0.389	0.75

Appendix G - Descriptive Statistics per Loci

Site	Locus	N	N_a	H_o	H_e	HWE	PIC
	Amp_9	17	8	0.294	0.613	**0.000**	0.58
	Amp_10	17	11	0.529	0.904	**0.000**	0.87
	Amp_11	17	15	0.412	0.923	**0.000**	0.89
	Amp_13	17	8	0.412	0.790	**0.000**	0.74
96	Amp_2	15	15	0.867	0.936	0.169	0.90
	Amp_3	15	10	0.800	0.722	0.921	0.67
	Amp_4	15	5	0.867	0.791	0.878	0.73
	Amp_5	15	10	0.800	0.853	0.037	0.80
	Amp_6	15	4	0.600	0.687	0.188	0.60
	Amp_8	15	9	0.733	0.878	0.021	0.83
	Amp_9	15	7	0.400	0.738	0.013	0.68
	Amp_10	15	6	0.133	0.618	**0.000**	0.57
	Amp_11	14	9	0.429	0.849	**0.000**	0.80
	Amp_13	15	8	0.533	0.786	**0.000**	0.73
9	Amp_2	18	18	0.944	0.927	0.184	0.89
	Amp_3	18	7	0.500	0.543	0.428	0.50
	Amp_4	18	8	0.833	0.867	0.171	0.82
	Amp_5	18	11	0.889	0.894	0.443	0.86
	Amp_6	18	5	0.444	0.722	**0.003**	0.66
	Amp_8	18	10	0.722	0.851	0.035	0.81
	Amp_9	18	6	0.500	0.708	0.065	0.65
	Amp_10	17	9	0.353	0.740	**0.001**	0.69
	Amp_11	18	14	0.556	0.892	**0.000**	0.86
	Amp_13	18	7	0.556	0.752	**0.004**	0.69
116	Amp_2	19	18	0.895	0.943	0.294	0.91
	Amp_3	19	5	0.316	0.290	1.000	0.27
	Amp_4	19	9	1.000	0.865	1.000	0.82
	Amp_5	19	11	0.737	0.794	0.507	0.75
	Amp_6	19	5	0.842	0.733	0.697	0.66
	Amp_8	19	10	0.789	0.836	0.299	0.80
	Amp_9	19	5	0.368	0.653	0.013	0.60
	Amp_10	11	11	0.364	0.922	**0.000**	0.87
	Amp_11	19	14	0.526	0.838	**0.000**	0.80
	Amp_13	19	7	0.474	0.804	**0.000**	0.76
113	Amp_2	16	15	0.813	0.911	0.192	0.87
	Amp_3	16	7	0.750	0.734	0.064	0.68
	Amp_4	16	8	0.938	0.853	0.920	0.80
	Amp_5	16	13	0.938	0.925	0.525	0.89
	Amp_6	16	6	0.750	0.698	0.178	0.62
	Amp_8	16	8	0.500	0.893	**0.000**	0.85
	Amp_9	16	8	0.438	0.669	**0.000**	0.63
	Amp_10	15	11	0.533	0.853	**0.000**	0.81
	Amp_11	16	11	0.500	0.893	**0.000**	0.85
	Amp_13	16	8	0.500	0.798	**0.000**	0.75
95	Amp_2	16	13	0.688	0.827	**0.000**	0.79
	Amp_3	16	7	0.813	0.808	0.427	0.75
	Amp_4	16	7	0.938	0.837	0.954	0.78
	Amp_5	16	10	0.750	0.641	0.976	0.60
	Amp_6	16	5	0.875	0.762	0.845	0.69
	Amp_8	16	8	0.875	0.867	0.631	0.82
	Amp_9	16	9	0.375	0.827	**0.000**	0.78
	Amp_10	15	8	0.600	0.800	0.052	0.74
	Amp_11	16	10	0.250	0.810	**0.000**	0.77
	Amp_13	16	7	0.438	0.778	**0.000**	0.71
6	Amp_2	17	10	0.706	0.797	0.094	0.75

Site	Locus	N	N_a	H_o	H_e	HWE	PIC
	Amp_3	17	6	0.471	0.565	0.195	0.52
	Amp_4	17	9	0.882	0.850	0.271	0.80
	Amp_5	17	7	0.471	0.497	0.180	0.46
	Amp_6	17	4	0.647	0.586	0.760	0.51
	Amp_8	17	10	0.824	0.879	0.210	0.84
	Amp_9	17	7	0.588	0.745	0.011	0.68
	Amp_10	16	6	0.375	0.750	**0.000**	0.69
	Amp_11	17	10	0.412	0.800	**0.000**	0.75
	Amp_13	17	5	0.353	0.729	**0.000**	0.65
59	Amp_2	18	15	1.000	0.929	1.000	0.90
	Amp_3	18	9	0.667	0.648	0.721	0.61
	Amp_4	18	9	0.667	0.817	0.056	0.77
	Amp_5	18	11	1.000	0.867	1.000	0.83
	Amp_6	18	6	0.556	0.670	0.049	0.59
	Amp_8	18	12	0.889	0.908	0.501	0.87
	Amp_9	18	11	0.333	0.824	**0.000**	0.79
	Amp_10	18	8	0.500	0.819	**0.000**	0.77
	Amp_11	18	18	0.444	0.933	**0.000**	0.90
	Amp_13	18	9	0.444	0.802	**0.002**	0.75
93	Amp_2	16	14	0.938	0.913	0.503	0.87
	Amp_3	16	8	0.563	0.637	0.218	0.60
	Amp_4	16	6	0.625	0.784	0.039	0.72
	Amp_5	16	8	0.813	0.835	0.553	0.78
	Amp_6	16	5	0.688	0.760	0.088	0.69
	Amp_8	16	8	0.688	0.851	0.076	0.80
	Amp_9	16	6	0.313	0.435	0.016	0.40
	Amp_10	16	10	0.625	0.833	**0.004**	0.79
	Amp_11	16	12	0.438	0.875	**0.000**	0.83
	Amp_13	16	6	0.313	0.558	**0.005**	0.52

Table G2. Summary statistics for *Isoperla grammatica* per loci, per site (10 loci, 12 sites). Sites with less than 10 samples were excluded (Site code 46). Including N = number of individuals; N_A = number of alleles per locus; H_O = observed heterozygosity; H_E = expected heterozygosity; *HWE* = p value of Hardy-Weinberg Equilibrium test (**bold** = significant disequilibrium after Bonferroni test (critical P value p ≤0.005)), and *PIC* = polymorphism information content.

Site	Locus	N	N_a	H_o	H_e	HWE	PIC
105	Iso_1	15	13	0.867	0.910	0.171	0.87
	Iso_2	15	21	0.867	0.972	**0.000**	0.94
	Iso_3	15	17	0.733	0.945	0.030	0.91
	Iso_4	15	16	0.600	0.947	**0.000**	0.91
	Iso_5	15	16	0.600	0.954	**0.000**	0.92
	Iso_6	15	16	0.467	0.933	**0.000**	0.90
	Iso_7	15	18	0.867	0.956	0.244	0.92
	Iso_8	15	19	0.733	0.970	**0.000**	0.93
	Iso_9	14	17	0.429	0.955	**0.000**	0.92
	Iso_10	15	10	0.467	0.883	**0.000**	0.84
106	Iso_1	20	16	0.750	0.885	0.257	0.85
	Iso_2	20	23	0.800	0.969	**0.000**	0.94
	Iso_3	20	22	0.800	0.964	0.026	0.94

Appendix G - Descriptive Statistics per Loci

Site	Locus	N	N_a	H_o	H_e	HWE	PIC
	Iso_4	20	21	0.750	0.954	**0.000**	0.93
	Iso_5	19	21	0.526	0.957	**0.000**	0.93
	Iso_6	20	16	0.700	0.894	**0.000**	0.86
	Iso_7	20	20	0.700	0.949	0.006	0.92
	Iso_8	20	22	0.850	0.962	0.035	0.93
	Iso_9	18	18	0.222	0.954	**0.000**	0.92
	Iso_10	20	15	0.300	0.899	**0.000**	0.87
108	Iso_1	20	17	1.000	0.942	1.000	0.91
	Iso_2	20	25	0.900	0.959	0.208	0.93
	Iso_3	20	14	0.900	0.914	0.445	0.88
	Iso_4	19	28	0.842	0.982	**0.000**	0.95
	Iso_5	18	22	0.333	0.976	**0.000**	0.95
	Iso_6	20	15	0.500	0.932	**0.000**	0.90
	Iso_7	19	26	0.947	0.980	0.145	0.95
	Iso_8	20	22	0.700	0.965	**0.000**	0.94
	Iso_9	17	16	0.294	0.936	**0.000**	0.90
	Iso_10	20	9	0.500	0.864	**0.000**	0.82
12	Iso_1	20	16	1.000	0.909	1.000	0.88
	Iso_2	20	25	0.900	0.972	0.024	0.95
	Iso_3	20	20	0.750	0.959	**0.000**	0.93
	Iso_4	20	20	0.850	0.940	**0.005**	0.91
	Iso_5	20	21	0.550	0.937	**0.000**	0.91
	Iso_6	20	17	0.450	0.944	**0.000**	0.92
	Iso_7	20	19	0.500	0.954	**0.000**	0.93
	Iso_8	20	21	0.500	0.956	**0.000**	0.93
	Iso_9	19	24	0.526	0.976	**0.000**	0.95
	Iso_10	20	9	0.250	0.815	**0.000**	0.77
97	Iso_1	17	15	0.941	0.941	0.584	0.91
	Iso_2	17	19	0.647	0.959	**0.000**	0.93
	Iso_3	17	20	0.941	0.945	0.507	0.91
	Iso_4	17	21	0.706	0.964	**0.000**	0.93
	Iso_5	17	17	0.235	0.941	**0.000**	0.91
	Iso_6	17	16	0.353	0.925	**0.000**	0.89
	Iso_7	17	23	0.882	0.971	0.425	0.94
	Iso_8	17	20	0.412	0.968	**0.000**	0.94
	Iso_9	15	17	0.200	0.963	**0.000**	0.93
	Iso_10	17	12	0.529	0.904	**0.000**	0.87
98	Iso_1	20	18	1.000	0.936	1.000	0.91
	Iso_2	20	24	0.750	0.973	**0.000**	0.95
	Iso_3	20	20	0.900	0.947	0.209	0.92
	Iso_4	20	25	0.750	0.962	**0.000**	0.93
	Iso_5	18	20	0.500	0.957	**0.000**	0.93
	Iso_6	20	20	0.400	0.955	**0.000**	0.93
	Iso_7	20	19	0.650	0.945	**0.000**	0.92
	Iso_8	20	19	0.700	0.954	**0.000**	0.93
	Iso_9	17	18	0.529	0.957	**0.000**	0.93
	Iso_10	20	11	0.700	0.873	0.081	0.84
112	Iso_1	20	18	0.900	0.947	0.191	0.92
	Iso_2	20	23	0.850	0.968	0.066	0.94
	Iso_3	20	21	0.900	0.953	0.122	0.92
	Iso_4	20	24	0.700	0.974	**0.000**	0.95
	Iso_5	20	21	0.600	0.951	**0.000**	0.92
	Iso_6	20	16	0.650	0.926	**0.000**	0.90
	Iso_7	20	21	0.750	0.960	**0.000**	0.93
	Iso_8	20	23	0.650	0.958	**0.000**	0.93

Appendix G - Descriptive Statistics per Loci

Site	Locus	N	N_a	H_o	H_e	HWE	PIC
	Iso_9	19	21	0.263	0.967	**0.000**	0.94
	Iso_10	20	12	0.400	0.890	**0.000**	0.85
114	Iso_1	19	18	0.842	0.925	0.007	0.89
	Iso_2	19	20	0.789	0.942	0.009	0.91
	Iso_3	19	18	0.947	0.930	0.797	0.90
	Iso_4	19	20	0.737	0.966	**0.000**	0.94
	Iso_5	19	18	0.474	0.939	**0.000**	0.91
	Iso_6	17	17	0.588	0.955	**0.000**	0.92
	Iso_7	18	20	0.722	0.927	**0.000**	0.90
	Iso_8	19	20	0.632	0.960	**0.000**	0.93
	Iso_9	18	23	0.500	0.973	**0.000**	0.94
	Iso_10	19	11	0.474	0.875	**0.000**	0.84
115	Iso_1	20	20	0.900	0.947	0.046	0.92
	Iso_2	20	25	0.950	0.976	0.470	0.95
	Iso_3	20	23	0.850	0.958	0.019	0.93
	Iso_4	20	23	0.750	0.960	**0.000**	0.93
	Iso_5	19	17	0.632	0.940	**0.000**	0.91
	Iso_6	19	25	0.684	0.974	**0.000**	0.95
	Iso_7	20	21	0.650	0.955	**0.000**	0.93
	Iso_8	20	23	0.700	0.972	**0.000**	0.95
	Iso_9	18	18	0.333	0.921	**0.000**	0.89
	Iso_10	20	11	0.550	0.878	**0.000**	0.84
10	Iso_1	20	15	0.750	0.903	0.018	0.87
	Iso_2	20	25	0.900	0.965	**0.000**	0.94
	Iso_3	20	15	0.800	0.918	0.083	0.89
	Iso_4	20	26	0.850	0.976	0.010	0.95
	Iso_5	19	20	0.684	0.953	**0.000**	0.92
	Iso_6	19	22	0.737	0.952	**0.000**	0.92
	Iso_7	20	17	0.650	0.940	**0.000**	0.91
	Iso_8	20	23	0.800	0.963	**0.000**	0.94
	Iso_9	19	24	0.579	0.974	**0.000**	0.95
	Iso_10	20	12	0.300	0.873	**0.000**	0.84
118	Iso_1	20	13	0.850	0.882	0.434	0.85
	Iso_2	20	25	0.950	0.972	0.625	0.95
	Iso_3	20	22	0.800	0.964	**0.000**	0.94
	Iso_4	20	25	0.750	0.973	**0.000**	0.95
	Iso_5	20	16	0.550	0.926	**0.000**	0.90
	Iso_6	20	17	0.400	0.905	**0.000**	0.88
	Iso_7	20	21	0.650	0.960	**0.000**	0.93
	Iso_8	20	23	0.550	0.967	**0.000**	0.94
	Iso_9	18	19	0.556	0.937	**0.000**	0.91
	Iso_10	20	10	0.350	0.882	**0.000**	0.85
93	Iso_1	19	12	0.632	0.757	0.128	0.72
	Iso_2	19	22	0.895	0.956	0.202	0.93
	Iso_3	18	15	0.889	0.892	0.033	0.86
	Iso_4	19	27	0.947	0.970	0.688	0.94
	Iso_5	19	18	0.632	0.888	**0.000**	0.86
	Iso_6	19	19	0.947	0.893	0.930	0.86
	Iso_7	19	19	0.526	0.949	**0.000**	0.92
	Iso_8	19	16	0.737	0.910	**0.000**	0.88
	Iso_9	17	14	0.471	0.930	**0.000**	0.90
	Iso_10	19	11	0.421	0.683	**0.000**	0.65

Appendix G - Descriptive Statistics per Loci

Table G3. Summary statistics for *Baetis rhodani* per loci, per site (13 loci, 10 sites, no sites excluded). Including N = number of individuals; N_A = number of alleles per locus; H_O = observed heterozygosity; H_E = expected heterozygosity; *HWE* = p value of Hardy-Weinberg Equilibrium test (**bold** = significant disequilibrium after Bonferroni test (critical P value p ≤0.0038)), and *PIC* = polymorphism information content.

Site	Locus	N	N_a	H_o	H_e	HWE	PIC
112	Brh-1	16	9	0.438	0.873	**0.000**	0.83
	Brh-2	16	5	0.375	0.714	**0.003**	0.65
	Brh-3	16	6	0.188	0.560	**0.000**	0.52
	Brh-4	16	10	0.875	0.853	0.659	0.81
	Brh-5	16	10	0.563	0.855	**0.000**	0.81
	Brh-6	16	8	0.625	0.823	0.105	0.77
	Brh-7	16	17	0.813	0.952	0.103	0.92
	B_1	16	8	0.625	0.736	0.201	0.67
	B_2	16	9	0.875	0.863	0.492	0.82
	B_3	16	10	0.750	0.883	0.106	0.84
	B_4	16	6	0.313	0.673	**0.001**	0.61
	B_5	16	7	0.313	0.855	**0.000**	0.81
	B_7	16	9	0.688	0.829	0.111	0.78
102	Brh-1	19	6	0.316	0.745	**0.000**	0.69
	Brh-2	18	8	0.389	0.775	**0.001**	0.72
	Brh-3	19	9	0.316	0.818	**0.000**	0.77
	Brh-4	20	12	0.700	0.871	**0.003**	0.83
	Brh-5	20	15	0.700	0.890	**0.000**	0.86
	Brh-6	20	8	0.700	0.760	0.332	0.70
	Brh-7	20	21	0.800	0.971	**0.000**	0.94
	B_1	20	9	0.750	0.782	0.132	0.73
	B_2	20	12	0.900	0.872	0.294	0.84
	B_3	20	11	0.750	0.906	0.026	0.87
	B_4	20	7	0.400	0.646	**0.000**	0.60
	B_5	19	11	0.632	0.829	**0.000**	0.78
	B_7	20	11	0.650	0.732	0.023	0.70
97	Brh-1	14	5	0.143	0.659	**0.000**	0.59
	Brh-2	16	4	0.313	0.683	**0.001**	0.60
	Brh-3	16	6	0.125	0.736	**0.000**	0.67
	Brh-4	16	9	0.875	0.853	0.629	0.81
	Brh-5	16	11	0.938	0.863	0.946	0.82
	Brh-6	16	6	0.625	0.776	0.019	0.71
	Brh-7	16	23	1.000	0.976	1.000	0.94
	B_1	16	8	0.750	0.780	0.130	0.72
	B_2	16	12	0.875	0.911	0.438	0.87
	B_3	16	15	0.813	0.921	**0.003**	0.88
	B_4	16	6	0.375	0.633	0.040	0.55
	B_5	16	8	0.563	0.794	0.025	0.74
	B_7	16	9	0.563	0.569	0.702	0.54
96	Brh-1	17	8	0.294	0.754	**0.000**	0.69
	Brh-2	16	6	0.500	0.752	**0.001**	0.69
	Brh-3	17	6	0.235	0.733	**0.000**	0.68
	Brh-4	17	13	0.824	0.897	0.017	0.86
	Brh-5	17	12	0.765	0.877	0.021	0.84
	Brh-6	17	12	0.647	0.868	0.007	0.83
	Brh-7	17	21	0.647	0.971	**0.000**	0.94

Appendix G - Descriptive Statistics per Loci

Site	Locus	N	N_a	H_o	H_e	HWE	PIC
	B_1	17	9	0.706	0.750	0.385	0.70
	B_2	17	10	0.882	0.877	0.508	0.84
	B_3	17	16	0.941	0.936	0.688	0.90
	B_4	17	5	0.353	0.597	**0.002**	0.55
	B_5	17	10	0.529	0.866	**0.001**	0.82
	B_7	17	10	0.706	0.756	0.398	0.70
9	Brh-1	10	6	0.300	0.784	**0.000**	0.71
	Brh-2	11	6	0.636	0.797	0.255	0.72
	Brh-3	11	8	0.455	0.784	**0.001**	0.73
	Brh-4	10	9	0.600	0.868	0.027	0.81
	Brh-5	11	9	0.727	0.805	0.286	0.75
	Brh-6	8	9	0.875	0.917	0.231	0.85
	Brh-7	11	18	0.909	0.983	0.079	0.94
	B_1	11	6	0.636	0.758	0.175	0.69
	B_2	11	9	0.818	0.892	0.419	0.84
	B_3	11	11	0.727	0.909	0.083	0.86
	B_4	11	5	0.455	0.649	0.102	0.55
	B_5	11	8	0.545	0.779	0.016	0.72
	B_7	11	6	0.636	0.632	0.317	0.57
106	Brh-1	18	8	0.444	0.798	**0.001**	0.75
	Brh-2	17	7	0.588	0.784	**0.005**	0.72
	Brh-3	18	7	0.222	0.697	**0.000**	0.65
	Brh-4	18	7	0.667	0.790	0.248	0.74
	Brh-5	18	10	0.778	0.829	0.423	0.78
	Brh-6	18	11	0.722	0.856	0.193	0.82
	Brh-7	18	21	0.778	0.957	**0.000**	0.93
	B_1	18	5	0.778	0.683	0.871	0.61
	B_2	18	11	0.611	0.783	0.030	0.74
	B_3	18	11	0.667	0.903	**0.001**	0.87
	B_4	18	5	0.333	0.583	**0.004**	0.52
	B_5	18	10	0.611	0.827	**0.000**	0.78
	B_7	17	5	0.294	0.729	**0.000**	0.66
113	Brh-1	15	6	0.133	0.756	**0.000**	0.69
	Brh-2	16	7	0.625	0.790	**0.003**	0.73
	Brh-3	16	4	0.188	0.286	0.091	0.26
	Brh-4	16	8	0.750	0.810	0.038	0.76
	Brh-5	16	11	0.688	0.853	0.018	0.81
	Brh-6	16	10	0.438	0.808	**0.000**	0.76
	Brh-7	16	20	0.813	0.966	0.009	0.93
	B_1	16	8	0.938	0.792	0.984	0.73
	B_2	16	12	0.875	0.869	0.383	0.83
	B_3	16	11	0.875	0.899	0.140	0.86
	B_4	16	5	0.625	0.639	0.081	0.58
	B_5	16	14	0.625	0.897	**0.000**	0.86
	B_7	16	6	0.750	0.750	0.372	0.68
115	Brh-1	16	9	0.375	0.835	**0.000**	0.78
	Brh-2	15	7	0.400	0.740	**0.002**	0.68
	Brh-3	16	7	0.125	0.772	**0.000**	0.72
	Brh-4	16	14	0.875	0.907	0.645	0.87
	Brh-5	14	10	0.857	0.857	0.413	0.81
	Brh-6	16	10	0.563	0.800	0.007	0.75
	Brh-7	16	22	0.938	0.976	0.269	0.94
	B_1	16	8	0.750	0.756	0.336	0.69
	B_2	16	11	1.000	0.877	1.000	0.83
	B_3	16	12	0.813	0.907	0.159	0.87

Appendix G - Descriptive Statistics per Loci

Site	Locus	N	N_a	H_o	H_e	HWE	PIC
	B_4	16	6	0.438	0.617	0.007	0.54
	B_5	16	12	0.563	0.911	**0.000**	0.87
	B_7	16	5	0.625	0.619	0.666	0.56
94	Brh-1	16	7	0.500	0.764	0.009	0.71
	Brh-2	16	8	0.438	0.825	**0.001**	0.77
	Brh-3	16	11	0.313	0.815	**0.000**	0.77
	Brh-4	16	8	0.938	0.849	0.939	0.80
	Brh-5	16	9	0.813	0.833	0.371	0.79
	Brh-6	16	12	0.750	0.841	0.361	0.80
	Brh-7	16	19	0.875	0.958	0.017	0.92
	B_1	16	4	0.688	0.718	0.245	0.65
	B_2	16	13	0.750	0.821	0.234	0.78
	B_3	16	13	0.750	0.925	0.009	0.89
	B_4	16	4	0.375	0.518	0.136	0.44
	B_5	16	9	0.438	0.841	**0.000**	0.79
	B_7	15	5	0.267	0.662	**0.000**	0.57
118	Brh-1	19	9	0.474	0.828	**0.000**	0.78
	Brh-2	20	6	0.450	0.782	**0.001**	0.73
	Brh-3	20	6	0.150	0.655	**0.000**	0.60
	Brh-4	20	12	0.850	0.885	0.113	0.85
	Brh-5	20	10	0.850	0.887	0.200	0.85
	Brh-6	20	11	0.500	0.796	**0.000**	0.75
	Brh-7	19	20	1.000	0.963	1.000	0.93
	B_1	20	7	0.700	0.703	0.533	0.63
	B_2	20	11	0.800	0.849	0.065	0.81
	B_3	20	11	0.700	0.899	**0.001**	0.86
	B_4	20	6	0.450	0.610	0.038	0.56
	B_5	20	11	0.650	0.883	**0.000**	0.85
	B_7	20	8	0.750	0.750	0.607	0.69

Appendix H - Chapter 3 Supporting Information

Containing extra information pertaining to Chapter 3, including values of environmental stressors and at which sites they are available; allelic richness values and the complete results of comparing allelic richness between sites. Also containing maps showing population structure for each species and altitude in Wales, UK.

Table of tables

Table H1 467

Table H2a-c 468

Table H3a-c 470

Table of figures

Figure H1 481

Figure H2 482

Table H1. Table showing all sites (listed from south to North of Wales) with genetic data for one species or more, sample size listed under species: *A* = *Amphinemura sulcicollis*, *I* = *Isoperla grammatica* and *B* = *Baetis rhodani*; and environmental variables: *Al* = Aluminium, *Cd* = Cadmium, *pH* and *TON* = Total Oxidised Nitrogen.

Site code	Site name	Year	Species			Environmental stressors			
			A	I	B	Al (µg/l)	Cd (µg/l)	pH (pH-units)	TON (mg/l)
104	Grwyne Fawr	2013	20	-	-	-	-	8.22	0.39
105	Honddu at Capel	2013	-	15	-	-	-	8.34	0.49
109	Ll7	2013	18	-	-	23.2	0.02	6.86	0.12
108	Ll6	2013	20	20	-	-	-	6.97	0.06
102	Gl1	2013	20	-	21	4.0	0.00	6.59	0.24
12	Cl1	2012	-	20	-	-	-	-	-
97	Cl1	2013	-	17	19	-	-	-	-
98	Cl4	2013	-	20	-	66.7	-	6.63	0.28
112	Nant Dar	2013	20	20	19	65.5	0.05	6.37	0.28
46	Nant Clawdd	2012	5	5	-	203.0	0.05	6.88	0.29
96	Brefi	2013	19	-	18	66.5	0.05	6.21	0.12
9	Brefi	2012	20	-	11	110.5	0.05	6.52	0.07
106	Ithon at Llandewi	2013	-	20	20	-	-	-	0.76
116	Nant Peiran	2013	20	-	-	212.5	0.15	5.15	0.15
114	Nant Glan dwr	2013	-	20	-	35.0	0.10	6.46	0.11
113	Nant Gelli Gethin	2013	20	-	18	-	-	-	-
10	Cerist (Afon)	2012	-	20	-	39.5	0.05	6.62	0.06
115	Nant Helygog	2013	-	20	19	49.5	0.05	6.68	0.09
94	Afon Fechan	2013	5	-	20	44.5	0.05	7.01	0.18
4	Afon Fechan	2012	1	-	-	31.5	0.15	7.04	0.08
95	Afon Pistyll	2013	22	-	-	138.0	0.15	4.75	0.01
6	Afon Pistyll	2012	21	-	-	148.5	0.05	4.60	0.02
59	Upper Llugwy	2012	20	-	-	-	-	-	0.11
55	Nant y Gwryd	2012	8	-	-	54.0	0.05	6.22	0.02
118	Nant y Gwryd	2013	-	20	21	34.5	0.10	6.06	0.09
93	Afon Colwyn	2013	19	20		91.5	0.20	5.33	0.06

Table H2a-c. Allelic richness per site, per loci, using the full set of loci for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. Adjusted for differences in sample size, based on a minimum sample size of a) 11, b) 14, c) 8 for per sites, and a) 16, b) 17, c) 160 for clusters for *A. sulcicollis*, *I. grammatica* and *B. rhodani*, respectively. For comparison with Table H4(a-c) which used a reduced set of loci.

Table H2a. *Amphinemura sulcicollis*.

