View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Hochschulschriftenserver - Universität Frankfurt am Main

using science to create a better place

Using molecular tools to differentiate closely related blackfly species of the genus *Simulium*

Science Report – SC040077

ea/br/e/sci/v1

SCHO0108BNQD-E-P

The Environment Agency is the leading public body protecting and improving the environment in England and Wales.

It's our job to make sure that air, land and water are looked after by everyone in today's society, so that tomorrow's generations inherit a cleaner, healthier world.

Our work includes tackling flooding and pollution incidents, reducing industry's impacts on the environment, cleaning up rivers, coastal waters and contaminated land, and improving wildlife habitats.

This report is the result of research commissioned and funded by the Environment Agency's Science Programme.

Published by:
Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury, Bristol, BS32 4UD
Tel: 01454 624400 Fax: 01454 624409
www.environment-agency.gov.uk

ISBN: 978-1-84432-853-6

© Environment Agency January 2008

All rights reserved. This document may be reproduced with prior permission of the Environment Agency.

The views and statements expressed in this report are those of the author alone. The views or statements expressed in this publication do not necessarily represent the views of the Environment Agency and the Environment Agency cannot accept any responsibility for such views or statements.

This report is printed on Cyclus Print, a 100% recycled stock, which is 100% post consumer waste and is totally chlorine free. Water used is treated and in most cases returned to source in better condition than removed.

Further copies of this report are available from:
The Environment Agency's National Customer Contact Centre by emailing:
enquiries@environment-agency.gov.uk
or by telephoning 08708 506506.

Author(s):
Day, J.C.

Keywords:
Molecular taxonomy, Simuliidae, DNA barcodes, mitochondrial DNA, cytochrome c oxidase I

Dissemination status:
Released to all regions
Publicly available

Research Contractor:
Centre for Ecology and Hydrology (CEH) – Oxford
Mansfield Road
Oxford
Oxfordshire
OX1 3SR
Tel: 01865 281 630

Environment Agency's Project Manager:
Kerry Walsh,
Science Department

Science Project Number:
SC040077

Product code: SCHO0108BNQD-E-P

Science at the Environment Agency

Science underpins the work of the Environment Agency. It provides an up-to-date understanding of the world about us and helps us to develop monitoring tools and techniques to manage our environment as efficiently and effectively as possible.

The work of the Environment Agency's Science Department is a key ingredient in the partnership between research, policy and operations that enables the Environment Agency to protect and restore our environment.

The science programme focuses on five main areas of activity:

- **Setting the agenda**, by identifying where strategic science can inform our evidence-based policies, advisory and regulatory roles;
- **Funding science**, by supporting programmes, projects and people in response to long-term strategic needs, medium-term policy priorities and shorter-term operational requirements;
- **Managing science**, by ensuring that our programmes and projects are fit for purpose and executed according to international scientific standards;
- **Carrying out science**, by undertaking research – either by contracting it out to research organisations and consultancies or by doing it ourselves;
- **Delivering information, advice, tools and techniques**, by making appropriate products available to our policy and operations staff.

Steve Killeen

Head of Science

Executive summary

Biodiversity data are the foundation for conservation and management and taxonomy provides the reference system, skills and tools used to identify organisms. Species level data such as species richness, composition and diversity are common metrics. However, species level identification of organisms tends to be neglected within ecological work, especially within monitoring programmes, but also in conservation biology (Giangrande, 2003). This is because collection of species level data is time consuming, with identification of species-specific characteristics traditionally involving lengthy examination of samples using microscopy. In addition it is costly and species level data is almost impossible to collect if the taxa involved are species rich and difficult to identify (Báldi 1999). Other reasons why species level identification is neglected include the fact that sample collection can damage organisms, so diagnostic morphological features are lost, or that individuals may be in a life history stage or of a sex that does not have diagnostic morphological characteristics. Furthermore, the numbers of available expert taxonomists needed for species identification are in decline and have been for several decades.

Species identification using molecular taxonomy where DNA is used as a marker is championed as a tool for resolving a range of morphological problems, such as the association of all life history stages, correlating male and female specimens to the same species and identifying partial specimens. Traditional taxonomy is built around morphological variations between species, with systematic inferences based upon shared physical characters. In molecular taxonomy on the other hand, proteins and genes are used to determine evolutionary relationships. 'DNA barcoding' aims to provide an efficient method for species-level identification and it is thought that it will provide a powerful tool for taxonomic and biodiversity research (Hajibabaei et al. 2007).