Loci	Site													Cluster			
	104	109	108	102	112	96	9	116	113	95	6	59	93	1	2	3	4
Amp_2	13.2	11.5	11.1	11.8	15.3	12.7	13.3	13.2	12.0	10.6	8.3	12.1	11.3	15.6	15.5	12.0	14.0
Amp_3	8.0	6.0	8.1	4.6	6.4	8.1	5.5	3.7	6.2	6.3	5.2	6.9	6.8	9.7	7.2	6.6	8.0
Amp_4	6.6	7.0	6.2	5.9	6.5	5.0	7.4	7.6	7.3	6.6	7.5	7.3	5.6	6.9	7.6	8.0	6.0
Amp_5	10.6	10.2	11.6	10.2	10.4	8.6	9.6	8.5	11.2	7.7	5.5	9.3	7.2	12.6	11.4	8.7	8.0
Amp_6	4.3	6.3	4.4	3.9	4.2	3.7	4.6	4.8	5.3	4.7	3.6	4.8	4.9	4.9	5.1	5.8	5.0
Amp_8	9.5	8.0	8.4	8.2	7.6	8.3	8.0	8.1	7.9	7.5	8.7	10.1	7.5	10.8	9.7	9.2	8.0
Amp_9	8.0	6.6	5.3	5.8	6.8	6.3	5.3	4.8	6.9	7.6	6.1	9.0	4.9	9.7	8.6	8.5	6.0
Amp_10	8.7	9.1	11.0	8.5	10.0	5.4	7.1	11.0	9.5	7.3	5.6	6.8	8.8	10.6	11.2	7.0	10.0
Amp_11	12.6	10.6	12.8	9.8	12.4	8.3	10.8	9.8	9.5	8.7	8.2	13.6	9.9	14.8	16.8	9.6	12.0
Amp_13	8.9	7.2	6.9	5.3	7.2	7.3	6.0	6.6	7.4	6.0	4.5	7.2	5.4	9.9	8.0	6.9	6.0
Mean	**9.0**	**8.3**	**8.6**	**7.4**	**8.7**	**7.4**	**7.8**	**7.8**	**8.3**	**7.3**	**6.3**	**8.7**	**7.2**	**10.6**	**10.1**	**8.2**	**8.3**
SE	**2.7**	**2.0**	**2.9**	**2.7**	**3.3**	**2.5**	**2.8**	**3.0**	**2.2**	**1.6**	**1.7**	**2.7**	**2.2**	**3.2**	**3.7**	**1.8**	**2.9**

Table H2b. *Isoperla grammatica*.

Loci	Site												Cluster	
	105	106	108	12	97	98	112	114	115	10	118	93	1	2
Iso_1	12.53	13.00	14.70	13.30	13.85	14.92	15.11	15.08	16.51	12.81	11.20	10.03	15.97	11.25
Iso_2	20.05	18.80	19.43	20.02	17.22	19.59	18.75	16.23	20.23	19.47	19.88	17.88	23.27	20.40
Iso_3	16.39	18.10	12.31	16.93	17.59	16.31	17.01	15.28	18.28	12.68	17.96	13.56	18.23	14.71
Iso_4	15.46	17.01	22.29	16.29	18.69	19.30	19.63	17.54	18.33	20.56	19.92	21.07	22.37	24.67

Loci	Site												Cluster	
	105	106	108	12	97	98	112	114	115	10	118	93	1	2
Iso_5	15.52	17.50	19.46	16.37	15.70	17.32	17.30	15.30	14.99	16.95	13.38	14.83	21.18	16.81
Iso_6	15.33	13.23	13.28	14.86	14.45	16.87	13.53	15.75	20.37	17.83	14.19	15.71	19.86	17.79
Iso_7	17.26	16.22	21.25	16.10	20.06	15.67	17.42	16.92	17.20	14.40	17.31	16.08	21.14	17.91
Iso_8	18.32	17.90	18.12	17.36	18.17	16.10	18.13	17.05	19.00	18.39	18.60	14.01	22.26	15.32
Iso_9	17.00	16.39	14.36	20.02	16.58	16.54	18.33	19.75	15.55	19.88	16.50	13.03	24.38	14.00
Iso_10	9.73	13.12	8.29	7.99	11.25	9.88	10.58	9.82	9.68	10.73	9.41	9.23	12.65	10.34
Mean	**15.76**	**16.13**	**16.35**	**15.92**	**16.36**	**16.25**	**16.58**	**15.87**	**17.01**	**16.37**	**15.83**	**14.54**	**20.13**	**16.32**
SE	**2.91**	**2.21**	**4.46**	**3.46**	**2.61**	**2.68**	**2.76**	**2.55**	**3.15**	**3.46**	**3.64**	**3.49**	**3.60**	**4.24**

Table H2c. *Baetis rhodani*.

Loci	Site										Cluster
	112	102	97	96	9	106	113	115	94	118	1
Brh-1	7.61	5.27	4.48	5.73	5.71	6.37	5.28	6.76	5.73	6.75	13.00
Brh-2	4.42	5.91	3.74	5.18	5.58	5.39	5.62	5.62	6.41	5.11	14.97
Brh-3	4.79	6.54	4.99	5.19	7.13	5.62	2.89	6.09	7.40	4.72	17.85
Brh-4	7.42	7.73	6.87	8.67	7.97	5.52	6.16	9.36	6.67	8.40	24.76
Brh-5	7.21	9.04	7.85	8.17	7.35	6.80	7.74	7.50	6.92	7.73	24.83
Brh-6	6.20	5.38	5.01	8.38	9.00	7.91	7.40	6.90	8.37	6.96	24.83
Brh-7	11.40	12.69	13.50	12.86	13.92	12.02	12.51	13.32	11.95	12.18	52.60
B_1	5.65	5.98	6.03	6.01	5.58	3.88	5.99	5.49	3.94	4.73	17.89
B_2	7.58	8.31	9.17	7.84	8.13	7.30	8.46	7.93	8.35	7.65	18.89
B_3	7.69	8.45	9.98	10.53	9.22	8.44	8.50	8.97	9.58	8.28	22.89
B_4	4.90	5.25	4.26	4.46	4.18	3.98	4.40	4.26	3.26	4.62	10.89
B_5	6.46	7.09	5.99	7.47	6.97	6.97	9.50	9.23	6.88	7.81	37.57
B_7	6.63	6.76	5.65	6.41	5.12	4.58	4.87	4.44	3.98	5.43	21.75
Mean	**6.76**	**7.26**	**6.73**	**7.45**	**7.37**	**6.52**	**6.87**	**7.38**	**6.88**	**6.95**	**23.28**
SE	**1.80**	**2.07**	**2.77**	**2.36**	**2.49**	**2.18**	**2.52**	**2.46**	**2.41**	**2.12**	**11.10**

Table H3a-c. Complete results from allelic richness comparisons between sites and clusters, with the full set of loci, for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. A linear mixed model (controlling for different loci) compared each site against each other, and each cluster against each other, in the program RStudio. The library 'nlme' was used to generate p-values. The model was re-levelled so each separate site was run as the reference site, therefore comparing every site to each other (please see R script in Appendix J). For comparison with Appendix I Tables I7a-c, which show results from reduced loci data set.

Table H3a. *Amphinemura sulcicollis*.

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter[#]
104	(Intercept)	9.031	0.799	108	11.304	0.000	***
	109	-0.774	0.536	108	-1.444	0.152	
	108	-0.443	0.536	108	-0.826	0.411	
	102	-1.629	0.536	108	-3.038	0.003	**
	112	-0.344	0.536	108	-0.641	0.523	
	96	-1.669	0.536	108	-3.112	0.002	**
	9	-1.254	0.536	108	-2.339	0.021	*
	116	-1.218	0.536	108	-2.271	0.025	*
	113	-0.728	0.536	108	-1.358	0.177	
	95	-1.724	0.536	108	-3.215	0.002	**
	6	-2.712	0.536	108	-5.058	0.000	***
	59	-0.309	0.536	108	-0.576	0.566	
	93	-1.796	0.536	108	-3.350	0.001	**
109	(Intercept)	8.257	0.799	108	10.335	0.000	***

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter[#]
	104	0.774	0.536	108	1.444	0.152	
	108	0.331	0.536	108	0.618	0.538	
	102	-0.855	0.536	108	-1.594	0.114	
	112	0.430	0.536	108	0.803	0.424	
	96	-0.895	0.536	108	-1.669	0.098	
	9	-0.480	0.536	108	-0.895	0.373	
	116	-0.444	0.536	108	-0.827	0.410	
	113	0.046	0.536	108	0.085	0.932	
	95	-0.950	0.536	108	-1.771	0.079	
	6	-1.938	0.536	108	-3.614	0.001	***
	59	0.465	0.536	108	0.868	0.387	
	93	-1.022	0.536	108	-1.906	0.059	
108	(Intercept)	8.588	0.799	108	10.750	0.000	***
	104	0.443	0.536	108	0.826	0.411	
	109	-0.331	0.536	108	-0.618	0.538	
	102	-1.186	0.536	108	-2.212	0.029	*
	112	0.099	0.536	108	0.184	0.854	
	96	-1.226	0.536	108	-2.287	0.024	*
	9	-0.811	0.536	108	-1.513	0.133	
	116	-0.775	0.536	108	-1.445	0.151	
	113	-0.286	0.536	108	-0.533	0.595	
	95	-1.281	0.536	108	-2.389	0.019	*
	6	-2.269	0.536	108	-4.232	0.000	***
	59	0.134	0.536	108	0.250	0.803	
	93	-1.353	0.536	108	-2.524	0.013	*
102	(Intercept)	7.402	0.799	108	9.265	0.000	***
	104	1.629	0.536	108	3.038	0.003	**
	109	0.855	0.536	108	1.594	0.114	
	108	1.186	0.536	108	2.212	0.029	*
	112	1.285	0.536	108	2.396	0.018	*
	96	-0.040	0.536	108	-0.075	0.941	
	9	0.375	0.536	108	0.699	0.486	
	116	0.411	0.536	108	0.767	0.445	
	113	0.900	0.536	108	1.679	0.096	
	95	-0.095	0.536	108	-0.177	0.860	
	6	-1.083	0.536	108	-2.020	0.046	*
	59	1.320	0.536	108	2.462	0.015	*
	93	-0.167	0.536	108	-0.312	0.756	
112	(Intercept)	8.687	0.799	108	10.874	0.000	***
	104	0.344	0.536	108	0.641	0.523	
	109	-0.430	0.536	108	-0.803	0.424	
	108	-0.099	0.536	108	-0.184	0.854	
	102	-1.285	0.536	108	-2.396	0.018	*
	96	-1.325	0.536	108	-2.471	0.015	*
	9	-0.910	0.536	108	-1.697	0.093	
	116	-0.874	0.536	108	-1.630	0.106	
	113	-0.385	0.536	108	-0.717	0.475	
	95	-1.380	0.536	108	-2.574	0.011	*
	6	-2.368	0.536	108	-4.416	0.000	***

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
	59	0.035	0.536	108	0.065	0.948	
	93	-1.452	0.536	108	-2.708	0.008	**
96	(Intercept)	7.362	0.799	108	9.215	0.000	***
	104	1.669	0.536	108	3.112	0.002	**
	109	0.895	0.536	108	1.669	0.098	
	108	1.226	0.536	108	2.287	0.024	*
	102	0.040	0.536	108	0.075	0.941	
	112	1.325	0.536	108	2.471	0.015	*
	9	0.415	0.536	108	0.774	0.441	
	116	0.451	0.536	108	0.841	0.402	
	113	0.941	0.536	108	1.754	0.082	
	95	-0.055	0.536	108	-0.103	0.919	
	6	-1.043	0.536	108	-1.945	0.054	
	59	1.360	0.536	108	2.536	0.013	*
	93	-0.127	0.536	108	-0.237	0.813	
9	(Intercept)	7.777	0.799	108	9.735	0.000	***
	104	1.254	0.536	108	2.339	0.021	*
	109	0.480	0.536	108	0.895	0.373	
	108	0.811	0.536	108	1.513	0.133	
	102	-0.375	0.536	108	-0.699	0.486	
	112	0.910	0.536	108	1.697	0.093	
	96	-0.415	0.536	108	-0.774	0.441	
	116	0.036	0.536	108	0.068	0.946	
	113	0.526	0.536	108	0.980	0.329	
	95	-0.470	0.536	108	-0.876	0.383	
	6	-1.458	0.536	108	-2.719	0.008	**
	59	0.945	0.536	108	1.763	0.081	
	93	-0.542	0.536	108	-1.011	0.314	
116	(Intercept)	7.813	0.799	108	9.780	0.000	***
	104	1.218	0.536	108	2.271	0.025	*
	109	0.444	0.536	108	0.827	0.410	
	108	0.775	0.536	108	1.445	0.151	
	102	-0.411	0.536	108	-0.767	0.445	
	112	0.874	0.536	108	1.630	0.106	
	96	-0.451	0.536	108	-0.841	0.402	
	9	-0.036	0.536	108	-0.068	0.946	
	113	0.489	0.536	108	0.913	0.364	
	95	-0.506	0.536	108	-0.944	0.347	
	6	-1.494	0.536	108	-2.787	0.006	**
	59	0.909	0.536	108	1.695	0.093	
	93	-0.578	0.536	108	-1.079	0.283	
113	(Intercept)	8.302	0.799	108	10.393	0.000	***
	104	0.728	0.536	108	1.358	0.177	
	109	-0.046	0.536	108	-0.085	0.932	
	108	0.286	0.536	108	0.533	0.595	
	102	-0.900	0.536	108	-1.679	0.096	
	112	0.385	0.536	108	0.717	0.475	
	96	-0.941	0.536	108	-1.754	0.082	
	9	-0.526	0.536	108	-0.980	0.329	

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter[#]
	116	-0.489	0.536	108	-0.913	0.364	
	95	-0.996	0.536	108	-1.857	0.066	
	6	-1.984	0.536	108	-3.699	0.000	***
	59	0.420	0.536	108	0.782	0.436	
	93	-1.068	0.536	108	-1.991	0.049	*
95	(Intercept)	7.307	0.799	108	9.146	0.000	***
	104	1.724	0.536	108	3.215	0.002	**
	109	0.950	0.536	108	1.771	0.079	
	108	1.281	0.536	108	2.389	0.019	*
	102	0.095	0.536	108	0.177	0.860	
	112	1.380	0.536	108	2.574	0.011	*
	96	0.055	0.536	108	0.103	0.919	
	9	0.470	0.536	108	0.876	0.383	
	116	0.506	0.536	108	0.944	0.347	
	113	0.996	0.536	108	1.857	0.066	
	6	-0.988	0.536	108	-1.843	0.068	
	59	1.415	0.536	108	2.639	0.010	**
	93	-0.072	0.536	108	-0.135	0.893	
6	(Intercept)	6.319	0.799	108	7.910	0.000	***
	104	2.712	0.536	108	5.058	0.000	***
	109	1.938	0.536	108	3.614	0.001	***
	108	2.269	0.536	108	4.232	0.000	***
	102	1.083	0.536	108	2.020	0.046	*
	112	2.368	0.536	108	4.416	0.000	***
	96	1.043	0.536	108	1.945	0.054	
	9	1.458	0.536	108	2.719	0.008	**
	116	1.494	0.536	108	2.787	0.006	**
	113	1.984	0.536	108	3.699	0.000	***
	95	0.988	0.536	108	1.843	0.068	
	59	2.403	0.536	108	4.482	0.000	***
	93	0.916	0.536	108	1.708	0.091	
59	(Intercept)	8.722	0.799	108	10.918	0.000	***
	104	0.309	0.536	108	0.576	0.566	
	109	-0.465	0.536	108	-0.868	0.387	
	108	-0.134	0.536	108	-0.250	0.803	
	102	-1.320	0.536	108	-2.462	0.015	*
	112	-0.035	0.536	108	-0.065	0.948	
	96	-1.360	0.536	108	-2.536	0.013	*
	9	-0.945	0.536	108	-1.763	0.081	
	116	-0.909	0.536	108	-1.695	0.093	
	113	-0.420	0.536	108	-0.782	0.436	
	95	-1.415	0.536	108	-2.639	0.010	**
	6	-2.403	0.536	108	-4.482	0.000	***
	93	-1.487	0.536	108	-2.774	0.007	**
93	(Intercept)	7.235	0.799	108	9.056	0.000	***
	104	1.796	0.536	108	3.350	0.001	**
	109	1.022	0.536	108	1.906	0.059	
	108	1.353	0.536	108	2.524	0.013	*
	102	0.167	0.536	108	0.312	0.756	

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
	112	1.452	0.536	108	2.708	0.008	**
	96	0.127	0.536	108	0.237	0.813	
	9	0.542	0.536	108	1.011	0.314	
	116	0.578	0.536	108	1.079	0.283	
	113	1.068	0.536	108	1.991	0.049	*
	95	0.072	0.536	108	0.135	0.893	
	6	-0.916	0.536	108	-1.708	0.091	
	59	1.487	0.536	108	2.774	0.007	**
cluster1	(Intercept)	10.550	0.948	27	11.126	0.000	***
	cluster2	-0.440	0.594	27	-0.741	0.465	
	cluster3	-2.320	0.594	27	-3.909	0.001	***
	cluster4	-2.250	0.594	27	-3.791	0.001	***
cluster2	(Intercept)	10.110	0.948	27	10.662	0.000	***
	cluster1	0.440	0.594	27	0.741	0.465	
	cluster3	-1.880	0.594	27	-3.167	0.004	**
	cluster4	-1.810	0.594	27	-3.049	0.005	**
cluster3	(Intercept)	8.230	0.948	27	8.679	0.000	***
	cluster1	2.320	0.594	27	3.909	0.001	***
	cluster2	1.880	0.594	27	3.167	0.004	**
	cluster4	0.070	0.594	27	0.118	0.907	
cluster4	(Intercept)	8.300	0.948	27	8.753	0.000	***
	cluster1	2.250	0.594	27	3.791	0.001	***
	cluster2	1.810	0.594	27	3.049	0.005	**
	cluster3	-0.070	0.594	27	-0.118	0.907	

#Key: * p-value<0.05, ** p-value <0.01, *** p-value <0.001. Blue = positive coefficient value, Red = negative coefficient value.

Table H3b. *Isoperla grammatica*.

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
105	(Intercept)	15.758	1.003	99	15.715	0.000	***
	106	0.369	0.800	99	0.462	0.645	
	108	0.590	0.800	99	0.737	0.463	
	12	0.164	0.800	99	0.205	0.838	
	97	0.598	0.800	99	0.748	0.456	
	98	0.491	0.800	99	0.614	0.541	
	112	0.820	0.800	99	1.026	0.308	
	114	0.113	0.800	99	0.141	0.888	
	115	1.256	0.800	99	1.570	0.120	
	10	0.611	0.800	99	0.764	0.447	
	118	0.077	0.800	99	0.096	0.924	
	93	-1.217	0.800	99	-1.521	0.132	
106	(Intercept)	16.127	1.003	99	16.083	0.000	***
	105	-0.369	0.800	99	-0.462	0.645	

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter[#]
	108	0.221	0.800	99	0.276	0.783	
	12	-0.205	0.800	99	-0.257	0.798	
	97	0.229	0.800	99	0.286	0.775	
	98	0.122	0.800	99	0.152	0.879	
	112	0.451	0.800	99	0.564	0.574	
	114	-0.257	0.800	99	-0.321	0.749	
	115	0.886	0.800	99	1.108	0.271	
	10	0.242	0.800	99	0.303	0.763	
	118	-0.293	0.800	99	-0.366	0.715	
	93	-1.586	0.800	99	-1.982	0.050	
108	(Intercept)	16.348	1.003	99	16.303	0.000	***
	105	-0.590	0.800	99	-0.737	0.463	
	106	-0.221	0.800	99	-0.276	0.783	
	12	-0.426	0.800	99	-0.532	0.596	
	97	0.008	0.800	99	0.011	0.992	
	98	-0.099	0.800	99	-0.123	0.902	
	112	0.231	0.800	99	0.288	0.774	
	114	-0.477	0.800	99	-0.597	0.552	
	115	0.666	0.800	99	0.832	0.407	
	10	0.021	0.800	99	0.027	0.979	
	118	-0.513	0.800	99	-0.642	0.523	
	93	-1.806	0.800	99	-2.258	0.026	*
12	(Intercept)	15.922	1.003	99	15.878	0.000	***
	105	-0.164	0.800	99	-0.205	0.838	
	106	0.205	0.800	99	0.257	0.798	
	108	0.426	0.800	99	0.532	0.596	
	97	0.434	0.800	99	0.543	0.589	
	98	0.327	0.800	99	0.409	0.684	
	112	0.656	0.800	99	0.820	0.414	
	114	-0.052	0.800	99	-0.064	0.949	
	115	1.091	0.800	99	1.364	0.176	
	10	0.447	0.800	99	0.559	0.577	
	118	-0.088	0.800	99	-0.110	0.913	
	93	-1.381	0.800	99	-1.726	0.088	
97	(Intercept)	16.356	1.003	99	16.311	0.000	***
	105	-0.598	0.800	99	-0.748	0.456	
	106	-0.229	0.800	99	-0.286	0.775	
	108	-0.008	0.800	99	-0.011	0.992	
	12	-0.434	0.800	99	-0.543	0.589	
	98	-0.107	0.800	99	-0.134	0.894	
	112	0.222	0.800	99	0.278	0.782	
	114	-0.486	0.800	99	-0.607	0.545	
	115	0.657	0.800	99	0.822	0.413	
	10	0.013	0.800	99	0.016	0.987	
	118	-0.522	0.800	99	-0.652	0.516	
	93	-1.815	0.800	99	-2.269	0.026	*
98	(Intercept)	16.249	1.003	99	16.205	0.000	***
	105	-0.491	0.800	99	-0.614	0.541	
	106	-0.122	0.800	99	-0.152	0.879	

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
	108	0.099	0.800	99	0.123	0.902	
	12	-0.327	0.800	99	-0.409	0.684	
	97	0.107	0.800	99	0.134	0.894	
	112	0.329	0.800	99	0.412	0.682	
	114	-0.379	0.800	99	-0.473	0.637	
	115	0.764	0.800	99	0.955	0.342	
	10	0.120	0.800	99	0.150	0.881	
	118	-0.415	0.800	99	-0.518	0.605	
	93	-1.708	0.800	99	-2.135	0.035	*
112	(Intercept)	16.579	1.003	99	16.533	0.000	***
	105	-0.820	0.800	99	-1.026	0.308	
	106	-0.451	0.800	99	-0.564	0.574	
	108	-0.231	0.800	99	-0.288	0.774	
	12	-0.656	0.800	99	-0.820	0.414	
	97	-0.222	0.800	99	-0.278	0.782	
	98	-0.329	0.800	99	-0.412	0.682	
	114	-0.708	0.800	99	-0.885	0.378	
	115	0.435	0.800	99	0.544	0.588	
	10	-0.209	0.800	99	-0.261	0.794	
	118	-0.744	0.800	99	-0.930	0.355	
	93	-2.037	0.800	99	-2.546	0.012	*
114	(Intercept)	15.871	1.003	99	15.827	0.000	***
	105	-0.113	0.800	99	-0.141	0.888	
	106	0.257	0.800	99	0.321	0.749	
	108	0.477	0.800	99	0.597	0.552	
	12	0.052	0.800	99	0.064	0.949	
	97	0.486	0.800	99	0.607	0.545	
	98	0.379	0.800	99	0.473	0.637	
	112	0.708	0.800	99	0.885	0.378	
	115	1.143	0.800	99	1.429	0.156	
	10	0.499	0.800	99	0.623	0.534	
	118	-0.036	0.800	99	-0.045	0.964	
	93	-1.329	0.800	99	-1.662	0.100	
115	(Intercept)	17.014	1.003	99	16.967	0.000	***
	105	-1.256	0.800	99	-1.570	0.120	
	106	-0.886	0.800	99	-1.108	0.271	
	108	-0.666	0.800	99	-0.832	0.407	
	12	-1.091	0.800	99	-1.364	0.176	
	97	-0.657	0.800	99	-0.822	0.413	
	98	-0.764	0.800	99	-0.955	0.342	
	112	-0.435	0.800	99	-0.544	0.588	
	114	-1.143	0.800	99	-1.429	0.156	
	10	-0.644	0.800	99	-0.805	0.423	
	118	-1.179	0.800	99	-1.474	0.144	
	93	-2.472	0.800	99	-3.090	0.003	**
10	(Intercept)	16.369	1.003	99	16.324	0.000	***
	105	-0.611	0.800	99	-0.764	0.447	
	106	-0.242	0.800	99	-0.303	0.763	
	108	-0.021	0.800	99	-0.027	0.979	

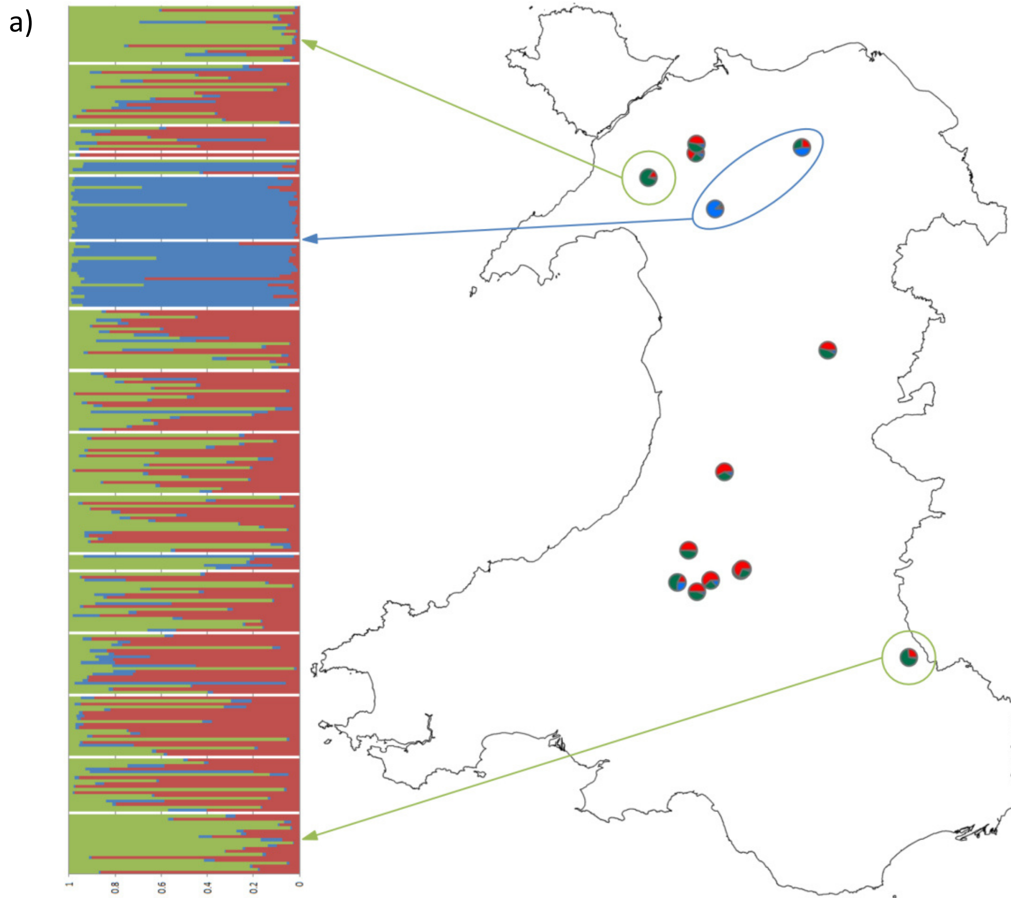
Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
	12	-0.447	0.800	99	-0.559	0.577	
	97	-0.013	0.800	99	-0.016	0.987	
	98	-0.120	0.800	99	-0.150	0.881	
	112	0.209	0.800	99	0.261	0.794	
	114	-0.499	0.800	99	-0.623	0.534	
	115	0.644	0.800	99	0.805	0.423	
	118	-0.535	0.800	99	-0.669	0.505	
	93	-1.828	0.800	99	-2.285	0.024	*
118	(Intercept)	15.835	1.003	99	15.791	0.000	***
	105	-0.077	0.800	99	-0.096	0.924	
	106	0.293	0.800	99	0.366	0.715	
	108	0.513	0.800	99	0.642	0.523	
	12	0.088	0.800	99	0.110	0.913	
	97	0.522	0.800	99	0.652	0.516	
	98	0.415	0.800	99	0.518	0.605	
	112	0.744	0.800	99	0.930	0.355	
	114	0.036	0.800	99	0.045	0.964	
	115	1.179	0.800	99	1.474	0.144	
	10	0.535	0.800	99	0.669	0.505	
	93	-1.293	0.800	99	-1.616	0.109	
93	(Intercept)	14.542	1.003	99	14.502	0.000	***
	105	1.217	0.800	99	1.521	0.132	
	106	1.586	0.800	99	1.982	0.050	
	108	1.806	0.800	99	2.258	0.026	*
	12	1.381	0.800	99	1.726	0.088	
	97	1.815	0.800	99	2.269	0.026	*
	98	1.708	0.800	99	2.135	0.035	*
	112	2.037	0.800	99	2.546	0.012	*
	114	1.329	0.800	99	1.662	0.100	
	115	2.472	0.800	99	3.090	0.003	**
	10	1.828	0.800	99	2.285	0.024	*
	118	1.293	0.800	99	1.616	0.109	
cluster1	(Intercept)	20.130	1.243	9	16.190	0.000	***
	cluster2	-3.810	1.043	9	-3.651	0.005	**
cluster2	(Intercept)	16.320	1.243	9	13.126	0.000	***
	cluster1	3.810	1.043	9	3.651	0.005	**