Cited strengths of a molecular based approach to species identification include the potential universality and objective nature of DNA data as taxonomic information, the usefulness of molecular data in animal groups characterized by morphological cryptic characters and the use of DNA sequence information to determine otherwise 'unidentifiable' biological material (such as incomplete specimens or immature specimens). Its aim is to increase the speed, precision and efficiency of field studies involving diverse and difficult to identify taxa and it has the potential to be automated to provide a rapid and consistently accurate supplementary identification system to traditional taxonomy.

This project was a proof-of-concept study that investigated the feasibility of using DNA barcodes to differentiate closely related blackfly species of the genus *Simulium*. The longer term objective would be to apply such molecular approaches to organisms used in water quality monitoring and to biodiversity studies to provide a quick, robust but practical and cost effective tool for species identification.

Great Britain is currently home to 33 morphospecies of blackfly many of which are morphologically close to other species and have been the cause of much systematic revision. In addition to evaluating the use of DNA barcodes in species identification, a non-destructive DNA extraction method was developed to preserve voucher specimens that will allow a complete morphological classification to be carried after DNA extraction.

Finding an effective DNA barcode for an individual species involves accurate taxonomic identification and the retention of voucher specimens for future morphological studies. A rapid non-destructive method for DNA extraction from small insects was developed where no clean-up step was required prior to amplification and it was possible to extract DNA of sufficient quality in minutes retaining diagnostic morphological characteristics.

For any molecular tool used for species discrimination, an important consideration is defining the specific genetic loci (e.g. the position of genes on a chromosome) to be monitored. All blackfly species in this study were successfully amplified with the standard barcoding coxI gene primer pair LCO1490 5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3' and HCO2198 5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3' (Folmer *et al.* 1994) and we did not need to optimise or redesign the primer sequence.

This study has strongly confirmed the ability of coxI barcodes to identify the majority of simuliid species in Britain. We sequenced (predominantly multiple) specimens of three genera: three *Prosimulium* species, one *Metacnephia* species and 18 *Simulium* species. Of the 32 species sequenced, 26 could subsequently be discriminated (81 per cent). These were from the subgenera *Boopthora*, *Eusimulium*, *Helichelia, Prosimulium* and *Wilhelmia*, as well as *Metacnephia amphora*, and clearly barcoding would be capable of unambiguously identifying all life stages: larvae, pupae and adults, and, potentially, egg masses. However, we were unable to resolve unambiguously all species from the subgenus *Nevermannia* and *Simulium*.

The most positive support for DNA barcoding came from the subgenus *Eusimulium*. Here, all the species of the *S. aureum* group are predominantly morphologically identical and can only truly be resolved using adult morphology and cytotaxonomy of the larvae. We have shown clear demarcations of these species using DNA barcodes and have also produced a preliminary population survey of sites in Britain.

DNA barcoding was able to identify 81 per cent of the British blackfly fauna and was successful in cases where traditional taxonomy would fail. The inability to resolve all the simuliid species illustrates a problem encountered with molecular taxonomy: a lack of resolution between closely related species. For these species, genus level identification would be the highest level of resolution. In situations where closely related species cannot be resolved an additional loci (gene target(s)) will be needed for species determination.

At present, although it is not possible to advocate a molecular system that can accurately identify all members within a blackfly community en masse. It is, however, possible to identify many species in certain life history stages that cannot be resolved based on morphology. The strength of a molecular taxonomic system depends on the resolving power of the identification. Despite the limitations of DNA barcodes illustrated herein, it is also clear that such sequences can provide strong support to resolve the unsubstantiated taxonomic status of certain species. Furthermore, such an approach does not require extensive taxonomist training. DNA barcoding is a powerful tool for identifying species, but is dependent upon a taxonomic framework within which to operate.