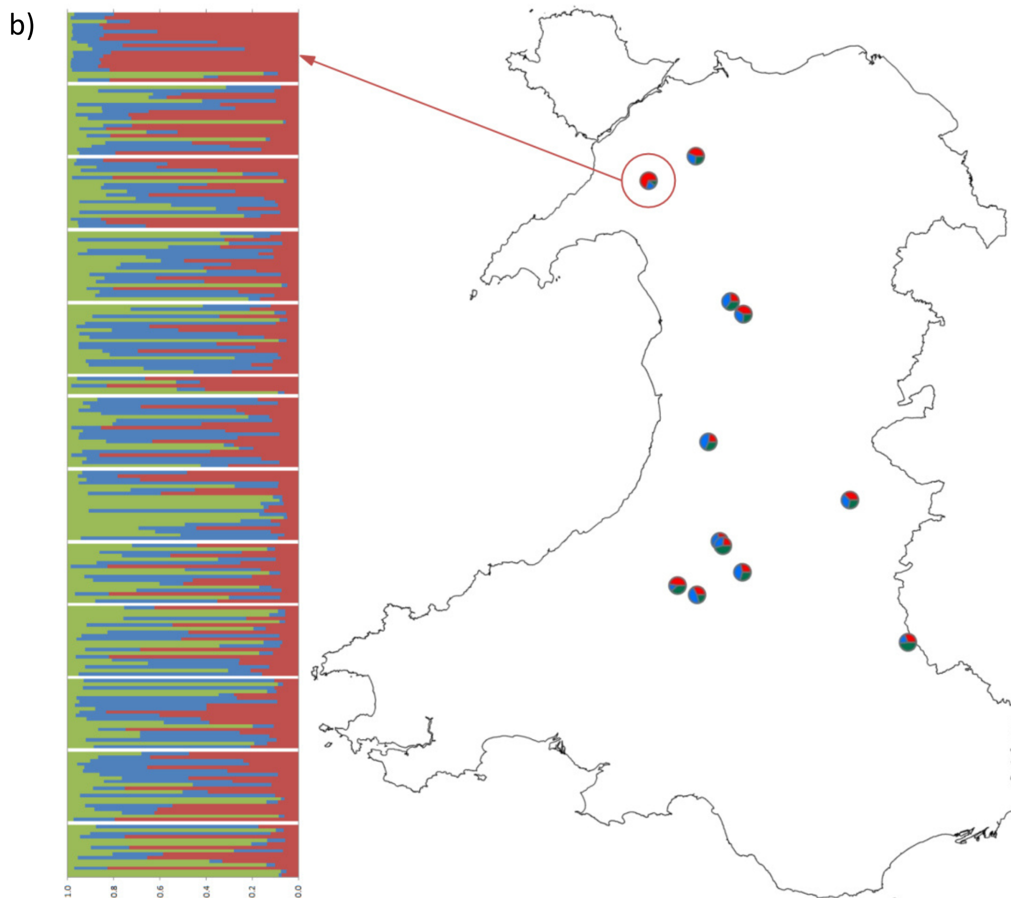
Table H3c. *Baetis rhodani*.

Reference site	Site	Coefficient Value	Std.Error	DF	t-value	p-value	Represented in chapter#
112	(Intercept)	6.764	0.647	108	10.449	0.000	***
	102	0.496	0.363	108	1.366	0.175	
	97	-0.034	0.363	108	-0.093	0.926	
	96	0.697	0.363	108	1.920	0.058	
	9	0.609	0.363	108	1.678	0.096	
	106	-0.244	0.363	108	-0.672	0.503	

Reference site	Site	Coefficient Value	Std.Error	DF	t-value	p-value	Represented in chapter[#]
	113	0.107	0.363	108	0.294	0.770	
	115	0.611	0.363	108	1.684	0.095	
	94	0.116	0.363	108	0.319	0.751	
	118	0.186	0.363	108	0.511	0.610	
102	(Intercept)	7.260	0.647	108	11.216	0.000	***
	112	-0.496	0.363	108	-1.366	0.175	
	97	-0.530	0.363	108	-1.459	0.148	
	96	0.201	0.363	108	0.554	0.581	
	9	0.113	0.363	108	0.312	0.756	
	106	-0.740	0.363	108	-2.039	0.044	*
	113	-0.389	0.363	108	-1.073	0.286	
	115	0.115	0.363	108	0.317	0.752	
	94	-0.380	0.363	108	-1.048	0.297	
	118	-0.310	0.363	108	-0.855	0.395	
97	(Intercept)	6.731	0.647	108	10.397	0.000	***
	112	0.034	0.363	108	0.093	0.926	
	102	0.530	0.363	108	1.459	0.148	
	96	0.731	0.363	108	2.013	0.047	*
	9	0.643	0.363	108	1.770	0.080	
	106	-0.211	0.363	108	-0.580	0.563	
	113	0.140	0.363	108	0.386	0.700	
	115	0.645	0.363	108	1.776	0.079	
	94	0.149	0.363	108	0.411	0.682	
	118	0.219	0.363	108	0.604	0.547	
96	(Intercept)	7.462	0.647	108	11.526	0.000	***
	112	-0.697	0.363	108	-1.920	0.058	
	102	-0.201	0.363	108	-0.554	0.581	
	97	-0.731	0.363	108	-2.013	0.047	*
	9	-0.088	0.363	108	-0.242	0.809	
	106	-0.941	0.363	108	-2.593	0.011	*
	113	-0.591	0.363	108	-1.627	0.107	
	115	-0.086	0.363	108	-0.237	0.813	
	94	-0.581	0.363	108	-1.602	0.112	
	118	-0.512	0.363	108	-1.409	0.162	
9	(Intercept)	7.374	0.647	108	11.390	0.000	***
	112	-0.609	0.363	108	-1.678	0.096	
	102	-0.113	0.363	108	-0.312	0.756	
	97	-0.643	0.363	108	-1.770	0.080	
	96	0.088	0.363	108	0.242	0.809	
	106	-0.853	0.363	108	-2.350	0.021	*
	113	-0.503	0.363	108	-1.384	0.169	
	115	0.002	0.363	108	0.006	0.995	
	94	-0.493	0.363	108	-1.359	0.177	
	118	-0.424	0.363	108	-1.167	0.246	
106	(Intercept)	6.520	0.647	108	10.072	0.000	***
	112	0.244	0.363	108	0.672	0.503	
	102	0.740	0.363	108	2.039	0.044	*
	97	0.211	0.363	108	0.580	0.563	
	96	0.941	0.363	108	2.593	0.011	*

Reference site	Site	Coefficient Value	Std.Error	DF	t-value	p-value	Represented in chapter#
	9	0.853	0.363	108	2.350	0.021	*
	113	0.351	0.363	108	0.966	0.336	
	115	0.855	0.363	108	2.356	0.020	*
	94	0.360	0.363	108	0.991	0.324	
	118	0.430	0.363	108	1.184	0.239	
113	(Intercept)	6.871	0.647	108	10.614	0.000	***
	112	-0.107	0.363	108	-0.294	0.770	
	102	0.389	0.363	108	1.073	0.286	
	97	-0.140	0.363	108	-0.386	0.700	
	96	0.591	0.363	108	1.627	0.107	
	9	0.503	0.363	108	1.384	0.169	
	106	-0.351	0.363	108	-0.966	0.336	
	115	0.505	0.363	108	1.390	0.167	
	94	0.009	0.363	108	0.025	0.980	
	118	0.079	0.363	108	0.218	0.828	
115	(Intercept)	7.376	0.647	108	11.394	0.000	***
	112	-0.611	0.363	108	-1.684	0.095	
	102	-0.115	0.363	108	-0.317	0.752	
	97	-0.645	0.363	108	-1.776	0.079	
	96	0.086	0.363	108	0.237	0.813	
	9	-0.002	0.363	108	-0.006	0.995	
	106	-0.855	0.363	108	-2.356	0.020	*
	113	-0.505	0.363	108	-1.390	0.167	
	94	-0.496	0.363	108	-1.365	0.175	
	118	-0.426	0.363	108	-1.172	0.244	
94	(Intercept)	6.880	0.647	108	10.628	0.000	***
	112	-0.116	0.363	108	-0.319	0.751	
	102	0.380	0.363	108	1.048	0.297	
	97	-0.149	0.363	108	-0.411	0.682	
	96	0.581	0.363	108	1.602	0.112	
	9	0.493	0.363	108	1.359	0.177	
	106	-0.360	0.363	108	-0.991	0.324	
	113	-0.009	0.363	108	-0.025	0.980	
	115	0.496	0.363	108	1.365	0.175	
	118	0.070	0.363	108	0.193	0.848	
118	(Intercept)	6.950	0.647	108	10.736	0.000	***
	112	-0.186	0.363	108	-0.511	0.610	
	102	0.310	0.363	108	0.855	0.395	
	97	-0.219	0.363	108	-0.604	0.547	
	96	0.512	0.363	108	1.409	0.162	
	9	0.424	0.363	108	1.167	0.246	
	106	-0.430	0.363	108	-1.184	0.239	
	113	-0.079	0.363	108	-0.218	0.828	
	115	0.426	0.363	108	1.172	0.244	
	94	-0.070	0.363	108	-0.193	0.848	


a)
1
0.8
0.6
0.4
0.2
0


b)
1.0
0.8
0.6
0.4
0.2
0.0

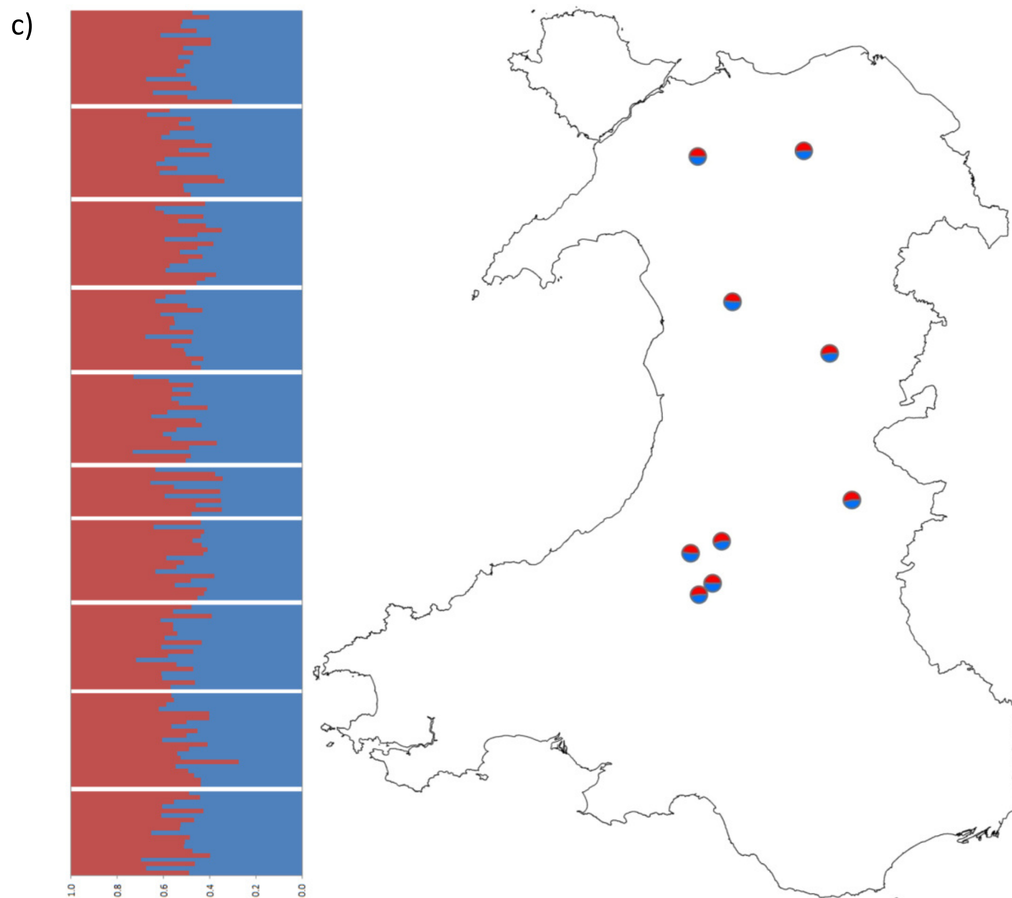


Figure H1. Showing STRUCTURE results over the area of Wales, UK where sampling for each species was completed, a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*.

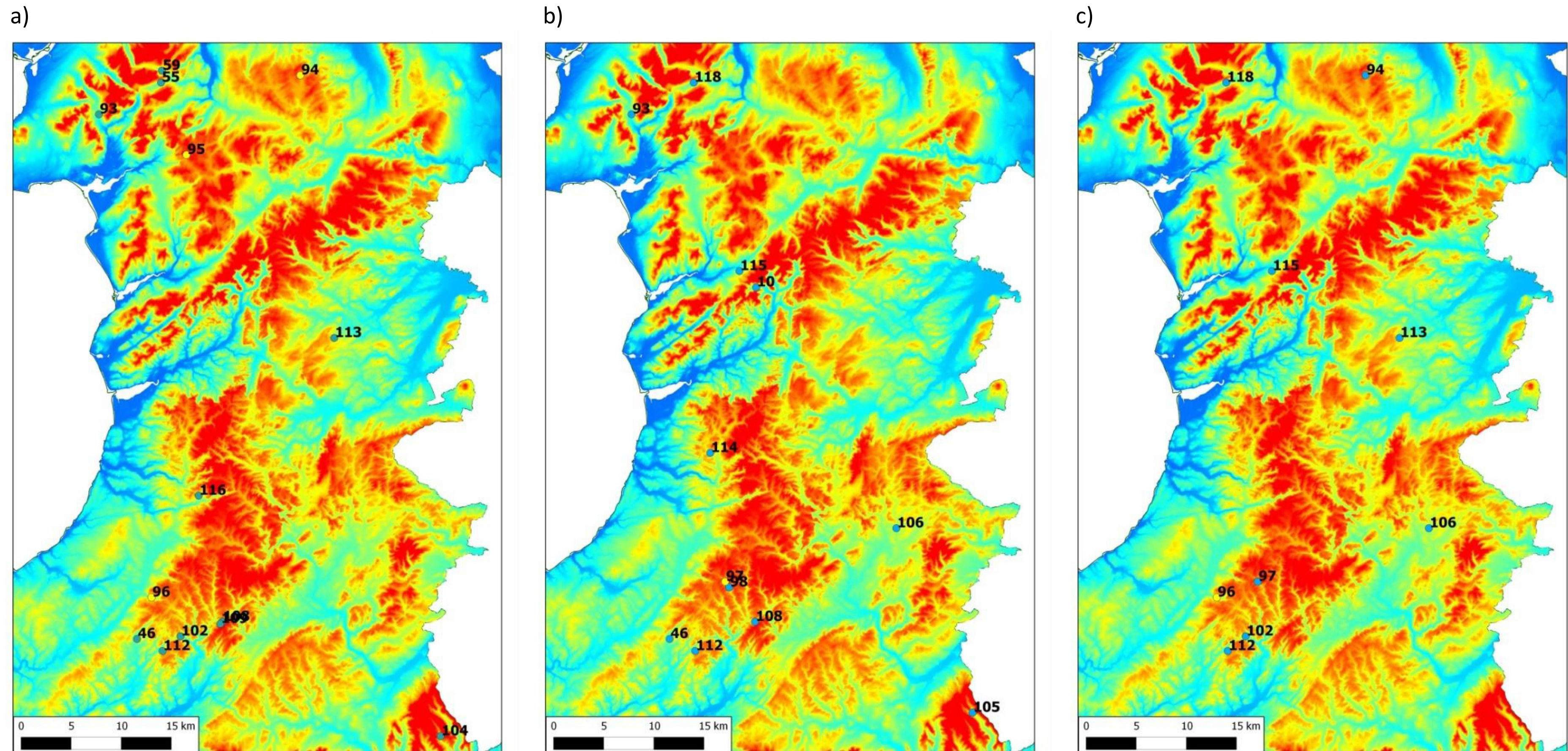


Figure H2. Showing elevation differences over the area of Wales, UK where sampling for each species was completed, a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. Scale of blue to red illustrates low altitude to high altitude.

Appendix I - Chapter 3 Reduced Loci

This appendix shows results using a reduced set of loci for each species repeating Tables 3.1 through to 3.5 and Appendix H, Tables H2 and H3 (i.e. Hardy-Weinberg equilibrium (HWE), descriptive statistics, FST values, allelic richness values and comparisons). Loci that were consistently out of HWE (in more than 50% of sites analysed) were removed, for *Amphinemura sulcicollis*; this left seven loci (Amp_2, Amp_3, Amp_4, Amp_5, Amp_6, Amp_8, Amp_9). For *Isoperla grammatica* there were only three loci (Iso_1, Iso_2 and Iso_3), and *Baetis rhodani* had nine loci (Brh-4, Brh-5, Brh-6, Brh-7, B_1, B_2, B_3, B_4, B_7).

Table of tables

Table I1a-c......484

Table I2a-c......487

Table I3a-b......489

Table I4a-c......490

Table I5a-c......492

Table I6a-b......493

Table I7a-c......494

Table I1a-c. Microsatellite diversity in a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*, using a reduced set of loci. *Mean* (across loci) and standard error (*SE*) values calculated per site. *N* = number of individuals; *Na* = number of alleles per locus; *Ho* = observed heterozygosity; *uHe* = expected heterozygosity; all calculated in Genalex. *HWE* = p value of Hardy-Weinberg Equilibrium test calculated per site using Genepop, *FIS* = the inbreeding co-efficient and *p* = the FIS p value. **Bold** represents significance (critical P value for HWE after Bonferroni correction = ≤0.0071, ≤0.0167, ≤0.0056, for *A. sulcicollis*, *I. grammatica* and *B. rhodani*, respectively). For comparison with Table 3.1a-c which used the full set of loci.

Table I1a. *Amphinemura sulcicollis*.

Site		N	Na	Ho	uHe	HWE	FIS	p
104 **(Cluster1)**	Mean	17.0	10.3	0.756	0.823	**0.000**	0.083	**0.020**
	SE	0.0	1.4	0.098	0.039			
109	Mean	16.0	9.1	0.777	0.831	0.014	0.068	0.053
	SE	0.0	1.1	0.084	0.032			
108	Mean	16.0	9.3	0.759	0.779	0.600	0.026	0.313
	SE	0.0	1.3	0.073	0.047			
102	Mean	17.0	8.1	0.706	0.789	**0.001**	0.108	**0.011**
	SE	0.0	1.4	0.059	0.046			
112	Mean	17.0	10.0	0.731	0.758	0.018	0.037	0.236
	SE	0.0	2.0	0.081	0.059			
96	Mean	15.0	8.6	0.724	0.801	**0.002**	0.099	**0.022**
	SE	0.0	1.4	0.064	0.035			
9	Mean	18.0	9.3	0.690	0.787	**0.004**	0.126	**0.001**
	SE	0.0	1.7	0.078	0.052			
116	Mean	19.0	9.0	0.707	0.731	0.266	0.033	0.253
	SE	0.0	1.8	0.099	0.081			
113	Mean	16.0	9.3	0.732	0.812	**0.000**	0.101	**0.014**
	SE	0.0	1.3	0.074	0.041			
95	Mean	16.0	8.4	0.759	0.796	**0.006**	0.047	0.176
	SE	0.0	1.0	0.071	0.028			
6	Mean	17.0	7.6	0.655	0.703	**0.002**	0.069	0.107
	SE	0.0	0.8	0.061	0.057			
59	Mean	18.0	10.4	0.730	0.809	**0.004**	0.100	**0.014**
	SE	0.0	1.1	0.093	0.042			
93 **(Cluster4)**	Mean	16.0	7.9	0.661	0.745	**0.003**	0.117	**0.005**
	SE	0.0	1.1	0.075	0.061			
Cluster2	Mean	165.0	16.7	0.722	0.801	**0.000**	0.091	**0.000**
	SE	0.0	3.2	0.061	0.039			
Cluster3	Mean	39.0	12.7	0.711	0.800	**0.000**	0.082	**0.002**
	SE	0.0	1.6	0.068	0.045			

Table I1b. *Isoperla grammatica*.

Site		N	Na	Ho	uHe	HWE	FIS	p
105	Mean	15.0	17.0	0.822	0.943	**0.000**	0.132	**0.005**
	SE	0.0	2.3	0.044	0.018			
106	Mean	20.0	20.3	0.783	0.939	**0.000**	0.170	**0.000**
	SE	0.0	2.2	0.017	0.027			
108	Mean	20.0	18.7	0.933	0.938	0.658	0.006	0.505
	SE	0.0	3.3	0.033	0.013			
12	Mean	20.0	20.3	0.883	0.947	**0.000**	0.068	**0.030**
	SE	0.0	2.6	0.073	0.019			
97	Mean	17.0	18.0	0.843	0.948	**0.000**	0.114	**0.007**
	SE	0.0	1.5	0.098	0.005			
98	Mean	20.0	20.7	0.883	0.952	**0.000**	0.074	**0.019**
	SE	0.0	1.8	0.073	0.011			
112	Mean	20.0	20.7	0.883	0.956	**0.000**	0.078	**0.019**
	SE	0.0	1.5	0.017	0.006			
114	Mean	19.0	18.7	0.860	0.932	**0.009**	0.080	**0.024**
	SE	0.0	0.7	0.046	0.005			
115	Mean	20.0	22.7	0.900	0.960	0.044	0.064	**0.025**
	SE	0.0	1.5	0.029	0.008			
10	Mean	20.0	18.3	0.817	0.929	**0.000**	0.123	**0.001**
	SE	0.0	3.3	0.044	0.019			
118	Mean	20.0	20.0	0.867	0.939	0.024	0.079	**0.020**
	SE	0.0	3.6	0.044	0.029			
93 **(Cluster2)**	Mean	18.7	16.3	0.805	0.868	0.049	0.059	0.148
	SE	0.3	3.0	0.087	0.059			
Cluster 1	Mean	216.0	50.3	0.863	0.948	**0.000**	0.090	**0.000**
	SE	0.0	5.6	0.013	0.013			

Table I1c. *Baetis rhodani*.

Site		N	Na	Ho	uHe	HWE	FIS	p
112	Mean	16.0	9.7	0.681	0.830	**0.000**	0.184	**0.000**
	SE	0.0	1.0	0.059	0.027			
102	Mean	20.0	11.8	0.706	0.825	**0.000**	0.149	**0.000**
	SE	0.0	1.4	0.045	0.034			
97	Mean	16.0	11.0	0.757	0.809	0.028	0.067	**0.032**
	SE	0.0	1.8	0.067	0.045			
96	Mean	17.0	12.0	0.719	0.837	**0.000**	0.144	**0.000**
	SE	0.0	1.5	0.057	0.039			
9	Mean	10.6	9.1	0.709	0.824	**0.001**	0.111	**0.011**
	SE	0.3	1.3	0.048	0.041			
106	Mean	17.9	9.6	0.625	0.790	**0.000**	0.207	**0.000**
	SE	0.1	1.7	0.062	0.038			
113	Mean	16.0	10.1	0.750	0.821	**0.000**	0.089	**0.016**
	SE	0.0	1.5	0.051	0.031			
115	Mean	15.8	10.9	0.762	0.813	0.027	0.053	0.121
	SE	0.2	1.7	0.062	0.042			
94	Mean	15.9	9.7	0.689	0.792	**0.000**	0.125	**0.001**

Site		N	Na	Ho	uHe	HWE	FIS	p
	SE	0.1	1.7	0.075	0.046			
118	Mean	19.9	10.7	0.733	0.816	**0.000**	0.098	**0.001**
	SE	0.1	1.4	0.058	0.037			
Cluster 1	Mean	165.0	24.6	0.712	0.823	**0.000**	0.129	**0.000**
	SE	0.4	3.9	0.048	0.035			

Table I2a-c. Pairwise genetic differentiation (FST, distance method) of a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*, between sites using a reduced set of loci. Values with a p value of <0.05 in **bold**. For comparison with Table 3.2a-c which used the full set of loci.

Table I2a. *Amphinemura sulcicollis*.

	104	109	108	102	112	96	9	116	113	95	6	59	93
104	0												
109	-0.001	0											
108	0.010	0.010	0										
102	0.007	0.006	**0.032**	0									
112	0.003	0.012	0.005	**0.030**	0								
96	-0.001	0.000	0.015	0.011	0.007	0							
9	**0.012**	0.011	0.015	**0.027**	0.008	0.009	0						
116	0.022	**0.023**	**0.037**	**0.028**	**0.026**	**0.025**	0.007	0					
113	**-0.010**	0.000	0.002	0.012	-0.004	-0.003	0.004	**0.022**	0				
95	**0.052**	**0.049**	**0.057**	**0.050**	**0.068**	**0.038**	**0.051**	**0.078**	**0.039**	0			
6	**0.081**	**0.071**	**0.066**	**0.071**	**0.078**	**0.058**	**0.052**	**0.085**	**0.062**	0.006	0		
59	**0.008**	0.009	**0.032**	0.012	0.014	-0.001	**0.017**	**0.023**	0.004	**0.038**	**0.067**	0	
93	0.019	**0.035**	**0.045**	**0.037**	**0.021**	**0.024**	**0.020**	**0.016**	0.014	**0.065**	**0.079**	**0.017**	0

Table I2b. *Isoperla grammatica*.

	105	106	108	12	97	98	112	114	115	10	118	93
105	0											
106	0.006	0										
108	0.002	0.010	0									
12	0.001	-0.005	0.002	0								
97	0.004	0.013	0.009	**0.013**	0							
98	0.001	0.007	-0.001	0.003	0.006	0						

	105	106	108	12	97	98	112	114	115	10	118	93
112	0.007	0.009	0.006	0.008	0.000	-0.002	0					
114	0.001	0.014	0.008	0.009	0.003	0.004	**0.012**	0				
115	0.003	0.004	0.005	0.003	-0.002	0.001	-0.001	**0.013**	0			
10	0.009	0.006	0.005	0.005	**0.019**	0.004	0.008	**0.020**	0.005	0		
118	-0.001	-0.001	0.004	-0.001	0.006	0.006	0.004	**0.014**	-0.001	0.006	0	
93	**0.023**	**0.024**	**0.022**	**0.021**	**0.046**	**0.026**	**0.041**	**0.045**	**0.029**	**0.021**	**0.016**	0

Table I2c. *Baetis rhodani*.

	112	102	97	96	9	106	113	115	94	118
112	0									
102	-0.003	0								
97	0.004	0.006	0							
96	0.000	0.002	-0.001	0						
9	0.007	0.004	0.004	0.002	0					
106	**0.025**	**0.021**	**0.020**	0.014	**0.025**	0				
113	0.017	**0.016**	0.014	0.005	0.005	-0.006	0			
115	0.001	0.001	-0.001	-0.002	-0.006	**0.021**	0.012	0		
94	**0.031**	**0.021**	**0.031**	**0.021**	0.021	-0.004	-0.001	**0.027**	0	
118	0.003	0.001	0.008	-0.006	0.012	**0.020**	**0.013**	0.009	**0.026**	0

Table I3a-b. Pairwise genetic differentiation (FST, distance method) of a) *Amphinemura sulcicollis* and b) *Isoperla grammatica* between clusters, using a reduced set of loci. Values with a p value of <0.05 in **bold**. For comparison with Table 3.3a-b which used the full set of loci.

Table I3a. *Amphinemura sulcicollis*.

	cluster1	cluster2	cluster3	cluster4
cluster1	0			
cluster2	0.001	0		
cluster3	**0.057**	**0.042**	0	
cluster4	**0.019**	**0.017**	**0.057**	0

Table I3b. *Isoperla grammatical*.

	cluster1	cluster2
cluster1	0	
cluster2	**0.02353**	0

Table I4a-c. Allelic richness per site, per loci, using a reduced set of loci for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. Adjusted for differences in sample size, based on a minimum sample size of a) 15, b) 15, c) 8 for per sites, and a) 16, b) 18, c) 163 for clusters, for *A. sulcicollis*, *I. grammatica* and *B. rhodani*, respectively. For comparison with Table H1a-c which used the full set of loci.

Table I4a. *Amphinemura sulcicollis*.

Loci	Site													Cluster			
	104	109	108	102	112	96	9	116	113	95	6	59	93	1	2	3	4
Amp_2	15.24	13.56	12.68	13.38	19.20	15.00	16.21	15.80	14.44	12.62	9.51	14.02	13.50	15.64	15.46	12.01	14.00
Amp_3	9.39	6.81	9.68	4.88	7.52	10.00	6.45	4.36	6.87	6.88	5.75	8.16	7.81	9.70	7.23	6.58	8.00
Amp_4	6.88	7.81	6.87	6.00	6.87	5.00	7.81	8.36	7.87	6.94	8.53	8.31	5.94	6.94	7.62	7.98	6.00
Amp_5	12.27	11.68	13.62	11.51	11.60	10.00	10.59	9.86	12.69	9.56	6.53	10.42	7.87	12.64	11.41	8.65	8.00
Amp_6	4.77	6.87	4.88	3.99	4.75	4.00	4.83	4.95	5.87	4.94	3.88	5.50	5.00	4.88	5.12	5.85	5.00
Amp_8	10.61	8.81	9.69	8.85	8.62	9.00	9.16	9.12	8.00	7.94	9.63	11.30	7.93	10.82	9.68	9.22	8.00
Amp_9	9.39	6.94	5.87	6.64	7.72	7.00	5.79	4.95	7.81	8.75	6.75	10.28	5.81	9.70	8.56	8.52	6.00
Mean	**9.79**	**8.93**	**9.04**	**7.89**	**9.47**	**8.57**	**8.69**	**8.20**	**9.08**	**8.23**	**7.23**	**9.71**	**7.69**	**10.05**	**9.29**	**8.40**	**7.86**
SE	**3.43**	**2.67**	**3.34**	**3.50**	**4.76**	**3.69**	**3.86**	**4.02**	**3.19**	**2.44**	**2.11**	**2.71**	**2.82**	**3.54**	**3.36**	**1.99**	**2.97**

Table I4b. *Isoperla grammatica*.

Loci	Site												Cluster	
	105	106	108	12	97	98	112	114	115	10	118	93	1	2
Iso_1	13.00	13.55	15.17	13.80	14.26	15.49	15.64	15.72	17.17	13.24	11.55	10.45	16.45	11.63
Iso_2	21.00	19.60	20.44	20.97	17.87	20.45	19.57	17.00	21.16	20.46	20.83	18.74	24.19	21.21
Iso_3	17.00	18.86	12.64	17.56	18.43	16.99	17.74	15.89	19.15	13.11	18.72	13.99	18.80	15.00
Mean	**17.00**	**17.34**	**16.08**	**17.44**	**16.85**	**17.65**	**17.65**	**16.21**	**19.16**	**15.60**	**17.03**	**14.39**	**19.81**	**15.95**
SE	**4.00**	**3.30**	**3.98**	**3.59**	**2.26**	**2.54**	**1.97**	**0.70**	**1.99**	**4.20**	**4.86**	**4.16**	**3.97**	**4.86**

Table I4c. *Baetis rhodani*.

Loci	Site										Cluster
	112	102	97	96	9	106	113	115	94	118	1
Brh-4	7.42	7.73	6.87	8.67	7.97	5.52	6.16	9.36	6.67	8.40	7.62
Brh-5	7.21	9.04	7.85	8.17	7.35	6.80	7.74	7.50	6.92	7.73	7.64
Brh-6	6.20	5.38	5.01	8.38	9.00	7.91	7.40	6.90	8.37	6.96	7.31
Brh-7	11.40	12.69	13.50	12.86	13.92	12.02	12.51	13.32	11.95	12.18	12.90
B_1	5.65	5.98	6.03	6.01	5.58	3.88	5.99	5.49	3.94	4.73	5.75
B_2	7.58	8.31	9.17	7.84	8.13	7.30	8.46	7.93	8.35	7.65	7.93
B_3	7.69	8.45	9.98	10.53	9.22	8.44	8.50	8.97	9.58	8.28	8.99
B_4	4.90	5.25	4.26	4.46	4.18	3.98	4.40	4.26	3.26	4.62	4.47
B_7	6.63	6.76	5.65	6.41	5.12	4.58	4.87	4.44	3.98	5.43	5.82
Mean	**7.18**	**7.73**	**7.59**	**8.15**	**7.83**	**6.71**	**7.34**	**7.58**	**7.00**	**7.33**	**7.60**
SE	**1.84**	**2.31**	**2.91**	**2.49**	**2.88**	**2.61**	**2.44**	**2.83**	**2.90**	**2.34**	**2.42**

Table I5a-c. Comparing the allelic richness between each site, using a reduced set of loci for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. Only significant differences in allelic richness are presented: * p-value <0.05, ** p-value <0.01, *** p-value <0.001. Blue = positive coefficient value, Red = negative coefficient value. The horizontal axis is the reference site, if the regression co-efficient value was positive (represented here with blue asterisks) this means that site on the vertical axis has higher mean allelic richness than the reference site above. If the value is negative (represented here with red asterisks) the site to the left of the value has a lower mean allelic richness compared to the reference site above. For comparison with Table 3.4a-c which used the full set of loci.

Table I5a. *Amphinemura sulcicollis*.

	104	109	108	102	112	96	9	116	113	95	6	59	93
104				*				*		*	**		**
109											*		
108											*		
102	*				*							*	
112				*							**		*
96													
9													
116	*												
113											*		
95	*												
6	**	*	*		**				*			**	
59				*							**		*
93	**				*							*	

Table I5b. *Isoperla grammatica*.

	105	106	108	12	97	98	112	114	115	10	118	93
105										*		
106												
108									*			
12												*
97												
98												*
112												*
114												
115			*									**
10									*			
118												
93				*		*	*		**			

Table I5c. *Baetis rhodani*.

	112	102	97	96	9	106	113	115	94	118
112				*						
102						*				
97						*				
96	*					***	*		**	*
9						**			*	
106		*	*	***	**			*		
113				*						
115						*				
94				**	*					
118				*						

Table I6a-b. Comparing the allelic richness between each cluster, using a reduced set of loci for a) *Amphinemura sulcicollis* and b) *Isoperla grammatica*. Only significant differences in allelic richness are presented: * p-value <0.05, ** p-value <0.01, *** p-value <0.001. Blue = positive coefficient value, Red = negative coefficient value. The horizontal axis is the reference cluster, if the regression co-efficient value was positive (represented here with blue asterisks) this means that cluster on the vertical axis has higher mean allelic richness than the reference cluster above. If the value is negative (represented here with red asterisks) the cluster to the left of the value has a lower mean allelic richness compared to the reference cluster above. For comparison with Table 3.5a-b which used the full set of loci.

Table I6a. *Amphinemura sulcicollis*.

	cluster1	cluster2	cluster3	cluster4
cluster1			*	**
cluster2				*
cluster3	*			
cluster4	**	*		

Table I6b. *Isoperla grammatica*.

	cluster1	cluster2
cluster1		*
cluster2	*	

Table I7a-c. Complete results from allelic richness comparisons between sites and clusters, with a reduced set of loci, for a) *Amphinemura sulcicollis*, b) *Isoperla grammatica* and c) *Baetis rhodani*. A linear mixed model (controlling for different loci) compared each site against each other, and each cluster against each other, in the program RStudio. The library 'nlme' was used to generate p-values. The model was re-levelled so each separate site was run as the reference site, therefore comparing every site to each other (please see R script in Appendix J). For comparison with Appendix H Table H3a-c, which show results from full loci data set.

Table I7a. *Amphinemura sulcicollis*.

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
104	(Intercept)	9.793	1.265	72	7.742	0.000	***
	109	-0.866	0.782	72	-1.108	0.272	
	108	-0.752	0.782	72	-0.962	0.339	
	102	-1.902	0.782	72	-2.432	0.018	*
	112	-0.324	0.782	72	-0.414	0.680	
	96	-1.222	0.782	72	-1.562	0.123	
	9	-1.101	0.782	72	-1.408	0.163	
	116	-1.592	0.782	72	-2.035	0.046	*
	113	-0.715	0.782	72	-0.915	0.363	
	95	-1.564	0.782	72	-1.999	0.049	*
	6	-2.568	0.782	72	-3.283	0.002	**
	59	-0.079	0.782	72	-0.102	0.919	
	93	-2.100	0.782	72	-2.685	0.009	**
109	(Intercept)	8.927	1.265	72	7.057	0.000	***
	104	0.866	0.782	72	1.108	0.272	
	108	0.114	0.782	72	0.146	0.885	
	102	-1.035	0.782	72	-1.324	0.190	
	112	0.543	0.782	72	0.694	0.490	
	96	-0.355	0.782	72	-0.454	0.651	
	9	-0.235	0.782	72	-0.300	0.765	
	116	-0.725	0.782	72	-0.927	0.357	
	113	0.151	0.782	72	0.193	0.847	
	95	-0.697	0.782	72	-0.891	0.376	
	6	-1.701	0.782	72	-2.175	0.033	*
	59	0.787	0.782	72	1.006	0.318	
	93	-1.233	0.782	72	-1.577	0.119	
108	(Intercept)	9.041	1.265	72	7.147	0.000	***
	104	0.752	0.782	72	0.962	0.339	
	109	-0.114	0.782	72	-0.146	0.885	
	102	-1.149	0.782	72	-1.469	0.146	
	112	0.429	0.782	72	0.548	0.585	
	96	-0.469	0.782	72	-0.600	0.551	
	9	-0.349	0.782	72	-0.446	0.657	
	116	-0.839	0.782	72	-1.073	0.287	
	113	0.037	0.782	72	0.047	0.962	

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
	95	-0.811	0.782	72	-1.037	0.303	
	6	-1.815	0.782	72	-2.321	0.023	*
	59	0.673	0.782	72	0.861	0.392	
	93	-1.347	0.782	72	-1.723	0.089	
102	(Intercept)	7.891	1.265	72	6.239	0.000	***
	104	1.902	0.782	72	2.432	0.018	*
	109	1.035	0.782	72	1.324	0.190	
	108	1.149	0.782	72	1.469	0.146	
	112	1.578	0.782	72	2.018	0.047	*
	96	0.680	0.782	72	0.870	0.387	
	9	0.800	0.782	72	1.023	0.310	
	116	0.310	0.782	72	0.396	0.693	
	113	1.186	0.782	72	1.517	0.134	
	95	0.338	0.782	72	0.432	0.667	
	6	-0.666	0.782	72	-0.851	0.397	
	59	1.822	0.782	72	2.330	0.023	*
	93	-0.198	0.782	72	-0.253	0.801	
112	(Intercept)	9.469	1.265	72	7.486	0.000	***
	104	0.324	0.782	72	0.414	0.680	
	109	-0.543	0.782	72	-0.694	0.490	
	108	-0.429	0.782	72	-0.548	0.585	
	102	-1.578	0.782	72	-2.018	0.047	*
	96	-0.898	0.782	72	-1.148	0.255	
	9	-0.778	0.782	72	-0.994	0.323	
	116	-1.268	0.782	72	-1.621	0.109	
	113	-0.392	0.782	72	-0.501	0.618	
	95	-1.240	0.782	72	-1.585	0.117	
	6	-2.244	0.782	72	-2.869	0.005	**
	59	0.244	0.782	72	0.312	0.756	
	93	-1.776	0.782	72	-2.271	0.026	*
96	(Intercept)	8.571	1.265	72	6.776	0.000	***
	104	1.222	0.782	72	1.562	0.123	
	109	0.355	0.782	72	0.454	0.651	
	108	0.469	0.782	72	0.600	0.551	
	102	-0.680	0.782	72	-0.870	0.387	
	112	0.898	0.782	72	1.148	0.255	
	9	0.120	0.782	72	0.154	0.878	
	116	-0.370	0.782	72	-0.473	0.638	
	113	0.506	0.782	72	0.647	0.520	
	95	-0.342	0.782	72	-0.437	0.663	
	6	-1.346	0.782	72	-1.721	0.090	
	59	1.142	0.782	72	1.460	0.149	
	93	-0.878	0.782	72	-1.123	0.265	
9	(Intercept)	8.692	1.265	72	6.871	0.000	***
	104	1.101	0.782	72	1.408	0.163	
	109	0.235	0.782	72	0.300	0.765	
	108	0.349	0.782	72	0.446	0.657	
	102	-0.800	0.782	72	-1.023	0.310	
	112	0.778	0.782	72	0.994	0.323	

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
	96	-0.120	0.782	72	-0.154	0.878	
	116	-0.490	0.782	72	-0.627	0.533	
	113	0.386	0.782	72	0.493	0.623	
	95	-0.462	0.782	72	-0.591	0.556	
	6	-1.466	0.782	72	-1.875	0.065	
	59	1.022	0.782	72	1.307	0.196	
	93	-0.999	0.782	72	-1.277	0.206	
116	(Intercept)	8.201	1.265	72	6.484	0.000	***
	104	1.592	0.782	72	2.035	0.046	*
	109	0.725	0.782	72	0.927	0.357	
	108	0.839	0.782	72	1.073	0.287	
	102	-0.310	0.782	72	-0.396	0.693	
	112	1.268	0.782	72	1.621	0.109	
	96	0.370	0.782	72	0.473	0.638	
	9	0.490	0.782	72	0.627	0.533	
	113	0.876	0.782	72	1.120	0.266	
	95	0.028	0.782	72	0.036	0.971	
	6	-0.976	0.782	72	-1.248	0.216	
	59	1.512	0.782	72	1.934	0.057	
	93	-0.508	0.782	72	-0.650	0.518	
113	(Intercept)	9.078	1.265	72	7.177	0.000	***
	104	0.715	0.782	72	0.915	0.363	
	109	-0.151	0.782	72	-0.193	0.847	
	108	-0.037	0.782	72	-0.047	0.962	
	102	-1.186	0.782	72	-1.517	0.134	
	112	0.392	0.782	72	0.501	0.618	
	96	-0.506	0.782	72	-0.647	0.520	
	9	-0.386	0.782	72	-0.493	0.623	
	116	-0.876	0.782	72	-1.120	0.266	
	95	-0.848	0.782	72	-1.084	0.282	
	6	-1.852	0.782	72	-2.368	0.021	*
	59	0.636	0.782	72	0.813	0.419	
	93	-1.384	0.782	72	-1.770	0.081	
95	(Intercept)	8.229	1.265	72	6.506	0.000	***
	104	1.564	0.782	72	1.999	0.049	*
	109	0.697	0.782	72	0.891	0.376	
	108	0.811	0.782	72	1.037	0.303	
	102	-0.338	0.782	72	-0.432	0.667	
	112	1.240	0.782	72	1.585	0.117	
	96	0.342	0.782	72	0.437	0.663	
	9	0.462	0.782	72	0.591	0.556	
	116	-0.028	0.782	72	-0.036	0.971	
	113	0.848	0.782	72	1.084	0.282	
	6	-1.004	0.782	72	-1.284	0.203	
	59	1.484	0.782	72	1.898	0.062	
	93	-0.536	0.782	72	-0.686	0.495	
6	(Intercept)	7.225	1.265	72	5.712	0.000	***
	104	2.568	0.782	72	3.283	0.002	**
	109	1.701	0.782	72	2.175	0.033	*

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter[#]
	108	1.815	0.782	72	2.321	0.023	*
	102	0.666	0.782	72	0.851	0.397	
	112	2.244	0.782	72	2.869	0.005	**
	96	1.346	0.782	72	1.721	0.090	
	9	1.466	0.782	72	1.875	0.065	
	116	0.976	0.782	72	1.248	0.216	
	113	1.852	0.782	72	2.368	0.021	*
	95	1.004	0.782	72	1.284	0.203	
	59	2.488	0.782	72	3.181	0.002	**
	93	0.468	0.782	72	0.598	0.552	
59	(Intercept)	9.714	1.265	72	7.679	0.000	***
	104	0.079	0.782	72	0.102	0.919	
	109	-0.787	0.782	72	-1.006	0.318	
	108	-0.673	0.782	72	-0.861	0.392	
	102	-1.822	0.782	72	-2.330	0.023	*
	112	-0.244	0.782	72	-0.312	0.756	
	96	-1.142	0.782	72	-1.460	0.149	
	9	-1.022	0.782	72	-1.307	0.196	
	116	-1.512	0.782	72	-1.934	0.057	
	113	-0.636	0.782	72	-0.813	0.419	
	95	-1.484	0.782	72	-1.898	0.062	
	6	-2.488	0.782	72	-3.181	0.002	**
	93	-2.020	0.782	72	-2.583	0.012	*
93	(Intercept)	7.693	1.265	72	6.082	0.000	***
	104	2.100	0.782	72	2.685	0.009	**
	109	1.233	0.782	72	1.577	0.119	
	108	1.347	0.782	72	1.723	0.089	
	102	0.198	0.782	72	0.253	0.801	
	112	1.776	0.782	72	2.271	0.026	*
	96	0.878	0.782	72	1.123	0.265	
	9	0.999	0.782	72	1.277	0.206	
	116	0.508	0.782	72	0.650	0.518	
	113	1.384	0.782	72	1.770	0.081	
	95	0.536	0.782	72	0.686	0.495	
	6	-0.468	0.782	72	-0.598	0.552	
	59	2.020	0.782	72	2.583	0.012	*
cluster1	(Intercept)	10.029	1.144	18	8.769	0.000	***
	cluster2	-0.729	0.604	18	-1.205	0.244	
	cluster3	-1.629	0.604	18	-2.694	0.015	*
	cluster4	-2.171	0.604	18	-3.592	0.002	**
cluster2	(Intercept)	9.300	1.144	18	8.132	0.000	***
	cluster1	0.729	0.604	18	1.205	0.244	
	cluster3	-0.900	0.604	18	-1.489	0.154	
	cluster4	-1.443	0.604	18	-2.387	0.028	*
cluster3	(Intercept)	8.400	1.144	18	7.345	0.000	***
	cluster1	1.629	0.604	18	2.694	0.015	*
	cluster2	0.900	0.604	18	1.489	0.154	
	cluster4	-0.543	0.604	18	-0.898	0.381	
cluster4	(Intercept)	7.857	1.144	18	6.871	0.000	***

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
	cluster1	2.171	0.604	18	3.592	0.002	**
	cluster2	1.443	0.604	18	2.387	0.028	*
	cluster3	0.543	0.604	18	0.898	0.381	

#Key: * p-value<0.05, ** p-value <0.01, *** p-value <0.001. Blue = positive coefficient value, Red = negative coefficient value.

Table I7b. *Isoperla grammatica*.

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
105	(Intercept)	17.000	1.930	22	8.808	0.000	***
	106	0.337	1.446	22	0.233	0.818	
	108	-0.918	1.446	22	-0.635	0.532	
	12	0.442	1.446	22	0.305	0.763	
	97	-0.149	1.446	22	-0.103	0.919	
	98	0.647	1.446	22	0.447	0.659	
	112	0.648	1.446	22	0.448	0.659	
	114	-0.794	1.446	22	-0.549	0.589	
	115	2.160	1.446	22	1.494	0.149	
	10	-1.397	1.446	22	-0.966	0.344	
	118	0.032	1.446	22	0.022	0.982	
	93	-2.606	1.446	22	-1.802	0.085	
106	(Intercept)	17.337	1.930	22	8.982	0.000	***
	105	-0.337	1.446	22	-0.233	0.818	
	108	-1.255	1.446	22	-0.868	0.395	
	12	0.105	1.446	22	0.072	0.943	
	97	-0.486	1.446	22	-0.336	0.740	
	98	0.310	1.446	22	0.214	0.832	
	112	0.311	1.446	22	0.215	0.832	
	114	-1.131	1.446	22	-0.782	0.443	
	115	1.823	1.446	22	1.261	0.221	
	10	-1.734	1.446	22	-1.199	0.243	
	118	-0.305	1.446	22	-0.211	0.835	
	93	-2.943	1.446	22	-2.036	0.054	
108	(Intercept)	16.082	1.930	22	8.332	0.000	***
	105	0.918	1.446	22	0.635	0.532	
	106	1.255	1.446	22	0.868	0.395	
	12	1.360	1.446	22	0.940	0.357	
	97	0.769	1.446	22	0.532	0.600	
	98	1.565	1.446	22	1.082	0.291	
	112	1.566	1.446	22	1.083	0.291	
	114	0.124	1.446	22	0.086	0.932	
	115	3.078	1.446	22	2.129	0.045	*
	10	-0.479	1.446	22	-0.331	0.744	
	118	0.950	1.446	22	0.657	0.518	

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
	93	-1.688	1.446	22	-1.167	0.256	
12	(Intercept)	17.442	1.930	22	9.037	0.000	***
	105	-0.442	1.446	22	-0.305	0.763	
	106	-0.105	1.446	22	-0.072	0.943	
	108	-1.360	1.446	22	-0.940	0.357	
	97	-0.591	1.446	22	-0.409	0.687	
	98	0.205	1.446	22	0.142	0.889	
	112	0.206	1.446	22	0.142	0.888	
	114	-1.235	1.446	22	-0.854	0.402	
	115	1.719	1.446	22	1.189	0.247	
	10	-1.839	1.446	22	-1.272	0.217	
	118	-0.409	1.446	22	-0.283	0.780	
	93	-3.048	1.446	22	-2.108	0.047	*
97	(Intercept)	16.851	1.930	22	8.730	0.000	***
	105	0.149	1.446	22	0.103	0.919	
	106	0.486	1.446	22	0.336	0.740	
	108	-0.769	1.446	22	-0.532	0.600	
	12	0.591	1.446	22	0.409	0.687	
	98	0.796	1.446	22	0.551	0.588	
	112	0.797	1.446	22	0.551	0.587	
	114	-0.644	1.446	22	-0.446	0.660	
	115	2.310	1.446	22	1.597	0.124	
	10	-1.248	1.446	22	-0.863	0.398	
	118	0.182	1.446	22	0.126	0.901	
	93	-2.457	1.446	22	-1.699	0.103	
98	(Intercept)	17.647	1.930	22	9.143	0.000	***
	105	-0.647	1.446	22	-0.447	0.659	
	106	-0.310	1.446	22	-0.214	0.832	
	108	-1.565	1.446	22	-1.082	0.291	
	12	-0.205	1.446	22	-0.142	0.889	
	97	-0.796	1.446	22	-0.551	0.588	
	112	0.001	1.446	22	0.001	1.000	
	114	-1.440	1.446	22	-0.996	0.330	
	115	1.514	1.446	22	1.047	0.307	
	10	-2.044	1.446	22	-1.413	0.172	
	118	-0.614	1.446	22	-0.425	0.675	
	93	-3.253	1.446	22	-2.250	0.035	*
112	(Intercept)	17.648	1.930	22	9.143	0.000	***
	105	-0.648	1.446	22	-0.448	0.659	
	106	-0.311	1.446	22	-0.215	0.832	
	108	-1.566	1.446	22	-1.083	0.291	
	12	-0.206	1.446	22	-0.142	0.888	
	97	-0.797	1.446	22	-0.551	0.587	
	98	-0.001	1.446	22	-0.001	1.000	
	114	-1.441	1.446	22	-0.997	0.330	
	115	1.513	1.446	22	1.046	0.307	
	10	-2.045	1.446	22	-1.414	0.171	
	118	-0.615	1.446	22	-0.426	0.675	
	93	-3.254	1.446	22	-2.250	0.035	*

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
114	(Intercept)	16.206	1.930	22	8.396	0.000	***
	105	0.794	1.446	22	0.549	0.589	
	106	1.131	1.446	22	0.782	0.443	
	108	-0.124	1.446	22	-0.086	0.932	
	12	1.235	1.446	22	0.854	0.402	
	97	0.644	1.446	22	0.446	0.660	
	98	1.440	1.446	22	0.996	0.330	
	112	1.441	1.446	22	0.997	0.330	
	115	2.954	1.446	22	2.043	0.053	
	10	-0.603	1.446	22	-0.417	0.681	
	118	0.826	1.446	22	0.571	0.574	
	93	-1.812	1.446	22	-1.253	0.223	
115	(Intercept)	19.160	1.930	22	9.927	0.000	***
	105	-2.160	1.446	22	-1.494	0.149	
	106	-1.823	1.446	22	-1.261	0.221	
	108	-3.078	1.446	22	-2.129	0.045	*
	12	-1.719	1.446	22	-1.189	0.247	
	97	-2.310	1.446	22	-1.597	0.124	
	98	-1.514	1.446	22	-1.047	0.307	
	112	-1.513	1.446	22	-1.046	0.307	
	114	-2.954	1.446	22	-2.043	0.053	
	10	-3.557	1.446	22	-2.460	0.022	*
	118	-2.128	1.446	22	-1.472	0.155	
	93	-4.766	1.446	22	-3.297	0.003	**
10	(Intercept)	15.603	1.930	22	8.084	0.000	***
	105	1.397	1.446	22	0.966	0.344	*
	106	1.734	1.446	22	1.199	0.243	
	108	0.479	1.446	22	0.331	0.744	
	12	1.839	1.446	22	1.272	0.217	
	97	1.248	1.446	22	0.863	0.398	
	98	2.044	1.446	22	1.413	0.172	
	112	2.045	1.446	22	1.414	0.171	
	114	0.603	1.446	22	0.417	0.681	
	115	3.557	1.446	22	2.460	0.022	
	118	1.429	1.446	22	0.989	0.334	
	93	-1.209	1.446	22	-0.836	0.412	
118	(Intercept)	17.032	1.930	22	8.824	0.000	***
	105	-0.032	1.446	22	-0.022	0.982	
	106	0.305	1.446	22	0.211	0.835	
	108	-0.950	1.446	22	-0.657	0.518	
	12	0.409	1.446	22	0.283	0.780	
	97	-0.182	1.446	22	-0.126	0.901	
	98	0.614	1.446	22	0.425	0.675	
	112	0.615	1.446	22	0.426	0.675	
	114	-0.826	1.446	22	-0.571	0.574	
	115	2.128	1.446	22	1.472	0.155	
	10	-1.429	1.446	22	-0.989	0.334	
	93	-2.638	1.446	22	-1.825	0.082	
93	(Intercept)	14.394	1.930	22	7.458	0.000	***

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
	105	2.606	1.446	22	1.802	0.085	
	106	2.943	1.446	22	2.036	0.054	
	108	1.688	1.446	22	1.167	0.256	
	12	3.048	1.446	22	2.108	0.047	*
	97	2.457	1.446	22	1.699	0.103	
	98	3.253	1.446	22	2.250	0.035	*
	112	3.254	1.446	22	2.250	0.035	*
	114	1.812	1.446	22	1.253	0.223	
	115	4.766	1.446	22	3.297	0.003	**
	10	1.209	1.446	22	0.836	0.412	
	118	2.638	1.446	22	1.825	0.082	
cluster1	(Intercept)	19.813	2.561	2	7.735	0.016	*
	cluster2	-3.867	0.532	2	-7.265	0.018	*
cluster2	(Intercept)	15.947	2.561	2	6.226	0.025	*
	cluster1	3.867	0.532	2	7.265	0.018	*

#Key: * p-value<0.05, ** p-value <0.01, *** p-value <0.001. Blue = positive coefficient value, Red = negative coefficient value.