Acknowledgements

This report would not have been possible without the collaboration of researchers and taxonomists in the laboratory and at other research organisations. They have made a concerted effort to solve many challenging problems regarding the molecular characterisation of macroinvertebrates over the past 15 months. Among them, I am especially indebted to Stephanie Hunter and Tim Goodall, who undertook the majority of the bench work presented in this report. The molecular characterisation of most species would not have been possible without the expertise and generosity of taxonomists from CEH Dorset, in particular Iwan Jones and Jon Bass, to whom I am extremely grateful for their time and patience. I am indebted to Rory Post from the Natural History Museum, who first introduced me to the world of collecting and identifying blackflies and who has been invaluable in providing expertise, as well as constant support and advice. I would also like to thank Roger Crosskey for the time spent prospecting on Stanmore Common and for patiently responding to my numerous questions. Thanks also go to Kerry Walsh, Richard Owen, Sarah Turner and Mark Bailey for assisting with the initial grant proposal and providing executive support throughout the project. This work was supported by a co-funded research grant from the Environment Agency and the Centre for Ecology and Hydrology. I am grateful to these funding agencies for their generous support.

Contents

1	Introduction	1
1.1	Molecular taxonomy and DNA barcoding	2
1.2	Objective	3
2	**A non-destructive DNA extraction: retaining voucher specimens for barcoding projects**	**4**
2.1	Introduction	4
2.2	Materials and methods	4
2.3	Results and discussion	5
3	**DNA barcoding the blackfly fauna of Britain (Diptera: Simuliidae)**	**8**
3.1	Introduction	8
3.2	Materials and methods	10
3.3	Results	11
3.4	Discussion	18
4	**Conclusion**	**23**
References		**24**
Glossary		**28**
Appendix		**31**

List of tables

TABLE 3.1 INVENTORY OF BRITISH BLACKFLIES (DIPTERA: SIMULIIDAE) 9
TABLE 3.2 MEAN PERCENTAGE BASE COMPOSITION (WITH STANDARD ERROR), COMPARING *COXI* SEQUENCES OF FIVE SIMULIID SUBGENERA 13
TABLE 3.3 GC CONTENT OF THE 1ST, 2ND AND 3RD CODON POSITIONS (WITH STANDARD ERROR) IN FIVE SIMULIID SUBGENERA 13

List of figures

Figure 2.1 Non-destruction dna extraction protocol 15
Figure 3.1 neighbor-joining tree of 256 coxi sequences from 32 blackfly species, using k2p distances 20
Figure 3.2 overlap between intra- and interspecific genetic variability for congruent sequences 21
Figure 3.3 k2p distance neighbor-joining tree of 45 coxi sequences from 10 species belonging to the subgenus simulium 23
Figure 3.4 k2p distance neighbor-joining tree of 28 coxi sequences from four members of the simulium aureum group belonging to the subgenera eusimulium 24
Figure 3.5 k2p distance neighbor-joining tree of 48 coxi sequences from eight species belonging to the subgenera nevermannia 25

1 Introduction

The Environment Agency has an important role to play in ecology and conservation. Despite improvements in both water and air quality over the last decade some of our wildlife is still under threat with subsequent loss in biodiversity and impacts on ecosystem functioning. Biodiversity loss and the speed at which it is occurring is a global issue with many biologists convinced that a "mass extinction" of plants and animals will take place during the next century. This phenomenon has been termed the “biodiversity crisis”. Biodiversity is threatened by habitat destruction, introduction of alien species, over exploitation of natural resources and climate change. Understanding and controlling such environmental impacts and preventing species decline are important aims of biodiversity conservation and management. Specifically the Environment Agency needs to protect the environment and promote conservation. To do this, understanding the ecological needs of many species, some of which are very important in highlighting the state of the environment around them, is vital.

Biodiversity data are the foundation for conservation and management and taxonomy provides the reference system and the skills and tools used to identify organisms. Scientifically there is no single most appropriate measure of biodiversity, although species richness, composition and diversity are common metrics. Species level identification, however, tends to be neglected within ecological work, especially within monitoring programmes, but also in conservation biology (Giangrande, 2003). This is because collection of species level data is time consuming, with identification of species-specific characteristics traditionally involving lengthy examination of samples using microscopy. In addition it is costly and species level data are almost impossible to collect if the taxa involved are species rich and difficult to identify (Báldi 1999).

As a time and cost effective alternative, the use of surrogate higher taxa has been adopted in many ecological studies. Here, taxa are identified to the highest possible category (genus, family, etc.) without loss of important information to detect changes in assemblages exposed to environmental stress. This is known as ‘taxonomic sufficiency’, a concept introduced by Ellis (1985), but its use in biodiversity and conservation scenarios has often been contentious (Báldi, 2003; De Biasi et al. 2003; Giangrande, 2003; Terlizzi et al. 2003; Heino & Soininen, 2007). According to Giangrande (2003), the implications of taxonomic sufficiency in conservation biology can lead to inaccuracy of biodiversity evaluation such as the *a priori* exclusion of some entities before their role in ecology has been understood.