Table I7c. *Baetis rhodani*.

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
112	(Intercept)	7.187	0.858	72	8.373	0.000	***
	102	0.546	0.394	72	1.385	0.171	
	97	0.404	0.394	72	1.026	0.308	
	96	0.961	0.394	72	2.439	0.017	*
	9	0.643	0.394	72	1.633	0.107	
	106	-0.472	0.394	72	-1.198	0.235	
	113	0.150	0.394	72	0.381	0.705	
	115	0.388	0.394	72	0.984	0.328	
	94	-0.184	0.394	72	-0.468	0.641	
	118	0.144	0.394	72	0.367	0.715	
102	(Intercept)	7.732	0.858	72	9.009	0.000	***
	112	-0.546	0.394	72	-1.385	0.171	
	97	-0.141	0.394	72	-0.358	0.721	
	96	0.416	0.394	72	1.055	0.295	
	9	0.098	0.394	72	0.248	0.805	
	106	-1.018	0.394	72	-2.583	0.012	*
	113	-0.396	0.394	72	-1.004	0.319	
	115	-0.158	0.394	72	-0.400	0.690	
	94	-0.730	0.394	72	-1.853	0.068	
	118	-0.401	0.394	72	-1.018	0.312	
97	(Intercept)	7.591	0.858	72	8.844	0.000	***
	112	-0.404	0.394	72	-1.026	0.308	
	102	0.141	0.394	72	0.358	0.721	

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter#
	96	0.557	0.394	72	1.413	0.162	
	9	0.239	0.394	72	0.606	0.546	
	106	-0.877	0.394	72	-2.225	0.029	*
	113	-0.254	0.394	72	-0.646	0.521	
	115	-0.017	0.394	72	-0.042	0.966	
	94	-0.589	0.394	72	-1.494	0.139	
	118	-0.260	0.394	72	-0.660	0.512	
96	(Intercept)	8.148	0.858	72	9.493	0.000	***
	112	-0.961	0.394	72	-2.439	0.017	*
	102	-0.416	0.394	72	-1.055	0.295	
	97	-0.557	0.394	72	-1.413	0.162	
	9	-0.318	0.394	72	-0.806	0.423	
	106	-1.433	0.394	72	-3.638	0.001	***
	113	-0.811	0.394	72	-2.058	0.043	*
	115	-0.573	0.394	72	-1.455	0.150	
	94	-1.146	0.394	72	-2.907	0.005	**
	118	-0.817	0.394	72	-2.073	0.042	*
9	(Intercept)	7.830	0.858	72	9.123	0.000	***
	112	-0.643	0.394	72	-1.633	0.107	
	102	-0.098	0.394	72	-0.248	0.805	
	97	-0.239	0.394	72	-0.606	0.546	
	96	0.318	0.394	72	0.806	0.423	
	106	-1.116	0.394	72	-2.831	0.006	**
	113	-0.493	0.394	72	-1.252	0.215	
	115	-0.256	0.394	72	-0.649	0.519	
	94	-0.828	0.394	72	-2.101	0.039	*
	118	-0.499	0.394	72	-1.266	0.210	
106	(Intercept)	6.714	0.858	72	7.823	0.000	***
	112	0.472	0.394	72	1.198	0.235	
	102	1.018	0.394	72	2.583	0.012	*
	97	0.877	0.394	72	2.225	0.029	*
	96	1.433	0.394	72	3.638	0.001	***
	9	1.116	0.394	72	2.831	0.006	**
	113	0.622	0.394	72	1.579	0.119	
	115	0.860	0.394	72	2.183	0.032	*
	94	0.288	0.394	72	0.730	0.468	
	118	0.617	0.394	72	1.565	0.122	
113	(Intercept)	7.337	0.858	72	8.548	0.000	***
	112	-0.150	0.394	72	-0.381	0.705	
	102	0.396	0.394	72	1.004	0.319	
	97	0.254	0.394	72	0.646	0.521	
	96	0.811	0.394	72	2.058	0.043	*
	9	0.493	0.394	72	1.252	0.215	
	106	-0.622	0.394	72	-1.579	0.119	
	115	0.238	0.394	72	0.603	0.548	
	94	-0.334	0.394	72	-0.849	0.399	
	118	-0.006	0.394	72	-0.014	0.989	
115	(Intercept)	7.574	0.858	72	8.825	0.000	***
	112	-0.388	0.394	72	-0.984	0.328	

Reference site	Site	Coefficient value	Std. error	DF	t-value	p-value	Represented in chapter[#]
	102	0.158	0.394	72	0.400	0.690	
	97	0.017	0.394	72	0.042	0.966	
	96	0.573	0.394	72	1.455	0.150	
	9	0.256	0.394	72	0.649	0.519	
	106	-0.860	0.394	72	-2.183	0.032	*
	113	-0.238	0.394	72	-0.603	0.548	
	94	-0.572	0.394	72	-1.452	0.151	
	118	-0.243	0.394	72	-0.618	0.539	
94	(Intercept)	7.002	0.858	72	8.158	0.000	***
	112	0.184	0.394	72	0.468	0.641	
	102	0.730	0.394	72	1.853	0.068	
	97	0.589	0.394	72	1.494	0.139	
	96	1.146	0.394	72	2.907	0.005	**
	9	0.828	0.394	72	2.101	0.039	*
	106	-0.288	0.394	72	-0.730	0.468	
	113	0.334	0.394	72	0.849	0.399	
	115	0.572	0.394	72	1.452	0.151	
	118	0.329	0.394	72	0.835	0.407	
118	(Intercept)	7.331	0.858	72	8.541	0.000	***
	112	-0.144	0.394	72	-0.367	0.715	
	102	0.401	0.394	72	1.018	0.312	
	97	0.260	0.394	72	0.660	0.512	
	96	0.817	0.394	72	2.073	0.042	*
	9	0.499	0.394	72	1.266	0.210	
	106	-0.617	0.394	72	-1.565	0.122	
	113	0.006	0.394	72	0.014	0.989	
	115	0.243	0.394	72	0.618	0.539	
	94	-0.329	0.394	72	-0.835	0.407	

[#]Key: * p-value<0.05, ** p-value <0.01, *** p-value <0.001. Blue = positive coefficient value, Red = negative coefficient value.

Appendix J - R Script

Contains example R script used in Chapters 3 and 4 so that all analysis could be repeated.

1. Comparing allelic richness between sites (Chapter 3)

Key: `Script in Red`, instructions in black

#Data Input

```
dframe1<-read.csv(file.choose(), header=T) # Choose file

names(dframe1)

str(dframe1) #loci and site are categorical therefore we
change them into factors
dframe1$site<-as.factor(dframe1$site)
dframe1$loci<-as.factor(dframe1$loci)
```

#let's check the structure again...

```
str(dframe1) #ahhhhh. That's better...
```

#Model

```
install.packages("nlme")
library(nlme)
model3 <- lme(Allelic.Richness ~ site, random = ~ 1|loci, data
= dframe1, na.action = na.exclude)
summary(model3)
```

#Re-level.

#The output will give you how site differs with the site that comes first numerically. If you #want to look at how the second site differs with everything else, you can re-level like this:

#E.g. site 2.

```
dframe1$site<-relevel(dframe1$site, ref="2")
model3 <- lme(Allelic.Richness ~ site, random = ~ 1|loci, data
= dframe1, na.action = na.exclude)
summary(model3)
```

#Repeat for all subsequent sites so that every site is the reference site

2. Correlating genetic diversity with environmental stressors (Chapter3)

Example using A. sulcicollis:

#Import data

```
amp<-  read.csv(file.choose())      # Amp.GDvschem

names(amp)
```

#Correlation between genetic diversity estimate, allelic richness, and pH

```
amp_ARpH<- lm(amp$a.allelic.rich ~ amp$pH)
plot(amp$pH, amp$a.allelic.rich)
abline(amp_ARpH)
summary(amp_ARpH)
```

#Graph for main thesis

```
plot(amp$pH, amp$a.allelic.rich, xlab="pH", ylab="Allelic
richness", main = "amp", pch = 21, cex.axis = 1.5,
cex.lab=1.5, cex= 2, ylim= c(6.7, 9), xlim= c(4.5, 8.5), bg =
"darkorchid1", col = "black")
abline(amp_ARpH, col ="black", lty=1, lwd = 2)
```

3. Making species indices (Chapter 4)

```
setwd("C:/Users/sbicg5/Desktop/Hannah") # set working directory

library(FD)
library(vegan)
library(ade4)
```

 # Invertebrate data

Loading species-level data (available in Appendix L)

```
spp<-read.table("species_data.txt",h=T,sep="\t")
```

Loading genus-level data (available in Appendix L)

```
gen<-read.table("genus_data.txt",h=T,sep="\t")
```

 # Trait data

Loading genus-level data (available in Appendix L)

```
tr<-read.table("trait_dat.txt",h=T,sep="\t")

# extracting qualitative values
spp[,1]->sites
spp[,2]->year
tr[,3]->gen.names

# removing first columns
spp[,-c(1:2)]->spp
gen[,-c(1:2)]->gen
tr[,-c(1:3)]->tr

# assigning row.names
rownames(tr)<-gen.names

# Transforming fuzzy-coded response traits into percentages
traits.blo<-c(2,3,4,8,4,5,5)

# transform fuzzy codes to percentages
tr<-prep.fuzzy.var(tr,traits.blo)

diversity(spp)->shannon.div # shannon diversity
specnumber(spp)->spp.ric  # species richness

# Functional diversity
gowdis(tr)->tr.dist # functional dissimilarity matrix
fdisp(tr.dist,as.matrix(gen))$FDis->fun.div # functional diversity

div<-data.frame(sites,year,shannon.div,spp.ric,fun.div)

write.table(div,"div.txt",sep="\t",row.names=F)
```

Appendix K - Species Diversity Data

Contains raw species abundance data used to create in species diversity indices analysis within Chapter 4. Table and Figures referred to in main text:

Table of tables

Table K1 509
Table K2 513
Table K3a-c 516
Table K4 524

Table of figures

Figure K1 525

Contributions:

- All species abundance data collection and identification was conducted by DURESS (Diversity of Upland Rivers for Ecosystem Service Sustainability: http://nerc-duress.org). Funded by the Natural Environment Research Council (NERC).
- Species indexes calculated in collaboration with Tano Gutiérrez Cánovas (DURESS post doc)

Table K1. Species abundance data (identified to species level where possible) for all sites used within Chapter 4 (though some taxa were identified at a very coarse taxonomic level e.g. Oligochaeta). The units are invertebrate abundance. *Species* are listed in alphabetical order. Within R script (Appendix J, section 3) this table is named "species_data.txt".

Species	Site name / site code (year)																					
	Cl1/12 (2012)	Cl1/97 (2013)	Cl4/14 (2012)	Cl4/98 (2013)	Gl1/21 (2012)	Gl1/102 (2013)	Ll6/38 (2012)	Ll6/108 (2013)	Ll7/39 (2012)	Ll7/109 (2013)	Afon Fechan/4 (2012)	Cerist (Afon)/10 (2012)	Nant Helygog/115 (2012)	Afon Pistyll/6 (2012)	Afon Colwyn/93 (2012)	Nant Y Gwyrd/55 (2012)	Upper Llugwy/59 (2012)	Nant Glan Dwr/114 (2012)	Nant Peiran/116 (2012)	Brefi/9 (2012)	Nant Clawdd/46 (2012)	Nant Dar/112 (2012)
Agabus_spp	0	0	3	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Agapetus_fuscipes	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Amphinemura_standfussi	0	0	0	0	0	0	2	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
Amphinemura_sulcicollis	0	3	0	0	18	14	0	39	20	16	6	1	1	454	1	2	41	1	18	36	84	157
Ancylus_fluviatilis	0	0	0	0	1	0	0	0	0	0	32	1	0	0	0	0	2	4	0	2	0	0
Atherix_sp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
Baetis_muticus	0	0	0	0	113	42	18	8	60	34	2	2	0	0	0	0	1	0	0	30	3	17
Baetis_rhodani	3	33	2	0	121	29	144	70	183	119	38	81	145	0	0	138	62	38	94	188	92	34
Baetis_scambus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0	0	0
Baetis_spp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	18	0	0	0	0
Brachyptera_risi	1	5	0	1	4	9	5	18	17	3	36	3	4	0	0	0	0	0	46	0	8	24
Caenis_rivulorum	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0
Ceratopogonidae	0	0	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Ceratopogoninae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	0	1
Chaetopteryx_villosa	0	0	0	0	0	0	0	0	0	0	0	0	4	62	1	0	0	4	0	2	1	0
Chironomidae	5	3	22	0	54	4	10	15	19	54	12	171	12	10	13	30	23	208	78	338	64	91
Chloroperla_tripunctata	0	0	0	0	0	1	0	0	5	1	0	9	34	0	0	0	0	0	0	14	3	11
Coleoptera_larvae	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Cordulegaster_boltonii	5	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Species	Site name / site code (year)																				
	Cl1/12 (2012)	Cl1/97 (2013)	Cl4/14 (2012)	Cl4/98 (2013)	Gl1/21 (2012)	Gl1/102 (2013)	Ll6/38 (2012)	Ll6/108 (2013)	Ll7/39 (2012)	Ll7/109 (2013)	Afon Fechan/4 (2012)	Cerist (Afon)/10 (2012)	Nant Helygog/115 (2012)	Afon Pistyll/6 (2012)	Afon Colwyn/93 (2012)	Nant Y Gwyrd/55 (2012)	Upper Llugwy/59 (2012)	Nant Glan Dwr/114 (2012)	Nant Peiran/116 (2012)	Brefi/9 (2012)	Nant Clawdd/46 (2012)	Nant Dar/112 (2012)
Corixidae	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Crunoecia_irrorata	0	0	0	0	12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
Diplectrona_felix	0	0	0	0	55	25	0	1	4	2	0	0	0	0	0	0	0	0	0	2	0	1
Diura_bicaudata	0	0	1	7	0	0	0	1	0	0	26	0	0	0	0	0	0	5	1	0	0	0
Dixidae	0	0	0	0	3	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Drusus_annulatus	1	6	3	1	0	0	3	11	6	9	22	33	1	0	0	0	0	7	0	8	0	0
Dytiscidae	1	1	1	2	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
Ecclisopteryx_guttulata	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
Ecdyonurus	0	0	0	0	0	0	3	1	2	0	0	1	8	0	0	0	23	0	0	56	50	5
Electrogena_lateralis	0	0	0	0	44	23	20	43	6	37	0	14	11	0	19	8	5	1	0	72	10	0
Elmis_aenea	4	22	0	0	1	1	11	10	33	12	80	18	2	0	0	6	4	160	1	66	3	0
Empididae	0	0	0	0	0	0	0	0	0	0	6	5	1	0	4	0	0	2	0	14	10	15
Erpobdella_octoculata	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Esolus_parallelepipedus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0
Gammarus_pulex	0	0	0	0	0	0	0	0	0	0	104	0	0	0	0	0	0	0	0	0	0	0
Glossosoma_boltoni	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Glossosoma_conforme	0	0	0	0	0	0	0	0	12	1	0	0	1	0	0	0	0	0	0	0	0	0
Gyrinidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0
Helophorus_flavipes	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
Heptagenia_sulphurea	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Hydracarina	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
Hydraena_gracilis	0	0	0	0	4	2	0	0	4	0	4	6	29	0	0	0	3	1	0	30	4	13
Hydrophilidae	1	7	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Hydropsyche_pellucidula	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0

Appendix K – Species Diversity Data

Species	Site name / site code (year)																				
	Cl1/12 (2012)	Cl1/97 (2013)	Cl4/14 (2012)	Cl4/98 (2013)	Gl1/21 (2012)	Gl1/102 (2013)	Ll6/38 (2012)	Ll6/108 (2013)	Ll7/39 (2012)	Ll7/109 (2013)	Afon Fechan/4 (2012)	Cerist (Afon)/10 (2012)	Nant Helygog/115 (2012)	Afon Pistyll/6 (2012)	Afon Colwyn/93 (2012)	Nant Y Gwyrd/55 (2012)	Upper Llugwy/59 (2012)	Nant Glan Dwr/114 (2012)	Nant Peiran/116 (2012)	Brefi/9 (2012)	Nant Clawdd/46 (2012)	Nant Dar/112 (2012)
Hydropsyche_siltalai	0	0	1	0	0	0	3	0	2	0	0	25	4	0	0	0	16	7	0	168	20	54
Hydroptilidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0
Isoperla_grammatica	1	3	5	35	55	21	25	54	40	27	2	85	31	0	85	12	6	229	3	80	31	29
Leptophlebia	7	29	68	11	0	0	0	2	2	0	0	0	0	14	0	0	1	5	1	0	0	0
Leuctra	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0
Leuctra_hippopus	0	17	0	2	0	1	3	0	2	0	4	2	0	2	5	0	1	9	37	0	6	0
Leuctra_inermis	3	34	2	12	37	64	63	156	42	37	32	233	39	4	10	14	17	32	13	390	2	289
Leuctra_nigra	0	1	9	1	90	30	7	5	14	2	4	1	0	20	2	0	0	6	0	12	24	3
Limnephilidae	6	30	45	14	18	21	2	2	28	20	14	1	17	0	5	0	4	143	0	6	7	45
Limnius_volckmari	0	0	0	0	0	1	0	0	1	0	0	7	0	0	1	14	2	19	0	18	1	6
Limoniidae	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	4	2	8	2	1	3
Nemoura_spp	0	1	9	0	62	8	1	3	1	1	10	0	0	6	0	0	0	0	0	0	1	17
Nemurella_pictetii	0	1	0	0	0	0	1	0	0	1	0	0	0	42	1	0	0	0	0	0	0	0
Odontocerum_albicorne	0	0	0	0	4	0	1	0	2	2	0	0	0	0	0	0	0	0	0	2	0	0
Oligochaeta	1	7	12	6	83	116	2	9	5	4	16	13	1	0	19	24	25	35	27	20	7	23
Oulimnius_spp	0	0	0	0	0	0	0	0	0	0	2	10	0	0	4	0	0	3	0	38	0	0
Oulimnius_tuberculatus	1	3	9	0	0	0	3	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Paraleptophlebia_submarginata	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48	0
Pediciidae	0	0	0	0	0	0	0	0	0	0	6	14	3	6	0	0	1	13	5	22	3	57
Perla_bipunctata	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
Perlodes_microcephalus	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	4	0	4
Philopotamus_montanus	0	0	0	0	3	0	0	0	0	0	2	0	23	0	0	0	0	0	0	0	21	12
Pisidium_spp	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
Plectrocnemia_conspersa	7	10	21	2	4	1	1	1	5	0	0	0	0	0	0	0	0	0	0	0	0	0

Species	Site name / site code (year)																				
	Cl1/12 (2012)	Cl1/97 (2013)	Cl4/14 (2012)	Cl4/98 (2013)	Gl1/21 (2012)	Gl1/102 (2013)	Ll6/38 (2012)	Ll6/108 (2013)	Ll7/39 (2012)	Ll7/109 (2013)	Afon Fechan/4 (2012)	Cerist (Afon)/10 (2012)	Nant Helygog/115 (2012)	Afon Pistyll/6 (2012)	Afon Colwyn/93 (2012)	Nant Y Gwyrd/55 (2012)	Upper Llugwy/59 (2012)	Nant Glan Dwr/114 (2012)	Nant Peiran/116 (2012)	Brefi/9 (2012)	Nant Clawdd/46 (2012)	Nant Dar/112 (2012)
Plectrocnemia_geniculata	0	14	7	6	4	2	0	3	2	0	0	0	0	0	0	0	0	0	0	0	0	0
Polycentropodidae	0	0	0	0	0	0	0	0	0	0	2	0	2	22	8	2	1	16	9	10	1	4
Protonemura_spp	0	1	0	0	2	15	1	2	5	4	0	1	0	0	0	0	2	0	0	0	1	8
Psychomyiidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Rhithrogena_semicolorata	0	1	0	0	3	20	133	22	139	19	32	34	96	0	0	0	39	0	0	210	76	85
Rhyacophila_dorsalis	0	0	0	0	0	1	1	3	2	0	0	6	3	2	1	0	1	6	10	18	3	7
Rhyacophila_fasciata	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Rhyacophila_munda	0	0	0	0	0	0	0	0	1	0	2	4	0	0	0	0	0	0	5	4	0	0
Scirtidae	3	2	0	0	20	3	4	2	11	1	2	0	2	0	0	0	8	0	4	4	6	103
Sericostoma_personatum	0	0	0	0	1	0	1	0	3	1	0	0	2	0	0	0	17	0	0	14	6	2
Silo_pallipes	0	0	0	0	8	3	5	1	17	2	2	3	6	0	0	0	0	0	0	0	3	10
Simuliidae	0	19	2	0	4	12	1	12	11	9	48	51	95	24	1	0	12	6	253	102	6	29
Siphlonurus_lacustris	8	7	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Siphonoperla_torrentium	1	12	10	26	101	55	16	9	3	3	0	3	0	92	16	8	11	32	13	40	5	24
Tipula_spp	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
Tipulidae	0	11	21	5	11	12	35	8	24	7	0	0	0	0	0	0	0	0	0	0	0	0
Triclad_indeterminate	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Velia_caprai	1	1	2	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Velia_spp	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
Wormaldia_occipitalis	0	0	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	3
Wormaldia_spp	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table K2. Species abundance data (identified to genus level) for all sites used within Chapter 4. The units are invertebrate abundance, *Genus* are listed in alphabetical order. Within R script (Appendix J, section 3) this table is named "genus _data.txt".

Genus	Site name / site code (year)																					
	Cl1/12 (2012)	Cl1/97 (2013)	Cl4/14 (2012)	Cl4/98 (2013)	Gl1/21 (2012)	Gl1/102 (2013)	Ll6/38 (2012)	Ll6/108 (2013)	Ll7/39 (2012)	Ll7/109 (2013)	Afon Fechan/4 (2012)	Cerist (Afon)/10 (2012)	Nant Helygog/115 (2012)	Afon Pistyll/6 (2012)	Afon Colwyn/93 (2012)	Nant Y Gwyrd/55 (2012)	Upper Llugwy/59 (2012)	Nant Glan Dwr/114 (2012)	Nant Peiran/116 (2012)	Brefi/9 (2012)	Nant Clawdd/46 (2012)	Nant Dar/112 (2012)
Agabus	0	0	3	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Agapetus	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Amphinemura	0	3	0	0	18	14	2	39	20	17	6	1	1	454	1	2	41	1	18	36	84	157
Ancylus	0	0	0	0	1	0	0	0	0	0	32	1	0	0	0	0	2	4	0	2	0	0
Atherix	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
Baetis	3	33	2	0	234	71	162	78	243	153	40	83	145	0	0	138	63	66	94	218	95	51
Brachyptera	1	5	0	1	4	9	5	18	17	3	36	3	4	0	0	0	0	0	46	0	8	24
Caenis	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0
Ceratopogonidae	0	0	1	0	3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	4	0	1
Chaetopteryx	0	0	0	0	0	0	0	0	0	0	0	0	4	62	1	0	0	4	0	2	1	0
Chironomidae	5	3	22	0	54	4	10	15	19	54	12	171	12	10	13	30	23	208	78	338	64	91
Chloroperla	0	0	0	0	0	1	0	0	5	1	0	9	34	0	0	0	0	0	0	14	3	11
Cordulegaster	5	2	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Corixidae	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Crunoecia	0	0	0	0	12	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	2
Diplectrona	0	0	0	0	55	25	0	1	4	2	0	0	0	0	0	0	0	0	0	2	0	1
Diura	0	0	1	7	0	0	0	1	0	0	26	0	0	0	0	0	0	5	1	0	0	0
Dixidae	0	0	0	0	3	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Drusus	1	6	3	1	0	0	3	11	6	9	22	33	1	0	0	0	0	7	0	8	0	0
Ecclisopteryx	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
Ecdyonurus	0	0	0	0	0	0	3	1	2	0	0	1	8	0	0	0	23	0	0	56	50	5

Genus	Site name / site code (year)																					
	Cl1/12 (2012)	Cl1/97 (2013)	Cl4/14 (2012)	Cl4/98 (2013)	Gl1/21 (2012)	Gl1/102 (2013)	Ll6/38 (2012)	Ll6/108 (2013)	Ll7/39 (2012)	Ll7/109 (2013)	Afon Fechan/4 (2012)	Cerist (Afon)/10 (2012)	Nant Helygog/115 (2012)	Afon Pistyll/6 (2012)	Afon Colwyn/93 (2012)	Nant Y Gwyrd/55 (2012)	Upper Llugwy/59 (2012)	Nant Glan Dwr/114 (2012)	Nant Peiran/116 (2012)	Brefi/9 (2012)	Nant Clawdd/46 (2012)	Nant Dar/112 (2012)
Electrogena	0	0	0	0	44	23	20	43	6	37	0	14	11	0	19	8	5	1	0	72	10	0
Elmis	4	22	0	0	1	1	11	10	33	12	80	18	2	0	0	6	4	160	1	66	3	0
Empididae	0	0	0	0	0	0	0	0	0	0	6	5	1	0	4	0	0	2	0	14	10	15
Erpobdella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Esolus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0
Gammarus	0	0	0	0	0	0	0	0	0	0	104	0	0	0	0	0	0	0	0	0	0	0
Glossosoma	0	0	0	0	0	0	0	0	12	1	0	1	1	0	0	0	0	0	0	0	0	0
Gyrinidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	0
Helophorus	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0
Heptagenia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Hydraena	0	0	0	0	4	2	0	0	4	0	4	6	29	0	0	0	3	1	0	30	4	13
Hydrophilidae	1	7	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Hydropsyche	0	0	1	0	0	0	3	0	2	0	0	25	4	0	0	0	17	7	0	168	20	54
Hydroptilidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0
Isoperla	1	3	5	35	55	21	25	54	40	27	2	85	31	0	85	12	6	229	3	80	31	29
Leptophlebia	7	29	68	11	0	0	0	2	2	0	0	0	0	14	0	0	1	5	1	0	0	0
Leuctra	3	52	11	15	127	95	73	161	58	39	40	236	39	26	17	14	18	47	50	402	42	292
Limnephilidae	6	30	45	14	18	21	2	2	28	20	14	1	17	0	5	0	4	143	0	6	7	45
Limnius	0	0	0	0	0	1	0	0	1	0	0	7	0	0	1	14	2	19	0	18	1	6
Limoniidae	0	0	0	0	0	0	0	0	0	0	0	1	0	2	0	0	4	2	8	2	1	3
Nemoura	0	1	9	0	62	8	1	3	1	1	10	0	0	6	0	0	0	0	0	0	1	17
Nemurella	0	1	0	0	0	0	1	0	0	1	0	0	0	42	1	0	0	0	0	0	0	0
Odontocerum	0	0	0	0	4	0	1	0	2	2	0	0	0	0	0	0	0	0	0	2	0	0
Oulimnius	1	3	9	0	0	0	3	0	1	0	2	10	0	0	4	0	0	3	0	38	0	0

Appendix K – Species Diversity Data

Genus	Site name / site code (year)																					
	Cl1/12 (2012)	Cl1/97 (2013)	Cl4/14 (2012)	Cl4/98 (2013)	Gl1/21 (2012)	Gl1/102 (2013)	Ll6/38 (2012)	Ll6/108 (2013)	Ll7/39 (2012)	Ll7/109 (2013)	Afon Fechan/4 (2012)	Cerist (Afon)/10 (2012)	Nant Helygog/115 (2012)	Afon Pistyll/6 (2012)	Afon Colwyn/93 (2012)	Nant Y Gwyrd/55 (2012)	Upper Llugwy/59 (2012)	Nant Glan Dwr/114 (2012)	Nant Peiran/116 (2012)	Brefi/9 (2012)	Nant Clawdd/46 (2012)	Nant Dar/112 (2012)
Paraleptophlebia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48	0
Pediciidae	0	0	0	0	0	0	0	0	0	0	6	14	3	6	0	0	1	13	5	22	3	57
Perla	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
Perlodes	0	0	0	0	0	0	0	1	3	0	0	0	0	0	0	0	0	0	0	4	0	4
Philopotamus	0	0	0	0	3	0	0	0	0	0	2	0	23	0	0	0	0	0	0	0	21	12
Pisidium	0	0	2	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
Plectrocnemia	7	24	28	8	8	3	1	4	7	0	0	0	0	0	0	0	0	0	0	0	0	0
Polycentropodidae	0	0	0	0	0	0	0	0	0	0	2	0	2	22	8	2	1	16	9	10	1	4
Protonemura	0	1	0	0	2	15	1	2	5	4	0	1	0	0	0	0	2	0	0	0	1	8
Psychomyiidae	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Rhithrogena	0	1	0	0	3	20	133	22	139	19	32	34	96	0	0	0	39	0	0	210	76	85
Rhyacophila	0	0	0	0	0	1	1	4	3	0	2	10	3	2	1	0	1	6	15	22	3	7
Scirtidae	3	2	0	0	20	3	4	2	11	1	2	0	2	0	0	0	8	0	4	4	6	103
Sericostoma	0	0	0	0	1	0	1	0	3	1	0	0	2	0	0	0	17	0	0	14	6	2
Silo	0	0	0	0	8	3	5	1	17	2	2	3	6	0	0	0	0	0	0	0	3	10
Simuliidae	0	19	2	0	4	12	1	12	11	9	48	51	95	24	1	0	12	6	253	102	6	29
Siphlonurus	8	7	3	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Siphonoperla	1	12	10	26	101	55	16	9	3	3	0	3	0	92	16	8	11	32	13	40	5	24
Tipula	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0
Tipulidae	0	11	21	5	11	12	35	8	24	7	0	0	0	0	0	0	0	0	0	0	0	0
Velia	1	1	2	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	2	0	0	0
Wormaldia	1	0	0	0	1	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0	0	3

Table K3a-c. Trait data used to calculate functional diversity used in Chapter 4. For each taxon, information about the taxonomic class, order, family and genus, when family was the finest identified level, the family name was also assigned to the genus column. Trait-level information is generally at genus or family level, allowing the user to extract all the relevant levels for traits for all the taxa at the same time. Within R script (Appendix J, section 3) this table is named "trait_data.txt".