Other reasons why species level identification is neglected include the fact that sample collection can damage organisms so diagnostic morphological features are lost, or that individuals may be in a life history stage or of a sex that does not have diagnostic morphological characters to separate it from other species. Furthermore, there has been a decline in the number of expert taxonomists available for species identification. Governments, through the Convention on Biological Diversity (CBD), have acknowledged the existence of a "taxonomic impediment" to the sound management of biodiversity (see

http://www.cbd.int/gti/default.shtml). This term is used to describe the gaps in knowledge in our taxonomic system, the shortage of trained taxonomists and curators and the impact these deficiencies have on our ability to manage and conserve biological diversity. The taxonomic impediment represents a major problem in light of the biodiversity crisis and the greater need for biodiversity studies. The Global Taxonomy Initiative (GTI), set up under the CBD, provides a framework to address the taxonomic impediment. Although existing taxonomic practices have served us well for many years they are clearly inadequate for the challenges that lie ahead in biodiversity and conservation management. One way to overcome the taxonomic impediment, although highly controversial (see Lipscomb et al., 2003; Seberg et al., 2003; Moritz & Cicero, 2004 and Ebach and Holdredge, 2005), is to facilitate species identification using molecular taxonomy where DNA is used as a marker for species identification. This concept, known as DNA Barcoding, aims to provide an efficient method for species-level identification and it is thought that it will provide a powerful tool for taxonomic and biodiversity research (Hajibabaei et al. 2007).

1.1 Molecular taxonomy and DNA barcoding

Traditional taxonomy is built around morphological variations between species, with systematic inferences based upon shared physical characters. In molecular taxonomy, on the other hand, proteins and genes are used to determine evolutionary relationships. It has long been accepted that DNA sequence diversity can be used to discriminate species, with mitochondrial DNA approaches, in particular, dominating molecular systematics since the late 1970s (Avise 1994). The percentage of DNA that is similar in different species is taken into consideration and compared with the morphological characters. Animal genomes are huge and present a potentially endless supply of evolutionary and systematic information.

Molecular taxonomy has a long history in morphologically ‘simple’ groups such as bacteria, viruses and fungi. Recently, several authors have argued that a similar model might profitably be used in the animal world (Baker *et al.* 2003, Blaxter and Floyd 2003, Hebert *et al.* 2003, Proudlove and Wood 2003, Tautz *et al.* 2002). Cited strengths include the potential universality and objective nature of DNA data as taxonomic information, the usefulness of molecular data in animal groups characterized by morphological cryptic characters and the use of DNA sequence information to determine otherwise ‘unidentifiable’ biological material (such as incomplete specimens or immature specimens).

Tautz *et al.* made the case for a DNA-based taxonomic system, proposing that it ‘will provide a new scaffold for our accumulated taxonomic knowledge and a reliable tool for species identification and description’ (Tautz *et al.* 2002). The main problem besetting a molecular-based taxonomic approach is the variety of markers employed by different laboratories for different taxonomic groups. It was Hebert *et al*. (2003) who recently proposed consolidating the disparity of genetic loci to a single gene sequence that would be sufficient to differentiate the vast majority of metazoan species. The cytochrome c oxidase subunit 1 (coxI) mitochondrial DNA gene has been advocated as the global bioidentification inventory system for animals. It is approximately 650 nucleotide base pairs (bp) long in most groups; a very short sequence compared with the 3 billion bp in the

human genome. These inventory systems could be used to estimate genetic variation and species diversity, even where no prior morphologically discriminating features are available.

DNA barcodes, as currently applied, are species identifiers, and empirical support for the barcoding concept comes from pilot projects involving invertebrates, fish and birds (Hebert *et al.* 2004a, Hebert *et al.* 2004b, Smith *et al.* 2005, Ward *et al.* 2005b). This discrimination of selected taxa assumes that DNA variation within species is much lower (10 fold or less) than between species (Hebert *et al.* 2004b). In addition to the pilot projects above, DNA barcoding systems are now being established for plants, macroalgae, fungi and protists (see Hajibabaei et al. 2007 and references there in). The utility of the DNA barcoding approach needs to be tested more broadly, particularly in biologically complex situations where lineages are composed of closely related species or are affected by a complicated evolutionary history of gene trees (Hebert *et al.* 2004a, Hebert *et al.* 2003, Hebert *et al.* 2004b, Smith *et al.* 2005, Ward *et al.* 2005a). This project aims to assess the ability of DNA barcoding using the coxI mitochondrial DNA gene to differentiate closely related blackfly species of the genus *Simulium* and support its use in taxonomic studies.