Table K3a

Order	Family	Genus	Life cycle duration (year)			Reproductive cycles (generation/year)			Aquatic stage		
			<1	>1	<1	1	>1	Egg	Larva	Nymph	Adult
Coleoptera	Dytiscidae	Agabus	0	3	0	3	1	3	3	0	2
Trichoptera	Glossosomatidae	Agapetus	3	0	0	3	1	3	3	3	0
Plecoptera	Nemouridae	Amphinemura	3	0	0	3	0	3	3	0	0
Gastropoda	Planorbidae	Ancylus	3	0	0	3	0	3	0	0	3
Diptera	Athericidae	Atherix	3	1	1	3	0	0	3	0	0
Ephemeroptera	Baetidae	Baetis	3	0	0	2	3	3	3	0	0
Plecoptera	Taeniopterygidae	Brachyptera	3	0	0	3	0	3	3	0	0
Ephemeroptera	Baetidae	Caenis	3	0	0	1	3	3	3	0	0
Diptera	Ceratopogonidae	Ceratopogonidae	3	0	0	2	2	1	2	2	0
Trichoptera	Limnephilidae	Chaetopteryx	3	0	0	3	0	3	3	3	0
Diptera	Chironomidae	Chironomidae	3	0	0	2	2	0	3	3	0
Plecoptera	Chloroperlidae	Chloroperla	1	3	3	1	0	3	3	0	0
Odonata	Cordulegastridae	Cordulegaster	0	3	3	0	0	2	3	0	0
Hemiptera	Corixidae	Corixidae	3	0	0	1	2	3	3	0	2
Trichoptera	Lepidostomatidae	Crunoecia	0	3	3	0	0	3	2	3	0
Trichoptera	Hydropsychidae	Diplectrona	3	1	1	3	0	3	3	3	0
Plecoptera	Perlodidae	Diura	1	3	3	2	0	3	3	0	0
Diptera	Dixidae	Dixidae	3	0	0	3	0	2	3	2	0
Trichoptera	Limnephilidae	Drusus	3	1	1	3	0	3	3	3	0
Trichoptera	Limnephilidae	Ecclisopteryx	1	0	0	3	0	3	3	3	0
Ephemeroptera	Heptageniidae	Ecdyonurus	3	1	1	3	0	3	3	0	0
Ephemeroptera	Heptageniidae	Electrogena	3	0	0	3	0	3	3	0	0
Coleoptera	Elmidae	Elmis	1	3	0	3	0	3	3	0	2

Appendix K – Species Diversity Data

Order	Family	Genus	Life cycle duration (year)			Reproductive cycles (generation/year)			Aquatic stage		
			<1	>1	<1	1	>1	Egg	Larva	Nymph	Adult
Diptera	Empididae	Empididae	3	0	0	3	2	0	3	3	0
Hirudinea	Erpobdellidae	Erpobdella	1	3	1	3	0	3	3	0	2
Coleoptera	Elmidae	Esolus	1	3	0	3	0	3	3	0	2
Amphipoda	Gammaridae	Gammarus	1	3	0	0	3	3	3	0	3
Trichoptera	Glossosomatidae	Glossosoma	3	0	0	2	1	3	3	3	0
Coleoptera	Gyrinidae	Gyrinidae	1	3	0	3	0	3	3	0	2
Coleoptera	Hydrophilidae	Helophorus	1	3	0	3	1	1	1	0	2
Ephemeroptera	Heptageniidae	Heptagenia	3	1	1	3	0	3	3	0	0
Coleoptera	Hydraenidae	Hydraena	1	3	0	3	1	3	0	0	3
Coleoptera	Hydrophilidae	Hydrophilidae	1	3	0	3	1	3	3	0	2
Trichoptera	Hydropsychidae	Hydropsyche	3	0	0	2	2	3	3	3	0
Trichoptera	Hydroptilidae	Hydroptilidae	3	0	0	1	2	3	3	3	0
Plecoptera	Perlodidae	Isoperla	1	3	2	3	0	3	3	0	0
Trichoptera	Leptophlebiidae	Leptophlebia	3	0	0	3	0	3	3	0	0
Plecoptera	Leuctridae	Leuctra	3	1	1	3	0	3	3	0	0
Trichoptera	Limnephilidae	Limnephilidae	2	0	0	3	0	3	3	3	0
Coleoptera	Elmidae	Limnius	1	3	0	3	0	3	3	0	2
Diptera	Limoniidae	Limoniidae	2	0	0	3	1	2	2	0	0
Plecoptera	Nemouridae	Nemoura	2	2	2	2	0	3	3	0	0
Plecoptera	Nemouridae	Nemurella	3	1	1	3	0	3	3	0	0
Trichoptera	Odontoceridae	Odontocerum	3	2	2	3	0	3	3	3	0
Coleoptera	Elmidae	Oulimnius	3	0	0	3	0	3	3	0	0
Trichoptera	Leptophlebiidae	Paraleptophlebia	3	0	0	3	1	2	2	0	0
Diptera	Pediciidae	Pediciidae	0	3	3	0	0	3	3	0	0
Plecoptera	Perlidae	Perla	0	3	1	3	0	3	3	0	0
Plecoptera	Perlidae	Perlodes	3	0	0	3	0	3	3	3	0
Trichoptera	Philopotamidae	Philopotamus	2	3	0	3	3	3	3	0	3
Bivalvia	Sphaeriidae	Pisidium	2	1	1	2	2	3	3	3	0
Trichoptera	Polycentropodidae	Plectrocnemia	3	0	0	2	3	3	3	3	0
Trichoptera	Polycentropodidae	Polycentropodidae	3	0	0	3	0	3	3	0	0
Plecoptera	Nemouridae	Protonemura	3	0	0	1	3	3	3	3	0
Trichoptera	Psychomyiidae	Psychomyiidae	3	1	1	3	0	3	3	0	0
Ephemeroptera	Heptageniidae	Rhithrogena	3	2	1	2	1	3	3	3	0

Order	Family	Genus	Life cycle duration (year)			Reproductive cycles (generation/year)			Aquatic stage		
			<1	>1	<1	1	>1	Egg	Larva	Nymph	Adult
Trichoptera	Rhyacophilidae	Rhyacophila	1	3	0	3	0	3	3	0	0
Coleoptera	Scirtidae	Scirtidae	3	1	2	3	1	3	3	3	0
Trichoptera	Sericostomatidae	Sericostoma	3	0	0	3	0	3	3	3	0
Trichoptera	Goeridae	Silo	3	1	0	2	3	2	3	3	0
Diptera	Simuliidae	Simuliidae	3	0	0	3	0	3	3	0	0
Ephemeroptera	Siphlonuridae	Siphlonurus	2	1	1	2	0	3	3	0	0
Plecoptera	Chloroperlidae	Siphonoperla	2	3	1	3	1	2	2	2	0
Diptera	Tipulidae	Tipula	2	3	1	3	1	2	2	2	0
Diptera	Tipulidae	Tipulidae	3	0	0	3	1	0	2	0	2
Hemiptera	Veliidae	Velia	0	3	3	0	0	3	3	3	0

Table K3b

Order	Family	Genus	Reproduction								Dispersion			
			Ovoviviparity	Isolated eggs, free	Isolated eggs, cemented	Clutches, cemented or fixed	Clutches, free	Clutches, in vegetation	Clutches, terrestrial	Asexual	Aquatic passive	Aquatic active	Aerial passive	Aerial active
Coleoptera	Dytiscidae	Agabus	0	0	0	3	0	1	0	0	1	1	0	3
Trichoptera	Glossosomatidae	Agapetus	0	0	0	3	0	0	0	0	1	2	1	1
Plecoptera	Nemouridae	Amphinemura	0	0	3	0	0	0	0	0	2	2	0	1
Gastropoda	Planorbidae	Ancylus	0	0	0	3	0	0	0	0	2	1	1	0
Diptera	Athericidae	Atherix	0	0	0	0	0	0	3	0	0	1	0	1
Ephemeroptera	Baetidae	Baetis	0	0	1	3	0	0	0	0	3	2	1	3
Plecoptera	Taeniopterygidae	Brachyptera	0	0	3	0	0	0	0	0	2	2	0	1
Ephemeroptera	Baetidae	Caenis	0	1	2	0	0	0	0	0	2	1	1	1
Diptera	Ceratopogonidae	Ceratopogonidae	0	0	1	3	0	0	0	0	3	1	0	2
Trichoptera	Limnephilidae	Chaetopteryx	0	0	0	3	0	0	0	0	0	1	0	3
Diptera	Chironomidae	Chironomidae	0	0	0	1	3	0	0	0	2	1	2	1
Plecoptera	Chloroperlidae	Chloroperla	0	0	2	2	0	0	0	0	3	2	0	1

Order	Family	Genus	Reproduction								Dispersion			
			Ovoviviparity	Isolated eggs, free	Isolated eggs, cemented	Clutches, cemented or fixed	Clutches, free	Clutches, in vegetation	Clutches, terrestrial	Asexual	Aquatic passive	Aquatic active	Aerial passive	Aerial active
Odonata	Cordulegastridae	Cordulegaster	0	3	0	0	0	0	1	0	2	0	0	3
Hemiptera	Corixidae	Corixidae	0	0	1	3	0	0	0	0	1	2	0	2
Trichoptera	Lepidostomatidae	Crunoecia	0	0	0	3	0	0	0	0	1	1	1	1
Trichoptera	Hydropsychidae	Diplectrona	0	0	0	3	0	0	0	0	3	1	1	3
Plecoptera	Perlodidae	Diura	0	0	0	1	3	0	0	0	1	2	0	1
Diptera	Dixidae	Dixidae	0	0	0	3	1	0	1	0	3	1	1	0
Trichoptera	Limnephilidae	Drusus	0	0	0	3	0	0	1	0	3	2	0	2
Trichoptera	Limnephilidae	Ecclisopteryx	0	0	0	3	0	0	1	0	0	2	0	2
Ephemeroptera	Heptageniidae	Ecdyonurus	0	2	0	3	0	0	0	0	3	1	1	3
Ephemeroptera	Heptageniidae	Electrogena	0	1	2	0	0	0	0	0	3	1	1	3
Coleoptera	Elmidae	Elmis	0	0	0	3	0	0	0	0	2	1	0	2
Diptera	Empididae	Empididae	0	0	0	0	0	0	2	0	1	1	0	2
Hirudinea	Erpobdellidae	Erpobdella	0	0	0	3	0	0	0	0	1	1	1	0
Coleoptera	Elmidae	Esolus	0	0	0	3	0	0	0	0	2	1	0	2
Amphipoda	Gammaridae	Gammarus	3	0	0	0	0	0	0	0	3	2	0	0
Trichoptera	Glossosomatidae	Glossosoma	0	0	0	3	0	0	0	0	1	2	1	1
Coleoptera	Gyrinidae	Gyrinidae	0	0	0	3	0	0	0	0	0	2	0	3
Coleoptera	Hydrophilidae	Helophorus	0	0	0	0	0	0	3	0	1	1	0	3
Ephemeroptera	Heptageniidae	Heptagenia	0	1	2	0	0	0	0	0	3	1	1	3
Coleoptera	Hydraenidae	Hydraena	0	0	0	0	0	0	3	0	2	1	0	3
Coleoptera	Hydrophilidae	Hydrophilidae	0	0	0	2	0	0	0	0	0	1	0	3
Trichoptera	Hydropsychidae	Hydropsyche	0	0	0	3	0	0	0	0	3	2	1	3
Trichoptera	Hydroptilidae	Hydroptilidae	0	0	0	3	0	0	0	0	1	1	2	1
Plecoptera	Perlodidae	Isoperla	0	0	1	2	0	0	0	0	2	2	0	2
Trichoptera	Leptophlebiidae	Leptophlebia	0	1	3	0	0	0	0	0	2	1	1	3
Plecoptera	Leuctridae	Leuctra	0	0	3	0	0	0	0	0	2	2	0	1
Trichoptera	Limnephilidae	Limnephilidae	0	0	0	2	0	0	0	0	1	2	0	2
Coleoptera	Elmidae	Limnius	0	0	0	3	0	0	0	0	2	1	0	2
Diptera	Limoniidae	Limoniidae	0	1	0	1	0	0	0	0	0	1	1	1
Plecoptera	Nemouridae	Nemoura	0	0	3	1	0	0	0	0	2	2	0	2

Order	Family	Genus	Reproduction								Dispersion			
			Ovoviviparity	Isolated eggs, free	Isolated eggs, cemented	Clutches, cemented or fixed	Clutches, free	Clutches, in vegetation	Clutches, terrestrial	Asexual	Aquatic passive	Aquatic active	Aerial passive	Aerial active
Plecoptera	Nemouridae	Nemurella	0	0	2	2	0	0	0	0	3	2	0	1
Trichoptera	Odontoceridae	Odontocerum	0	0	0	0	0	0	3	0	3	1	1	2
Coleoptera	Elmidae	Oulimnius	0	0	3	0	0	0	0	0	2	1	1	3
Trichoptera	Leptophlebiidae	Paraleptophlebia	0	1	0	0	1	0	0	0	2	1	0	1
Diptera	Pediciidae	Pediciidae	0	0	3	0	2	0	0	0	1	2	0	1
Plecoptera	Perlidae	Perla	0	0	3	0	0	0	0	0	1	2	0	1
Plecoptera	Perlidae	Perlodes	0	0	0	3	0	0	0	0	1	1	1	1
Trichoptera	Philopotamidae	Philopotamus	3	0	0	0	0	0	0	0	3	1	1	0
Bivalvia	Sphaeriidae	Pisidium	0	0	0	2	1	0	0	0	2	1	1	2
Trichoptera	Polycentropodidae	Plectrocnemia	0	0	0	3	0	0	0	0	2	1	1	2
Trichoptera	Polycentropodidae	Polycentropodidae	0	0	3	0	0	0	0	0	2	2	0	1
Plecoptera	Nemouridae	Protonemura	0	0	0	3	0	0	0	0	1	1	1	1
Trichoptera	Psychomyiidae	Psychomyiidae	0	1	1	3	0	0	0	0	3	2	1	3
Ephemeroptera	Heptageniidae	Rhithrogena	0	0	3	0	1	0	0	0	3	2	1	2
Trichoptera	Rhyacophilidae	Rhyacophila	0	0	0	0	0	0	3	0	0	1	0	3
Coleoptera	Scirtidae	Scirtidae	0	0	0	3	0	0	0	0	2	1	1	1
Trichoptera	Sericostomatidae	Sericostoma	0	0	0	3	0	0	0	0	1	1	1	2
Trichoptera	Goeridae	Silo	0	1	0	3	1	0	1	0	2	2	3	1
Diptera	Simuliidae	Simuliidae	0	1	3	0	0	1	0	0	1	2	1	3
Ephemeroptera	Siphlonuridae	Siphlonurus	0	0	3	0	0	0	0	0	2	2	0	1
Plecoptera	Chloroperlidae	Siphonoperla	0	3	0	0	0	0	1	0	0	1	0	1
Diptera	Tipulidae	Tipula	0	3	0	0	0	0	1	0	0	1	0	1
Diptera	Tipulidae	Tipulidae	0	0	0	3	0	0	3	0	1	3	0	1
Hemiptera	Veliidae	Velia	0	0	0	3	0	0	0	0	1	2	1	1

Table K3c

Order	Family	Genus	Resistance form					Respiration type				
			Eggs, statoblasts	Cocoons	House	Diapause or dormancy	None	Tegument	Gill	Plastron	Spiracle	Hydrostatic vesicle
Coleoptera	Dytiscidae	Agabus	0	0	0	0	3	1	0	0	3	0
Trichoptera	Glossosomatidae	Agapetus	0	2	0	2	0	3	0	0	0	0
Plecoptera	Nemouridae	Amphinemura	1	0	0	2	2	3	2	0	0	0
Gastropoda	Planorbidae	Ancylus	0	0	0	0	2	3	0	0	0	0
Diptera	Athericidae	Atherix	0	0	0	0	3	0	1	0	0	0
Ephemeroptera	Baetidae	Baetis	2	0	0	0	2	1	2	0	0	0
Plecoptera	Taeniopterygidae	Brachyptera	1	0	0	2	0	3	0	0	0	0
Ephemeroptera	Baetidae	Caenis	2	0	0	1	2	1	3	0	0	0
Diptera	Ceratopogonidae	Ceratopogonidae	1	0	0	0	3	0	3	0	0	0
Trichoptera	Limnephilidae	Chaetopteryx	0	0	0	2	1	2	2	0	0	0
Diptera	Chironomidae	Chironomidae	0	0	0	0	3	3	1	0	0	0
Plecoptera	Chloroperlidae	Chloroperla	0	0	0	0	3	3	0	0	0	0
Odonata	Cordulegastridae	Cordulegaster	0	0	0	0	3	1	3	0	0	0
Hemiptera	Corixidae	Corixidae	0	0	0	1	0	1	0	2	3	0
Trichoptera	Lepidostomatidae	Crunoecia	0	0	0	0	0	2	2	0	0	0
Trichoptera	Hydropsychidae	Diplectrona	0	0	0	0	3	2	0	0	0	0
Plecoptera	Perlodidae	Diura	3	0	0	2	0	3	0	0	0	0
Diptera	Dixidae	Dixidae	0	0	0	2	0	0	0	0	3	0
Trichoptera	Limnephilidae	Drusus	0	0	0	0	1	3	2	0	0	0
Trichoptera	Limnephilidae	Ecclisopteryx	0	0	0	0	0	3	2	0	0	0
Ephemeroptera	Heptageniidae	Ecdyonurus	2	0	0	1	2	1	3	0	0	0
Ephemeroptera	Heptageniidae	Electrogena	2	0	0	1	2	1	3	0	0	0
Coleoptera	Elmidae	Elmis	0	0	0	0	3	1	3	3	0	0
Diptera	Empididae	Empididae	0	0	0	0	0	1	0	0	1	0
Hirudinea	Erpobdellidae	Erpobdella	0	0	0	0	3	3	0	0	0	0
Coleoptera	Elmidae	Esolus	0	0	0	0	3	1	3	3	0	0
Amphipoda	Gammaridae	Gammarus	0	0	0	0	3	0	3	0	0	0
Trichoptera	Glossosomatidae	Glossosoma	0	2	0	2	0	3	0	0	0	0
Coleoptera	Gyrinidae	Gyrinidae	0	0	0	0	3	1	3	0	3	0
Coleoptera	Hydrophilidae	Helophorus	0	0	0	0	3	0	0	1	3	0
Ephemeroptera	Heptageniidae	Heptagenia	1	0	0	2	2	1	3	0	0	0

Order	Family	Genus	Resistance form					Respiration type				
			Eggs, statoblasts	Cocoons	House	Diapause or dormancy	None	Tegument	Gill	Plastron	Spiracle	Hydrostatic vesicle
Coleoptera	Hydraenidae	Hydraena	0	0	0	0	3	0	0	3	2	0
Coleoptera	Hydrophilidae	Hydrophilidae	0	0	0	0	3	1	0	1	3	0
Trichoptera	Hydropsychidae	Hydropsyche	0	0	0	0	3	2	3	0	0	0
Trichoptera	Hydroptilidae	Hydroptilidae	0	0	0	0	3	3	1	0	0	0
Plecoptera	Perlodidae	Isoperla	2	0	0	1	1	3	0	0	0	0
Trichoptera	Leptophlebiidae	Leptophlebia	0	0	0	0	3	1	3	0	0	0
Plecoptera	Leuctridae	Leuctra	1	0	0	0	3	3	0	0	0	0
Trichoptera	Limnephilidae	Limnephilidae	0	0	0	1	2	2	2	0	0	0
Coleoptera	Elmidae	Limnius	0	0	0	0	3	1	3	3	0	0
Diptera	Limoniidae	Limoniidae	0	0	0	0	2	0	0	0	3	0
Plecoptera	Nemouridae	Nemoura	0	0	0	1	3	3	0	0	0	0
Plecoptera	Nemouridae	Nemurella	0	0	0	1	2	3	0	0	0	0
Trichoptera	Odontoceridae	Odontocerum	0	0	0	0	3	2	3	0	0	0
Coleoptera	Elmidae	Oulimnius	0	0	0	0	3	1	3	0	0	0
Trichoptera	Leptophlebiidae	Paraleptophlebia	0	0	0	0	2	0	0	0	3	0
Diptera	Pediciidae	Pediciidae	2	0	0	0	2	0	3	0	0	0
Plecoptera	Perlidae	Perla	2	0	0	0	3	3	0	0	0	0
Plecoptera	Perlidae	Perlodes	0	0	0	2	2	3	0	0	0	0
Trichoptera	Philopotamidae	Philopotamus	1	0	0	1	0	1	3	0	0	0
Bivalvia	Sphaeriidae	Pisidium	0	0	0	0	3	3	0	0	0	0
Trichoptera	Polycentropodidae	Plectrocnemia	0	0	0	0	3	3	0	0	0	0
Trichoptera	Polycentropodidae	Polycentropodidae	1	0	0	1	3	3	2	0	0	0
Plecoptera	Nemouridae	Protonemura	0	0	0	0	3	3	0	0	0	0
Trichoptera	Psychomyiidae	Psychomyiidae	2	0	0	0	2	1	3	0	0	0
Ephemeroptera	Heptageniidae	Rhithrogena	1	0	0	2	3	2	2	0	0	0
Trichoptera	Rhyacophilidae	Rhyacophila	0	0	0	0	3	1	3	0	0	0
Coleoptera	Scirtidae	Scirtidae	0	0	0	1	3	2	2	0	0	0
Trichoptera	Sericostomatidae	Sericostoma	0	0	0	0	3	2	3	0	0	0
Trichoptera	Goeridae	Silo	1	0	0	1	0	3	1	0	2	0
Diptera	Simuliidae	Simuliidae	3	0	0	0	0	1	2	0	0	0
Ephemeroptera	Siphlonuridae	Siphlonurus	0	0	0	0	3	3	0	0	0	0
Plecoptera	Chloroperlidae	Siphonoperla	0	2	0	0	2	2	1	0	3	0
Diptera	Tipulidae	Tipula	0	2	0	0	2	2	1	0	3	0

Appendix K – Species Diversity Data

Order	Family	Genus	Resistance form					Respiration type				
			Eggs, statoblasts	Cocoons	House	Diapause or dormancy	None	Tegument	Gill	Plastron	Spiracle	Hydrostatic vesicle
Diptera	Tipulidae	Tipulidae	0	0	0	3	0	0	0	0	3	0
Hemiptera	Veliidae	Velia	0	0	0	0	3	3	1	0	0	0

Table K4. Species diversity indices, created using information in Table K1-K3, as described within R script (Appendix J, section 3). Within R script this table is named “div.txt”.

Site name	Site code	Year	Shannon diversity	Species richness	Functional diversity
CI1	12	2012	2.74	21	0.14
CI1	97	2013	2.86	29	0.14
CI4	14	2012	2.53	26	0.14
CI4	98	2013	2.21	15	0.15
GI1	21	2012	2.80	36	0.14
GI1	102	2013	2.73	32	0.14
LI6	38	2012	2.32	34	0.13
LI6	108	2013	2.47	31	0.11
LI7	39	2012	2.70	41	0.13
LI7	109	2013	2.51	29	0.12
Afon Fechan	4	2012	2.75	30	0.17
Cerist (Afon)	10	2012	2.39	33	0.14
Nant Helygog	115	2012	2.40	28	0.14
Afon Pistyll	6	2012	1.53	16	0.10
Afon Colwyn	93	2012	2.09	20	0.13
Nant y Gwryd	55	2012	1.64	11	0.10
Upper Llugwy	59	2012	2.82	33	0.13
Nant Glan dwr	114	2012	2.36	32	0.15
Nant Peiran	116	2012	2.03	20	0.13
Brefi	9	2012	2.74	37	0.14
Nant Clawdd	46	2012	2.82	37	0.13
Nant Dar	112	2012	2.73	36	0.13

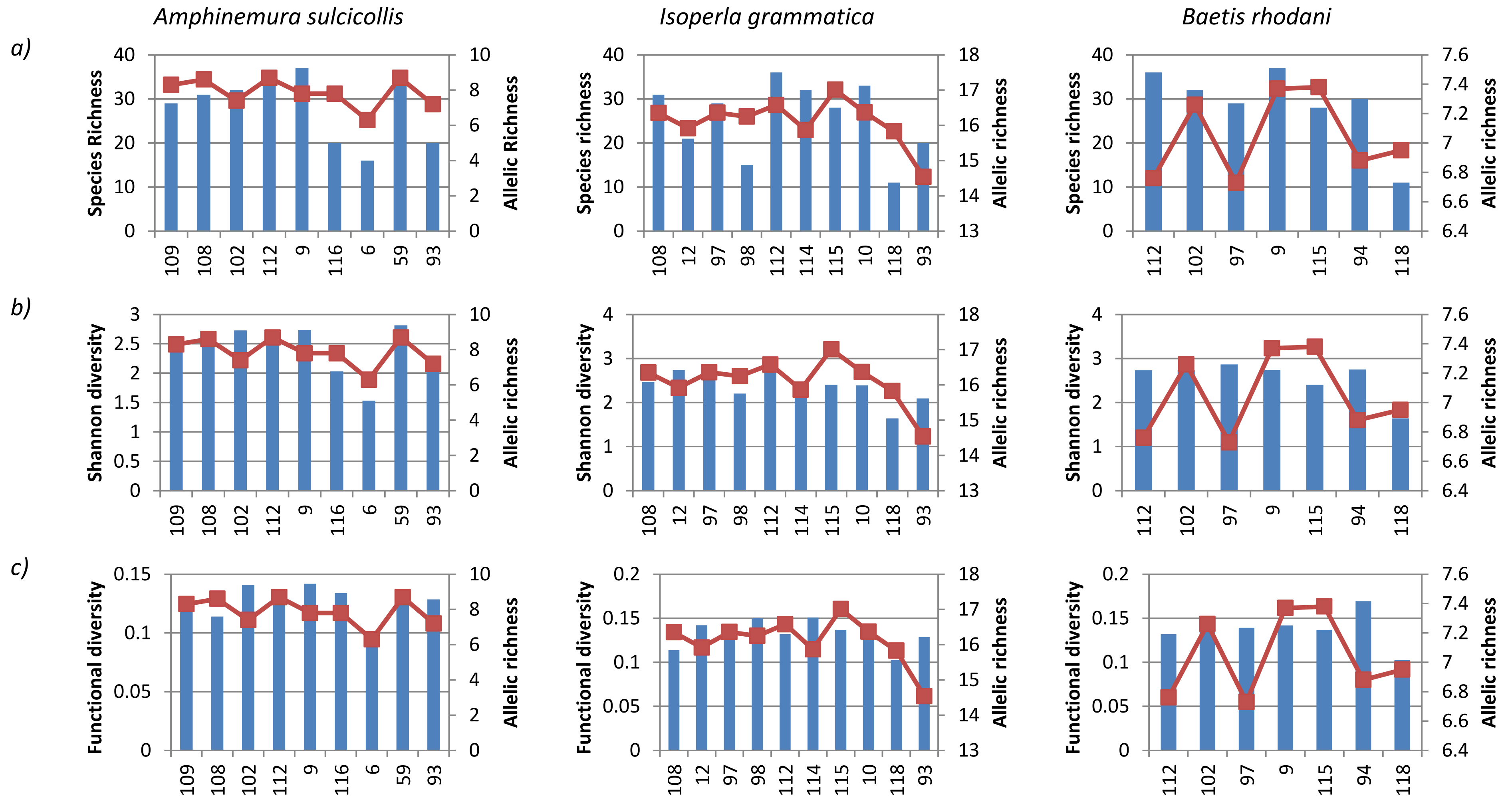


Figure K1. Showing mean allelic richness plotted against a) Species richness, b) Shannon diversity and c) Functional diversity for all three species.

Appendix L - Chapter 5 Supporting Information

Contains supplementary information detailing how data was split into clusters for demographic history analysis within Chapter 5, and results from BOTTLENECK and MSVAR not included in the main chapter.

Table of tables

Table L1 529
Table L2a-c. 531
Table L3a-c. 536

Table of figures

Figure L1 528
Figure L2 528
Figure L3 529
Figure L4 534
Figure L5 535
Figure L6 536

How genotyping data was split into clusters

To summarise, within Chapter 3 genetic differentiation was assessed and each species was separated into clusters using data from pairwise FST (Table 3.2a-c), STRUCTURE (Figure 3.1 – 3.3) and comparing genetic diversity (Table 3.4 a-c). *Amphinemura sulcicollis* was split into four clusters (Figure L1), *Isoperla grammatica* was split into two clusters (Figure L2) and *Baetis rhodani*, as it contained very weak genetic structure, was assumed to contain just one cluster (Figure L3).

Data was separated into clusters by looking at the output from STRUCTURE assuming two clusters (K = 2; after averaging iterations using CLUMP). Individuals were separated into clusters based on the q-value, an individual was assigned to a cluster only if it had a q-value of more than 0.8 belonging to one cluster. If it was a mixed (*ad hoc*) cluster like Cluster 2 for *A. sulcicollis*, and Cluster 1 for *I. grammatica* and *B. rhodani*, then individuals were used that did not fit into each cluster (i.e. they were mixed, had a q-value of less than 0.8). Using this method Table L1 shows the sample sizes for all clusters and representative sites. For larger clusters fifty individuals are randomly selected. For 'per site' analysis, all samples were used regardless of STRUCTURE results. MSVAR simulations could not be run for all sites, as with BOTTLENECK and MPval, so representative sites from each of the clusters were chosen (shown in Figure L1 - L3 and Table L1).

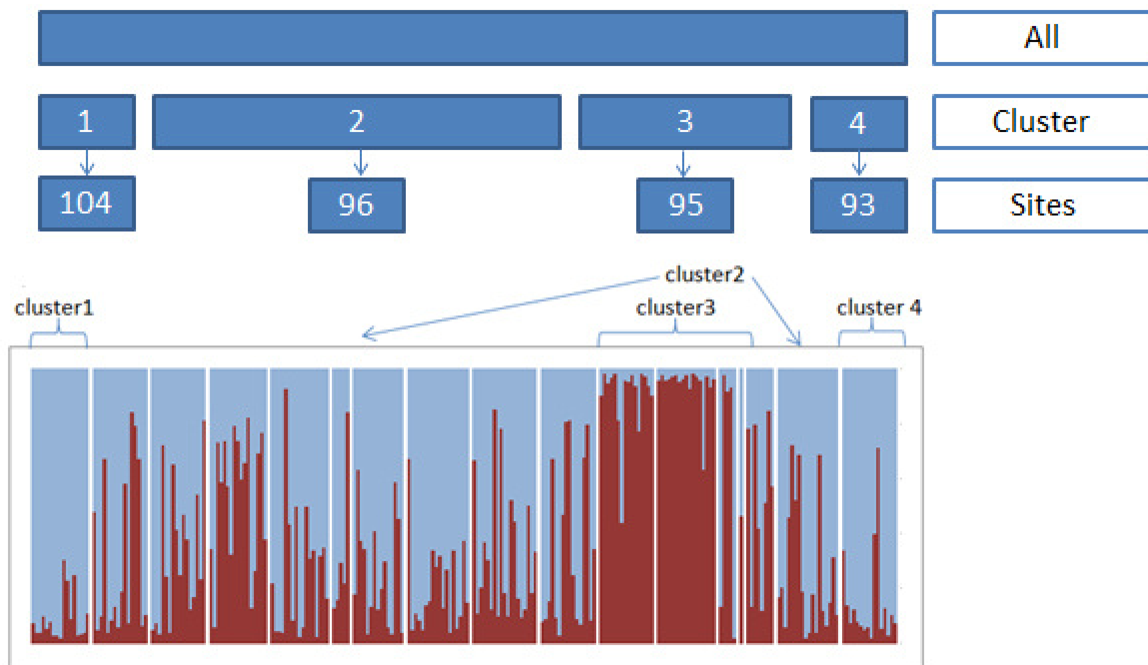


Figure L1. Showing how the four clusters were separated in *Amphinemura sulcicollis*. Contains a STRUCTURE plot assuming two clusters (k = 2) and above shows clusters and sites chosen to represent each cluster, used to assess demographic history using MSVAR.

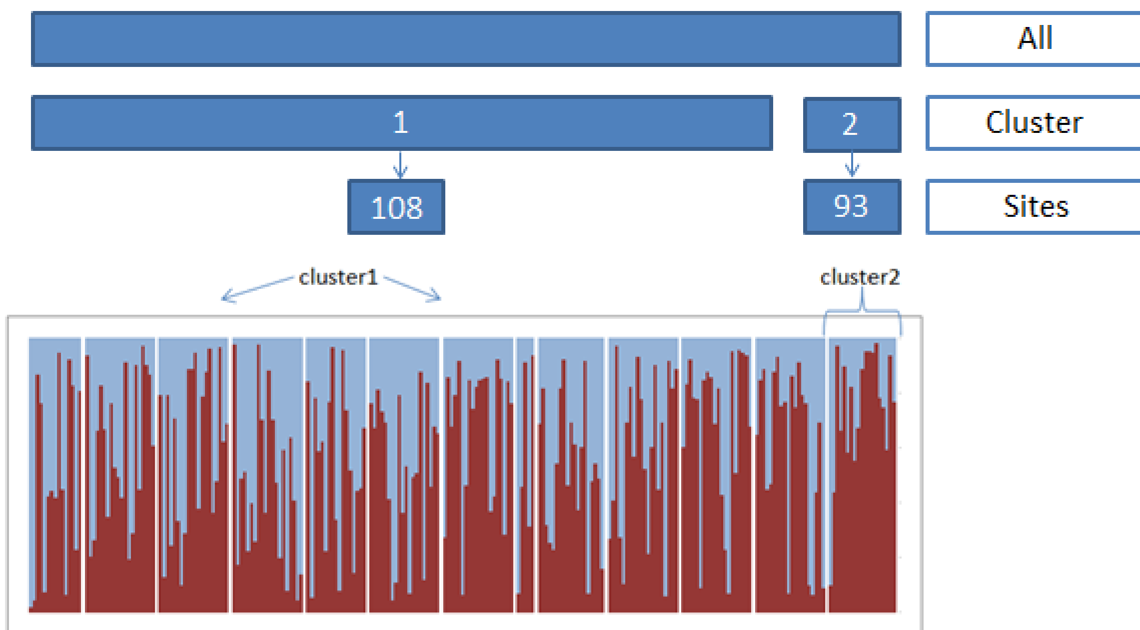


Figure L2. Showing how the two clusters were separated in *Isoperla grammatica*. Contains a STRUCTURE plot assuming two clusters (k = 2) and above shows clusters and sites chosen to represent each cluster, used to assess demographic history using MSVAR.

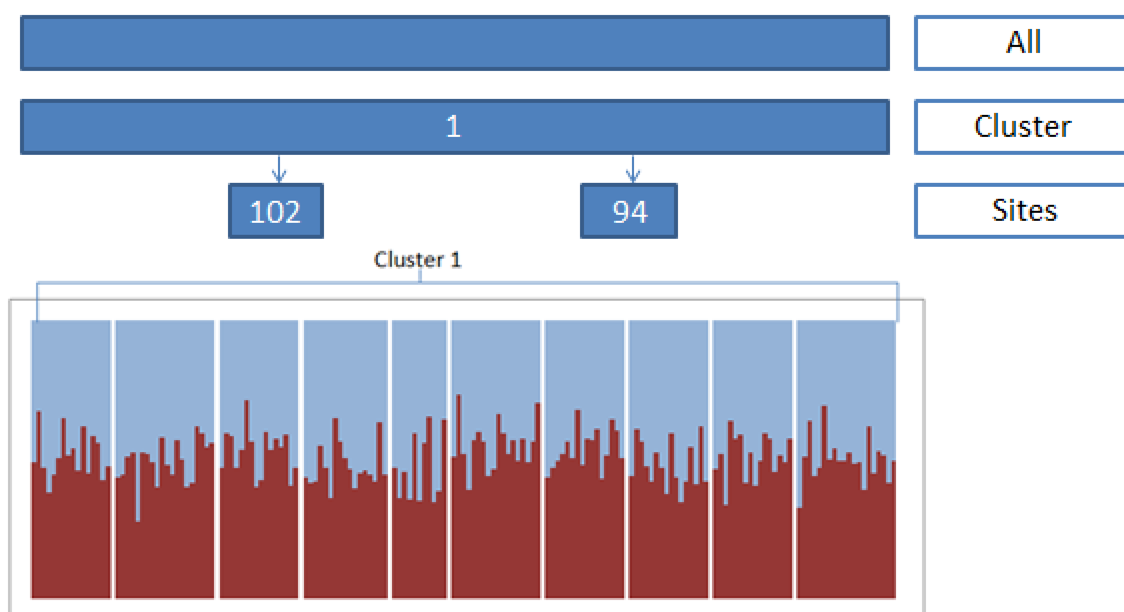


Figure L3. Showing the sole cluster for *Baetis rhodani*. Contains a STRUCTURE plot assuming two clusters (k = 2) and above shows the whole dataset as one cluster and sites chosen to represent this cluster, used to assess demographic history using MSVAR.

Table L1. Details explaining how data was split into clusters for demographic history analysis of *Amphinemura sulcicollis, Isoperla grammatica* and *Baetis rhodani*, including final sample size for each cluster and representative site chosen for MSVAR.

	Amphinemura sulcicollis				*Isoperla grammatica*		*Baetis rhodani*
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 1	Cluster 2	Cluster 1
Relating to STRUCTURE plot (Figure L1 - L3)	Blue south (q = >0.8)	Mixed (q = <0.8)	Red north (q = >0.8)	Blue north (q = >0.8)	Mixed (q = <0.8)	Red north(q = >0.8)	Mixed (q = <0.8)
Sites incorporated into that cluster	104	109, 108, 102, 112, 46, 96, 9, 116, 113, 55, 59	95, 6, 94, 4	93	105, 106, 108, 12, 97, 98, 112, 46, 114, 115, 10, 118	93	all
Total sample size	17	165	39	16	216	19	166
Final sample size of cluster (after using q-value method)	14	(80) Random 50 selected	33	13	(119) Random 50 selected	9	(166) Random 50 selected

	Amphinemura sulcicollis				*Isoperla grammatica*		*Baetis rhodani*
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 1	Cluster 2	Cluster 1
Representative sites chosen for MSVAR	104	96	95	93	108	93	102, 94
Representative sites sample size	17	15	16	16	20	19	20, 16

Table L2a-c. Showing results from the Sign test (under the infinite allele (IAM), stepwise mutation (SMM) , and two-phase (TPM) model (TPM assumed variance = 30 and % of one step mutations =70%) and S.M.M) and the Wilcoxon test (under IAM and SMM) using BOTTLENECK for a) *Amphinemura sulcicollis*; b) *Isoperla grammatica* and c) *Baetis rhodani*. For Sign test: *He excess* = expected number of loci with heterozygosity excess; *H def* = number of loci with heterozygosity deficiency; *H ex* = number of loci with heterozygosity excess and *P* = probability. For the Wilcoxon test: *P H def* = probability of heterozygosity deficiency; *P H excess* = probability of heterozygosity excess and *P (H excess & Def)* = two-tailed probability of heterozygosity deficiency or excess. **Bold** = significance (p = <0.05).

Table L2a. *Amphinemura sulcicollis*.

Site	Sign tests												Wilcoxon test					
	IAM				TPM				SMM				IAM			SMM		
	He excess	H def	H ex	P	He excess	H def	H ex	P	He excess	H def	H ex	P	P H def	P H excess	P (H excess & def)	P H def	P H excess	P (H excess & def)
104	6.08	3	7	0.404	6.01	3	7	0.383	5.85	5	5	0.405	0.813	0.216	0.432	0.188	0.839	0.375
109	6.09	4	6	0.595	6.06	4	6	0.602	5.97	6	4	0.172	0.958	0.053	0.105	0.161	0.862	0.322
108	6.04	3	7	0.393	6.05	7	3	0.051	5.98	9	1	**0.002**	0.813	0.216	0.432	**0.001**	0.999	**0.003**
102	6.01	2	8	0.168	5.97	3	7	0.375	5.89	5	5	0.395	0.958	0.053	0.105	0.348	0.688	0.695
112	5.99	4	6	0.630	6.05	6	4	0.158	5.99	7	3	0.056	0.577	0.461	0.922	**0.012**	0.991	**0.024**
96	6.00	2	8	0.168	6.00	5	5	0.368	6.00	7	3	0.055	0.884	0.138	0.275	0.065	0.947	0.131
9	6.01	4	6	0.616	6.13	6	4	0.145	5.95	7	3	0.059	0.754	0.278	0.557	**0.012**	0.991	**0.024**
116	5.96	3	7	0.373	6.01	5	5	0.365	5.87	6	4	0.188	0.688	0.348	0.695	0.080	0.935	0.160
113	6.02	3	7	0.388	6.08	6	4	0.153	5.97	7	3	0.057	0.862	0.161	0.322	0.065	0.947	0.131
95	6.06	3	7	0.396	6.07	5	5	0.349	5.89	6	4	0.185	0.722	0.313	0.625	0.080	0.935	0.160
6	5.92	4	6	0.613	5.95	5	5	0.380	5.93	8	2	**0.014**	0.615	0.423	0.846	**0.005**	0.997	**0.010**
59	6.02	4	6	0.612	6.01	7	3	0.054	5.86	9	1	**0.002**	0.839	0.188	0.375	**0.002**	0.999	**0.005**
93	6.02	4	6	0.613	6.00	6	4	0.165	5.89	7	3	0.063	0.461	0.577	0.922	**0.016**	0.988	**0.032**
Cluster 1	6.10	3	7	0.408	6.07	4	6	0.599	5.98	4	6	0.628	0.920	0.097	0.193	0.278	0.754	0.557
Cluster 2	6.06	3	7	0.397	5.96	6	4	0.173	5.92	9	1	**0.002**	0.958	0.053	0.105	**0.003**	0.998	**0.007**

Site	Sign tests												Wilcoxon test					
	IAM				TPM				SMM				IAM			SMM		
	He excess	H def	H ex	P	He excess	H def	H ex	P	He excess	H def	H ex	P	P H def	P H excess	P (H excess & def)	P H def	P H excess	P (H excess & def)
Cluster 3	6.03	1	9	**0.048**	5.95	5	5	0.378	5.94	8	2	**0.014**	0.984	**0.042**	0.084	**0.002**	0.999	**0.005**
Cluster 4	5.97	5	5	0.374	6.01	5	5	0.366	5.95	7	3	0.058	0.313	0.722	0.625	**0.042**	0.984	0.084

Table L2b. *Isoperla grammatica*.

Site	Sign tests												Wilcoxon test					
	IAM				TPM				SMM				IAM			SMM		
	He excess	H def	H ex	P	He excess	H def	H ex	P	He excess	H def	H ex	P	P H def	P H excess	P (H excess & def)	P H def	P H excess	P (H excess & def)
105I9	6.11	0	10	**0.007**	6.13	3	7	0.416	5.73	6	4	0.213	1.000	**0.000**	**0.001**	0.348	0.688	0.695
106I9	5.95	2	8	0.160	6.29	4	6	0.543	5.93	6	4	0.177	0.947	0.065	0.131	0.138	0.884	0.275
108I9*	6.01	2	8	0.168	6.34	1	9	0.071					0.997	**0.005**	**0.010**			
12I9	6.03	3	7	0.388	6.30	4	6	0.541	5.84	5	5	0.409	0.988	**0.016**	**0.032**	0.246	0.784	0.492
97I9	6.05	1	9	**0.049**	6.08	3	7	0.403	5.98	6	4	0.169	0.997	**0.005**	**0.010**	0.461	0.577	0.922
98I9	6.04	1	9	**0.049**	6.12	3	7	0.414	5.98	6	4	0.170	0.997	**0.005**	**0.010**	0.278	0.754	0.557
112I9	6.05	1	9	**0.050**	6.15	3	7	0.421	5.89	5	5	0.394	1.000	**0.001**	**0.002**	0.423	0.615	0.846
114I9	5.96	4	6	0.624	6.28	5	5	0.298	6.09	6	4	0.151	0.884	0.138	0.275	0.138	0.884	0.275
115I9	6.11	1	9	0.053	6.25	4	6	0.554	5.95	7	3	0.058	0.991	**0.012**	**0.024**	0.138	0.884	0.275
10I9	6.07	2	8	0.179	6.24	4	6	0.555	5.84	7	3	0.067	0.995	**0.007**	**0.014**	0.053	0.958	0.105
118I9	6.02	2	8	0.170	6.33	3	7	0.467	5.97	4	6	0.627	0.997	**0.005**	**0.010**	0.278	0.754	0.557
93I9	6.03	8	2	**0.012**	6.06	8	2	**0.011**	5.99	9	1	**0.002**	**0.016**	0.988	**0.032**	**0.001**	0.999	**0.003**
Cluster 1	6.00	0	10	**0.006**	6.21	5	5	0.316	6.15	9	1	**0.001**	1.000	**0.000**	**0.001**	**0.007**	0.995	**0.014**
Cluster 2	5.93	6	4	0.178	5.97	8	2	**0.012**	6.45	8	2	**0.005**	0.216	0.813	0.432	**0.012**	0.991	**0.024**

* Results for site 108 could not be found under the S. M. M model in either sign test or Wilcoxon test.

Table L2c. *Baetis rhodani*.

Site	Sign tests												Wilcoxon test					
	IAM				TPM				SMM				IAM			SMM		
	He excess	H def	H ex	P	He excess	H def	H ex	P	He excess	H def	H ex	P	P H def	P H excess	P (H excess & def)	P H def	P H excess	P (H excess & def)
112B9	7.83	10	3	0.174	7.94	7	6	0.394	7.70	7	6	0.246	0.980	**0.024**	**0.048**	0.084	0.927	0.168
102B9	7.87	5	8	0.590	7.77	10	3	**0.008**	7.75	11	2	**0.001**	0.830	0.188	0.376	**0.003**	0.998	**0.005**
97B9	7.82	2	11	0.059	7.86	7	6	0.218	7.68	10	3	**0.009**	0.966	**0.040**	0.080	**0.004**	0.997	**0.009**
96B8	7.82	5	8	0.579	7.75	7	6	0.238	7.71	11	2	**0.002**	0.905	0.108	0.216	**0.001**	0.999	**0.002**
09B5	7.56	6	7	0.480	7.95	7	6	0.203	7.65	9	4	**0.039**	0.393	0.632	0.787	**0.016**	0.987	**0.033**
106B9	7.85	3	10	0.175	7.82	8	5	0.095	7.67	11	2	**0.002**	0.960	**0.047**	0.094	**0.002**	0.999	**0.003**
113B9	7.80	5	8	0.575	7.77	7	6	0.235	7.70	11	2	**0.002**	0.892	0.122	0.244	**0.000**	1.000	**0.001**
115B9	7.83	3	10	0.173	7.88	10	3	**0.007**	7.67	11	2	**0.002**	0.936	0.073	0.146	**0.001**	1.000	**0.001**
94B9	7.82	4	9	0.358	7.76	7	6	0.237	7.63	10	3	**0.010**	0.847	0.170	0.340	**0.020**	0.984	**0.040**
118B9	7.83	4	9	0.359	7.76	6	7	0.434	7.70	9	4	**0.036**	0.984	**0.020**	**0.040**	**0.029**	0.976	0.057
Cluster 1	7.80	6	7	0.425	7.79	10	3	**0.008**	7.64	13	**0**	**0.000**	0.446	0.580	0.893	**0.000**	1.000	**0.000**

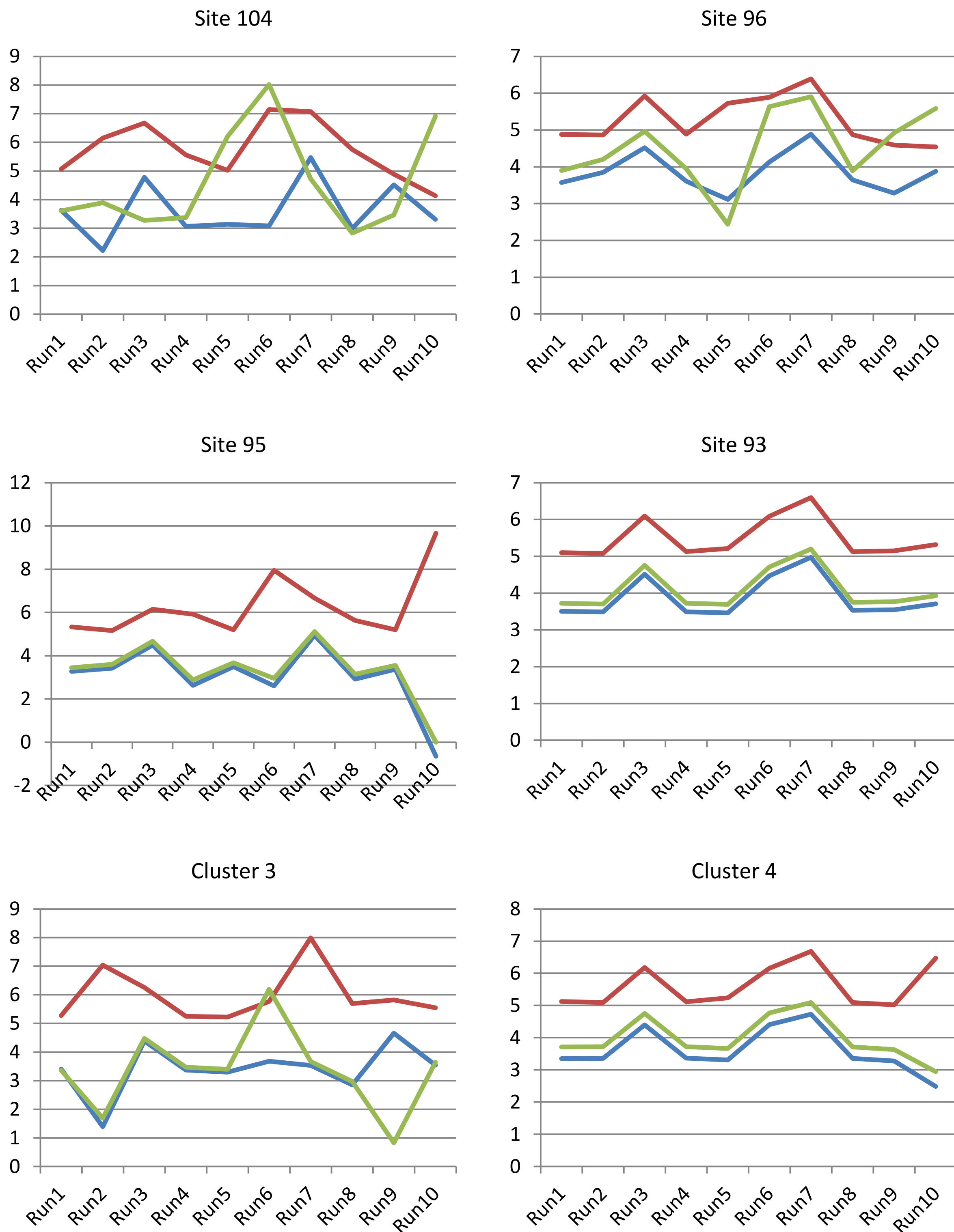


Figure L4. Showing mean values of N_1 (log10 effective population size at time Ta) = red line; N_0 (log10 current effective population size) = blue line and Ta (log10 Time between N_0 and N_1) = green line, for Amphinemura *sulcicollis* for four representative sites and two clusters, each having ten independent runs.

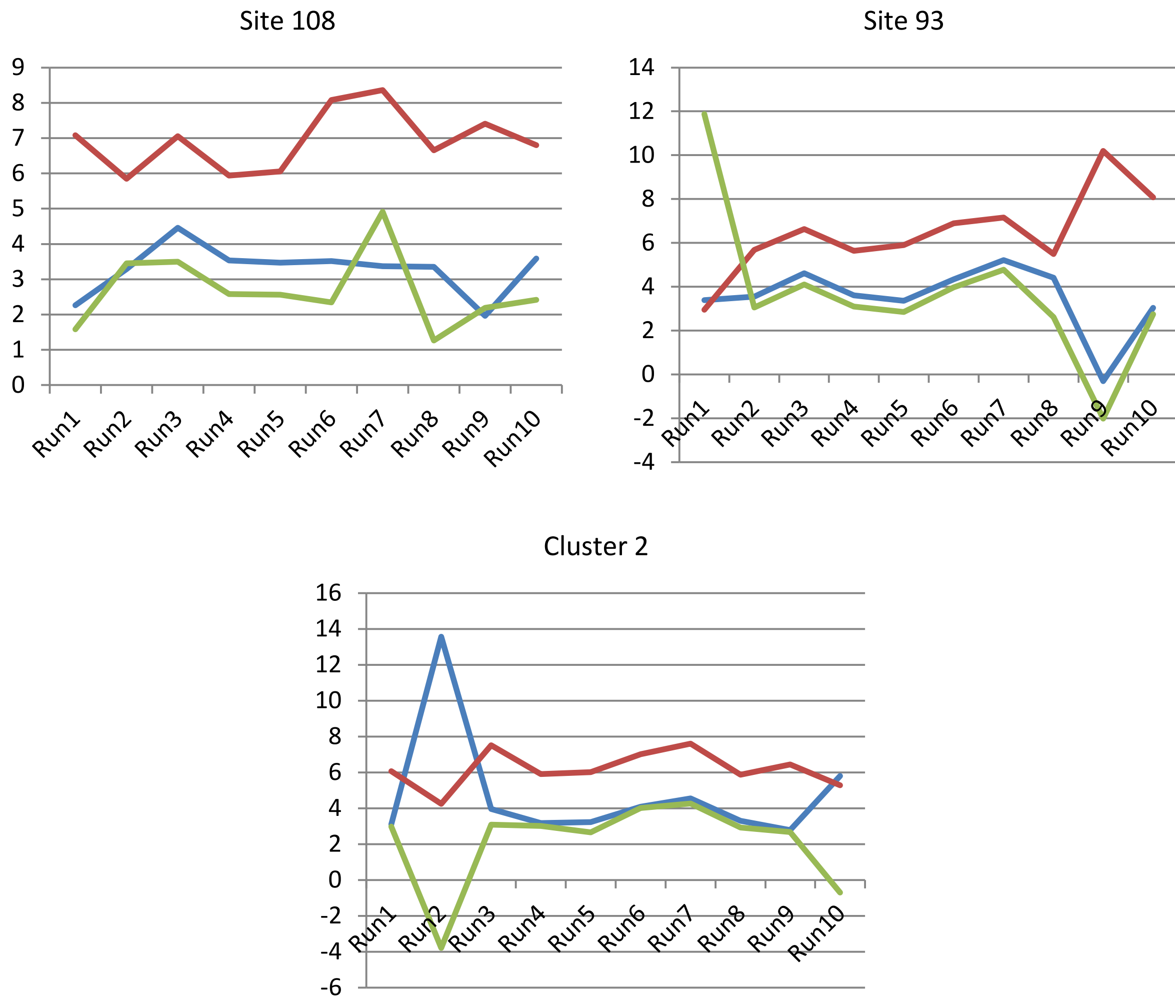


Figure L5. Showing mean values of N_1 (log10 effective population size at time Ta) = red line; N_0 (log10 current effective population size) = blue line and Ta (log 10 Time between N_0 and N_1) = green line, for *Isoperla grammatica* for two representative sites and one clusters, each having ten independent runs.

Figure L6. Showing mean values of N_1 (log10 effective population size at time Ta) = red line; N_0 (log10 current effective population size) = blue line and Ta (log10 Time between N_0 and N_1) = green line, for *Baetis rhodani* for two representative sites, 102 and 94, having 3 and 9 independent runs, respectively.

Table L3a-c. Showing summary statistics for N_0 (current effective population size), N_1 (effective population size at time *Ta*) and *Ta* (Time between N_0 and N_1), including *Min*, *1stQu*, *Median*, *Mean*, *3rdQu* and *Max*. Calculated from MSVAR outputs using RStudio for a) *Amphinemura sulcicollis*; b) *Isoperla grammatica* and c) *Baetis rhodani*. Grey = Runs which were considered outliers, **bold** = mean values used to plot Figures 5.2 - 5.4.

Table L3a. *Amphinemura sulcicollis*.

Site			Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
104	N_0	Min.	-9.94	-29.39	-16.47	-10.80	-28.40	-28.23	-19.70	-12.83	-23.41	-24.41
		1stQu.	1.86	-2.44	1.55	1.15	-1.91	-2.10	1.97	0.92	-0.34	-0.58
		Median	3.63	1.72	4.52	3.07	2.77	2.73	5.35	3.06	3.97	3.32
		Mean	**3.63**	**2.22**	**4.78**	**3.07**	**3.14**	**3.09**	**5.47**	**3.00**	**4.51**	**3.31**
		3rdQu.	5.39	6.13	7.64	4.97	7.67	7.71	8.75	5.06	8.67	7.15
		Max	19.96	43.88	39.52	21.45	42.13	41.69	35.81	25.35	49.13	42.09
	N_1	Min.	-14.18	-35.03	-24.13	-15.15	-33.25	-31.06	-23.99	-15.67	-37.70	-40.67
		1stQu.	3.32	2.02	3.77	3.66	0.26	2.18	3.87	3.69	0.91	0.25
		Median	5.11	6.54	6.68	5.59	5.21	7.03	7.02	5.73	5.22	4.58
		Mean	**5.07**	**6.15**	**6.68**	**5.56**	**5.03**	**7.14**	**7.08**	**5.75**	**4.89**	**4.13**
		3rdQu.	6.88	10.79	9.59	7.53	9.98	12.06	10.21	7.86	9.34	8.62
		Max	22.10	42.32	31.33	24.54	46.70	48.12	37.07	21.00	43.96	38.51
	Ta	Min.	-15.93	-44.19	-29.29	-15.12	-37.50	-37.55	-30.69	-16.21	-39.12	-36.58
		1stQu.	1.57	-2.48	-0.03	1.02	-1.14	0.88	0.24	0.42	-3.21	1.22

Site			Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
		Median	3.45	2.90	3.42	3.12	6.62	8.26	4.52	2.72	3.11	5.90
		Mean	**3.61**	**3.89**	**3.28**	**3.38**	**6.20**	**8.02**	**4.73**	**2.82**	**3.46**	**6.91**
		3rdQu.	5.41	9.84	6.73	5.37	14.06	15.66	8.90	5.00	9.97	11.86
		Max	23.90	45.42	40.18	24.37	49.31	53.27	41.55	24.53	48.07	50.77
96	N_0	Min.	-9.59	-23.05	-17.03	-8.88	-25.45	-23.51	-16.09	-7.78	-21.93	-26.69
		1stQu.	1.85	-0.04	1.80	1.90	-0.72	0.72	2.12	1.94	-0.26	-0.48
		Median	3.57	3.69	4.54	3.62	2.96	4.21	4.92	3.64	3.35	3.61
		Mean	**3.57**	**3.85**	**4.52**	**3.61**	**3.11**	**4.13**	**4.89**	**3.65**	**3.28**	**3.88**
		3rdQu.	5.30	7.48	7.29	5.32	6.69	7.67	7.70	5.34	6.84	7.81
		Max	18.66	43.67	22.98	13.67	41.28	24.73	32.92	19.28	26.66	41.21
	N_1	Min.	-13.90	-37.70	-15.05	-7.02	-31.65	-28.86	-19.08	-6.59	-29.46	-34.66
		1stQu.	3.16	1.28	3.17	3.15	2.14	2.36	3.56	3.18	1.05	0.64
		Median	4.89	4.95	5.91	4.90	5.76	5.88	6.36	4.88	4.73	4.79
		Mean	**4.88**	**4.87**	**5.92**	**4.89**	**5.73**	**5.89**	**6.39**	**4.88**	**4.59**	**4.54**
		3rdQu.	6.62	8.62	8.68	6.58	9.37	9.41	9.18	6.57	8.30	8.80
		Max	25.49	40.60	28.02	18.70	46.75	43.30	38.76	15.29	37.05	38.82
	Ta	Min.	-14.51	-37.72	-20.84	-9.22	-36.13	-17.90	-21.88	-15.64	-28.32	-35.78
		1stQu.	2.09	-0.03	2.06	2.20	-1.47	1.71	2.86	2.19	0.70	0.04
		Median	3.88	4.07	4.90	3.94	2.55	5.27	5.70	3.91	4.45	5.07
		Mean	**3.90**	**4.20**	**4.96**	**3.95**	**2.44**	**5.64**	**5.91**	**3.89**	**4.92**	**5.59**
		3rdQu.	5.67	8.23	7.77	5.66	6.48	9.01	8.64	5.63	8.47	10.81
		Max	24.10	44.26	33.17	23.69	47.07	51.97	39.49	17.25	47.67	54.05
95	N_0	Min.	-7.02	-18.94	-11.89	-8.76	-19.54	-19.08	-14.92	-9.11	-17.21	-25.33
		1stQu.	1.56	0.06	1.79	0.76	0.12	-0.90	2.25	1.15	-0.02	-4.04
		Median	3.28	3.40	4.48	2.68	3.48	2.59	4.93	2.93	3.38	-0.65
		Mean	**3.27**	**3.42**	**4.49**	**2.63**	**3.49**	**2.60**	**4.94**	**2.92**	**3.37**	**-0.65**
		3rdQu.	4.98	6.78	7.18	4.53	6.86	6.09	7.64	4.70	6.77	2.74
		Max	13.72	24.47	21.92	16.02	24.95	26.45	22.62	15.03	24.91	20.89
	N_1	Min.	-5.37	-17.57	-10.67	-5.80	-17.66	-15.68	-13.38	-5.34	-15.84	-13.56
		1stQu.	3.63	1.79	3.44	4.04	1.83	4.50	3.98	3.87	1.81	6.27
		Median	5.34	5.14	6.15	5.88	5.18	7.95	6.66	5.62	5.21	9.66
		Mean	**5.33**	**5.16**	**6.15**	**5.92**	**5.20**	**7.95**	**6.66**	**5.63**	**5.20**	**9.66**
		3rdQu.	7.04	8.52	8.84	7.78	8.56	11.40	9.35	7.39	8.59	13.05
		Max	15.68	26.04	23.83	17.65	26.52	31.32	24.25	17.63	26.26	31.74
	Ta	Min.	-6.74	-18.81	-11.92	-8.06	-19.44	-18.58	-14.52	-8.76	-16.97	-24.50
		1stQu.	1.73	0.23	1.95	1.04	0.29	-0.52	2.42	1.37	0.15	-3.40
		Median	3.45	3.58	4.66	2.90	3.66	2.96	5.10	3.14	3.57	0.00
		Mean	**3.44**	**3.60**	**4.66**	**2.88**	**3.67**	**2.96**	**5.11**	**3.14**	**3.55**	**0.00**
		3rdQu.	5.16	6.96	7.37	4.74	7.04	6.44	7.81	4.91	6.94	3.37
		Max	13.82	24.88	22.02	16.39	25.06	26.51	22.50	15.13	24.44	21.45
93	N_0	Min.	-6.63	-16.69	-13.32	-8.66	-17.93	-16.48	-10.91	-6.75	-17.00	-16.67
		1stQu.	1.81	0.14	1.81	1.79	0.12	1.12	2.26	1.87	0.18	0.36
		Median	3.51	3.49	4.51	3.49	3.47	4.50	4.97	3.52	3.54	3.70
		Mean	**3.51**	**3.49**	**4.51**	**3.49**	**3.46**	**4.47**	**4.97**	**3.53**	**3.55**	**3.71**
		3rdQu.	5.19	6.83	7.21	5.20	6.83	7.83	7.68	5.20	6.91	7.06
		Max	14.24	23.82	20.95	14.21	25.99	27.09	22.67	13.41	26.31	24.45
	N_1	Min.	-4.88	-14.97	-11.93	-7.27	-16.36	-14.94	-9.25	-5.45	-15.45	-14.61

Site			Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
		1stQu.	3.41	1.73	3.39	3.42	1.89	2.74	3.89	3.46	1.77	1.97
		Median	5.10	5.09	6.09	5.13	5.21	6.10	6.59	5.12	5.13	5.29
		Mean	**5.10**	**5.08**	**6.09**	**5.13**	**5.21**	**6.09**	**6.59**	**5.13**	**5.15**	**5.32**
		3rdQu.	6.79	8.43	8.80	6.83	8.59	9.44	9.30	6.80	8.51	8.67
		Max	15.84	25.08	22.89	15.71	29.15	28.71	24.23	15.34	27.59	25.97
	Ta	Min.	-6.29	-	-	-8.28	-	-	-	-6.17	-	-16.48
				16.72	13.07		17.66	16.41	10.51		16.47	
		1stQu.	2.02	0.35	2.05	2.02	0.34	1.36	2.49	2.08	0.40	0.58
		Median	3.72	3.71	4.75	3.72	3.71	4.73	5.20	3.74	3.75	3.91
		Mean	**3.72**	**3.70**	**4.75**	**3.72**	**3.69**	**4.71**	**5.20**	**3.75**	**3.77**	**3.93**
		3rdQu.	5.41	7.05	7.45	5.43	7.07	8.06	7.91	5.43	7.13	7.28
		Max	14.47	24.11	21.58	14.44	26.35	27.29	22.92	13.91	26.25	24.57
Cluster 3	N_0	Min.	-6.73	-	-	-7.21	-	-	-	-	-	-18.01
				20.23	11.78		20.71	17.27	15.25	10.02	18.93	
		1stQu.	1.70	-2.16	1.69	1.68	-0.07	0.19	0.80	1.14	1.20	0.14
		Median	3.40	1.39	4.38	3.36	3.30	3.70	3.52	2.87	4.64	3.53
		Mean	**3.40**	**1.39**	**4.39**	**3.37**	**3.30**	**3.68**	**3.53**	**2.85**	**4.66**	**3.54**
		3rdQu.	5.11	4.92	7.09	5.07	6.65	7.15	6.27	4.57	8.13	6.92
		Max	13.54	22.62	21.83	14.68	25.55	25.48	24.90	13.70	31.14	26.86
	N_1	Min.	-5.51	-	-	-4.86	-	-	-	-5.15	-	-16.11
				16.94	10.27		18.95	15.51	10.82		17.97	
		1stQu.	3.58	3.54	3.56	3.55	1.86	2.28	5.24	3.98	2.42	2.14
		Median	5.27	7.03	6.25	5.25	5.23	5.79	7.99	5.69	5.80	5.54
		Mean	**5.27**	**7.04**	**6.25**	**5.24**	**5.22**	**5.76**	**7.99**	**5.69**	**5.82**	**5.55**
		3rdQu.	6.97	10.57	8.96	6.93	8.58	9.22	10.72	7.40	9.23	8.92
		Max	15.92	28.44	24.31	16.00	27.86	27.69	30.50	16.05	30.67	28.81
	Ta	Min.	-7.95	-	-	-6.96	-	-	-	-	-	-18.08
				19.73	11.46		20.65	17.28	14.86	10.02	21.71	
		1stQu.	1.66	-1.83	1.79	1.77	0.03	2.07	0.93	1.27	-2.62	0.23
		Median	3.37	1.69	4.48	3.46	3.41	5.93	3.65	2.98	0.83	3.62
		Mean	**3.36**	**1.69**	**4.48**	**3.47**	**3.40**	**6.19**	**3.67**	**2.97**	**0.83**	**3.64**
		3rdQu.	5.09	5.18	7.20	5.16	6.75	10.08	6.40	4.67	4.27	7.01
		Max	13.76	22.65	22.02	14.49	25.70	36.34	25.21	13.60	24.70	26.96
Cluster 4	N_0	Min.	-6.88	-	-	-6.67	-	-	-	-7.14	-	-21.01
				20.74	14.03		17.23	18.06	12.43		21.62	
		1stQu.	1.66	0.00	1.69	1.67	-0.07	1.06	2.01	1.66	-0.10	-1.03
		Median	3.35	3.36	4.41	3.37	3.30	4.41	4.74	3.35	3.29	2.52
		Mean	**3.35**	**3.36**	**4.40**	**3.36**	**3.31**	**4.40**	**4.73**	**3.36**	**3.28**	**2.49**
		3rdQu.	5.04	6.72	7.12	5.05	6.69	7.76	7.43	5.04	6.63	6.00
		Max	14.59	23.12	23.21	14.97	24.49	26.55	22.48	14.15	28.47	26.41
	N_1	Min.	-5.48	-	-	-4.96	-	-	-	-5.90	-	-14.82
				19.09	11.99		15.75	16.36	10.41		20.09	
		1stQu.	3.43	1.73	3.47	3.43	1.86	2.80	3.96	3.40	1.65	2.97
		Median	5.12	5.09	6.20	5.13	5.24	6.16	6.69	5.09	5.03	6.47
		Mean	**5.13**	**5.10**	**6.18**	**5.12**	**5.24**	**6.16**	**6.68**	**5.10**	**5.02**	**6.47**
		3rdQu.	6.81	8.46	8.91	6.80	8.63	9.51	9.38	6.79	8.38	9.97
		Max	16.05	25.18	24.65	17.03	26.02	28.44	24.24	15.78	30.33	29.19
	Ta	Min.	-6.68	-	-	-6.43	-	-	-	-7.24	-	-20.34
				20.35	13.47		16.78	17.85	12.27		21.64	
		1stQu.	2.02	0.35	2.04	2.03	0.28	1.41	2.38	2.02	0.25	-0.55
		Median	3.71	3.71	4.77	3.73	3.65	4.77	5.11	3.71	3.66	2.97
		Mean	**3.72**	**3.72**	**4.76**	**3.72**	**3.66**	**4.77**	**5.09**	**3.72**	**3.64**	**2.95**
		3rdQu.	5.41	7.08	7.48	5.41	7.04	8.14	7.80	5.40	7.00	6.43
		Max	14.87	23.25	23.54	15.41	24.96	27.22	22.63	14.41	29.16	26.82

Table L3b. *Isoperla grammatica*.

Site			Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
108	N_0	Min.	-9.97	-21.31	-13.79	-7.06	-17.82	-17.47	-20.51	-8.93	-20.59	-18.24
		1stQu.	0.51	-0.18	1.71	1.82	0.05	0.08	0.47	1.47	-1.53	0.18
		Median	2.25	3.34	4.46	3.54	3.48	3.50	3.44	3.30	1.98	3.61
		Mean	**2.26**	**3.29**	**4.46**	**3.54**	**3.47**	**3.52**	**3.37**	**3.35**	**1.96**	**3.59**
		3rdQu.	4.04	6.80	7.21	5.25	6.87	6.93	6.37	5.18	5.48	7.01
		Max	12.96	23.37	21.12	14.51	25.77	25.88	24.55	17.32	23.99	25.43
	N_1	Min.	-4.56	-18.01	-10.66	-4.78	-16.79	-13.85	-17.75	-7.07	-13.93	-14.87
		1stQu.	5.31	2.40	4.34	4.24	2.67	4.67	5.38	4.85	3.84	3.35
		Median	7.08	5.84	7.06	5.95	6.05	8.08	8.48	6.65	7.42	6.80
		Mean	**7.08**	**5.85**	**7.05**	**5.94**	**6.05**	**8.08**	**8.37**	**6.66**	**7.41**	**6.80**
		3rdQu.	8.84	9.30	9.77	7.64	9.44	11.49	11.54	8.45	10.97	10.21
		Max	18.70	27.87	24.83	17.14	29.13	30.57	36.30	18.33	30.81	29.52
	Ta	Min.	-10.75	-21.95	-14.08	-7.66	-18.90	-18.58	-14.09	-14.92	-19.21	-18.96
		1stQu.	-0.18	-0.33	0.78	0.88	-0.85	-1.08	0.67	-0.59	-1.25	-1.03
		Median	1.56	3.20	3.50	2.58	2.58	2.33	3.92	1.34	2.21	2.42
		Mean	**1.58**	**3.45**	**3.50**	**2.58**	**2.56**	**2.35**	**4.91**	**1.27**	**2.19**	**2.42**
		3rdQu.	3.35	6.92	6.21	4.29	5.98	5.78	7.90	3.21	5.67	5.85
		Max	12.33	31.43	20.53	13.76	23.96	23.49	40.96	13.34	23.89	24.85
93	N_0	Min.	-133000.00	-19.17	-12.32	-6.98	-17.25	-17.17	-11.02	-6.99	-21.23	-15.67
		1stQu.	1.91	0.18	1.89	1.92	-0.05	0.98	2.51	2.35	-4.05	-0.35
		Median	3.60	3.54	4.62	3.60	3.36	4.39	5.22	4.26	-0.54	3.04
		Mean	**3.38**	**3.55**	**4.61**	**3.60**	**3.36**	**4.34**	**5.22**	**4.41**	**-0.31**	**3.03**
		3rdQu.	5.31	6.91	7.30	5.29	6.77	7.64	7.93	6.34	3.17	6.40
		Max	315500.00	25.06	21.01	14.47	23.79	26.86	21.99	17.49	34.17	24.75
	N_1	Min.	-31380.00	-17.18	-9.87	-5.23	-15.45	-15.48	-9.28	-5.63	-10.17	-12.30
		1stQu.	3.92	2.32	3.91	3.96	2.49	3.50	4.45	3.77	6.79	4.71
		Median	5.60	5.66	6.63	5.63	5.90	6.91	7.16	5.49	10.15	8.07
		Mean	**2.95**	**5.68**	**6.62**	**5.64**	**5.90**	**6.89**	**7.16**	**5.49**	**10.19**	**8.08**
		3rdQu.	7.32	9.04	9.32	7.33	9.31	10.22	9.86	7.22	13.61	11.44
		Max	38210.00	26.88	23.29	16.04	25.76	28.94	23.70	15.63	31.93	29.92
	Ta	Min.	-33330.00	-20.27	-12.74	-7.19	-17.73	-17.63	-11.49	-9.16	-36.03	-17.45
		1stQu.	1.42	-0.33	1.38	1.41	-0.57	0.60	2.05	0.83	-5.32	-0.65
		Median	3.11	3.03	4.12	3.10	2.85	4.00	4.77	2.67	-1.54	2.72
		Mean	**11.88**	**3.04**	**4.10**	**3.10**	**2.85**	**3.97**	**4.77**	**2.62**	**-2.02**	**2.74**
		3rdQu.	4.82	6.41	6.80	4.78	6.25	7.29	7.49	4.45	2.08	6.14
		Max	530400.00	24.63	20.90	13.77	23.74	26.61	21.70	13.70	20.10	25.01
Cluster 2	N_0	Min.	-6.78	-7.64	-11.72	-7.31	-14.88	-16.82	-11.81	-7.64	-16.37	-14.02
		1stQu.	1.45	9.41	1.19	1.50	-0.10	0.69	1.88	1.56	-0.62	2.38
		Median	3.07	13.56	3.94	3.16	3.23	4.01	4.53	3.27	2.82	5.78
		Mean	**3.10**	**13.56**	**3.96**	**3.18**	**3.22**	**4.08**	**4.55**	**3.29**	**2.78**	**5.80**
		3rdQu.	4.76	17.77	6.68	4.83	6.59	7.49	7.24	5.02	6.21	9.20
		Max	13.48	34.66	18.41	12.88	26.71	22.88	19.18	12.84	20.89	23.15
	N_1	Min.	-3.56	-15.62	-7.66	-3.95	-12.65	-13.22	-8.96	-4.67	-12.07	-13.53
		1stQu.	4.41	0.70	4.79	4.25	2.61	3.61	4.90	4.13	3.01	1.85

Site			Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9	Run10
		Median	6.03	4.29	7.55	5.88	5.96	6.94	7.58	5.84	6.46	5.35
		Mean	**6.06**	**4.25**	**7.51**	**5.91**	**6.01**	**7.00**	**7.59**	**5.87**	**6.43**	**5.28**
		3rdQu.	7.74	7.79	10.20	7.57	9.38	10.41	10.29	7.56	9.85	8.74
		Max	16.64	24.79	22.77	15.72	30.40	26.43	23.21	15.64	24.73	23.58
	Ta	Min.	-7.01	-24.88	-11.81	-7.41	-15.48	-16.60	-11.81	-8.09	-16.35	-18.95
		1stQu.	1.33	-7.41	0.31	1.35	-0.68	0.63	1.57	1.20	-0.73	-4.16
		Median	2.96	-3.78	3.09	2.99	2.64	3.94	4.23	2.91	2.72	-0.70
		Mean	**2.98**	**-3.80**	**3.08**	**3.00**	**2.66**	**4.02**	**4.26**	**2.93**	**2.67**	**-0.70**
		3rdQu.	4.65	-0.16	5.78	4.66	6.04	7.42	6.95	4.65	6.14	2.73
		Max	13.56	17.81	18.48	12.76	26.38	22.43	19.66	13.00	20.27	18.68

Table L3c. *Baetis rhodani*.

Site			Run1	Run2	Run3	Run4	Run5	Run6	Run7	Run8	Run9
102	N_0	Min.		-15.97			-21.82	-30.02			
		1stQu.		0.219			-1.452	1.118			
		Median		3.591			1.945	4.525			
		Mean		**3.599**			**1.941**	**4.52**			
		3rdQu.		6.98			5.35	7.951			
		Max		26.05			23.7	28.48			
	N_1	Min.		-14.66			-17.26	-23.52			
		1stQu.		1.916			3.454	2.769			
		Median		5.289			6.832	6.179			
		Mean		**5.295**			**6.845**	**6.199**			
		3rdQu.		8.675			10.26	9.623			
		Max		26.9			28.18	35.58			
	Ta	Min.		-16.21			-21.53	-17.85			
		1stQu.		0.09532			-1.701	1.168			
		Median		3.473			1.697	4.587			
		Mean		**3.48**			**1.713**	**4.687**			
		3rdQu.		6.87			5.143	8.062			
		Max		25.79			23.57	38.69			
94	N_0	Min.	-10.42	-27.41	-19.49	-10.06	-27.58	-31.17	-19.09	-10.25	-25.65
		1stQu.	-0.3777	-2.906	0.3725	0.9192	-1.881	0.2134	0.06082	1.594	1.274
		Median	1.703	1.923	4.417	2.828	3.365	5.324	3.721	3.485	6.258
		Mean	**1.772**	**2.532**	**4.96**	**2.86**	**4.004**	**5.661**	**4.425**	**3.588**	**6.882**
		3rdQu.	3.833	7.215	8.932	4.739	9.251	10.73	8.015	5.423	11.95
		Max	19.47	46.9	40.57	21.43	44.78	46.45	35.18	24.36	49.44
	N_1	Min.	-15.12	-41.11	-33.9	-16.15	-39.69	-32.68	-23.96	-16.26	-31.5
		1stQu.	4.25	-0.6633	2.572	3.99	0.4418	2.751	4.911	3.349	1.443
		Median	6.63	4.907	6.503	5.911	5.295	7.082	8.452	5.305	5.589
		Mean	**6.363**	**4.715**	**6.369**	**5.871**	**5.22**	**7.047**	**8.201**	**5.182**	**5.361**
		3rdQu.	8.852	10.26	10.32	7.809	10.04	11.43	11.77	7.223	9.615
		Max	25.48	43.74	37.06	24.91	47.21	47.85	38.25	21.41	40.28
	Ta	Min.	-16.02	-40.58	-30.04	-16.03	-40.14	-42.93	-31.82	-15.16	-41.39
		1stQu.	-0.4021	0.5151	-1.803	0.5398	-3.416	-4.616	-1.941	0.6589	-7.681
		Median	1.979	9.695	5.66	2.589	4.603	2.876	2.642	2.936	-0.8796
		Mean	**2.808**	**8.169**	**5.295**	**2.919**	**5.01**	**3.444**	**3.77**	**3.162**	**-0.2078**
		3rdQu.	5.195	16.46	12.5	4.838	13.58	11.14	9.041	5.346	6.388
		Max	26.49	47.09	46.48	24.63	49.24	45.92	40.61	26.82	46.12

‘"Er ... Good morning, O Deep Thought," said Loonquawl nervously, "do you have ... er, that is..."
"An answer for you?" interrupted Deep Thought majestically. "Yes. I have."
The two men shivered with expectancy. Their waiting had not been in vain.
"There really is one?" breathed Phouchg.
"There really is one," confirmed Deep Thought.
"To Everything? To the great Question of Life, the Universe and Everything?"
"Yes."
Both of the men had been trained for this moment, their lives had been a preparation for it, they had been selected at birth as those who would witness the answer, but even so they found themselves gasping and squirming like excited children.
"And you're ready to give it to us?" urged Loonquawl.
"I am."
"Now?"
"Now," said Deep Thought.
They both licked their dry lips.
"Though I don't think," added Deep Thought, "that you're going to like it."
"Doesn't matter!" said Phouchg. "We must know it! Now!"
"Now?" inquired Deep Thought.
"Yes! Now..."
"Alright," said the computer and settled into silence again. The two men fidgeted. The tension was unbearable.
"You're really not going to like it," observed Deep Thought.
"Tell us!"
"Alright," said Deep Thought. "The Answer to the Great Question..."
"Yes...!"
"Of Life, the Universe and Everything..." said Deep Thought.
"Yes...!"
"Is." said Deep Thought, and paused.
"Yes...!"
"Is."
"Yes...!!!...?"

"Forty-two," said Deep Thought, with infinite majesty and calm.’

Douglas Adams, Hitchhikers guide to the galaxy