Further information on DNA barcoding initiatives can be found at the Consortium for the Barcode of Life website at http://www.barcoding.si.edu/DNABarCoding.htm and the Canadian Centre for Barcoding website at http://www.dnabarcoding.ca/. The sites contain information on countries and personnel involved in different barcoding initiatives, laboratory protocols, links to public reference libraries of species identifiers which could be used to assign unknown specimens to known species, publications and latest news, events and developments on DNA barcoding.

1.2 Objective

This study investigated the feasibility of using DNA barcodes to differentiate closely related blackfly species of the genus *Simulium*. Great Britain is currently home to 33 morphospecies of blackfly many of which are morphologically close to other species and have been the cause of much systematic revision. In addition to evaluating the use of DNA barcodes in species identification, a non-destructive DNA extraction method was developed to preserve voucher specimens that will allow a complete morphological classification to be carried after DNA extraction.

If this project could demonstrate that molecular approaches (Environment Agency, 2003) could be used to improve species identification of difficult to identify taxa then their scope and potential application could be extended to target other organisms assessed during water quality monitoring and biodiversity surveys with the aim of reducing cost, increasing efficiency and providing more comprehensive data of improved quality.

2 A non-destructive DNA extraction: retaining voucher specimens for barcoding projects

2.1 Introduction

Finding an effective DNA barcode for an individual species involves accurate taxonomic identification and the retention of voucher specimens for future morphological studies. Once a species division has been identified, returning to such voucher specimens may yield reliable diagnostic characters, which may previously have been regarded as a morphological variation within the species. Unfortunately, extracting DNA from samples of small insects (such as Simuliidae) often destroys them, particularly in the immature stages, thus sacrificing the entire specimen in order to obtain sufficient DNA for PCR and subsequent sequencing. Therefore, linking the voucher specimen with the corresponding DNA barcode requires photo-documentation prior to DNA extraction.

Preserving the entire voucher specimen would allow a complete morphological classification to be carried out after DNA extraction. In addition, many extractions utilise noxious chemicals such as phenol or chloroform and are labour intensive. Here we present a rapid, non-destructive, chemical-free DNA extraction method, which we have used to determine coxI barcodes for the Blandford fly *Simulium posticatum* Meigen and *Simulium erythrocephalum*.

2.2 Materials and methods

Simulium posticatum was collected on 2 May 2006 from two sites in Oxfordshire: the River Cherwell at Upper Heyford (O.S. grid reference SP492260) and the River Evenlode at Long Hanborough (O.S. grid reference SP420148). *Simulium erythrocephalum* was collected on 6 October 2006 from Upper Heyford. Larvae and pupae were manually removed from trailing vegetation before being fixed in 95 per cent ethanol. In the case of *S. erythrocephalum*, eight adults were reared from pupae using a standard protocol (Davies 1968). Prior to DNA extraction, the specimens were washed in sterile distilled water and air dried. The mass of each *S. posticatum* blackfly was recorded and then the fly was submerged in 50µl of sterile distilled water in a 1.5ml Eppendorf tube and transferred to ice.

Samples were placed in a sonicating water bath (Dawe Sonicleaner, Dawe Instruments) for 30s periods and chilled on ice for 60s between consecutive sonications. A range of treatment times were tested, from one sonication (30s) to eight sonications (240s). For each treatment time, eight individual larvae and eight individual pupae were used for the

DNA extraction and the results were compared with the corresponding number of non-sonication controls.

Each set of eight *S. posticatum* specimens consisted of four representatives from each of the two populations. After sonication, the specimens were checked for the preservation of diagnostic characters, including cephalic apotome markings, the postgenal cleft of the larval head, and the gill filaments and cocoon of the pupae (Bass 1998). Voucher specimens were stored in ethanol at -20°C. The concentration and purity of the DNA was determined using a GeneQuant pro spectrophotometer (Biochrom Ltd). For *S. erythrocephalum*, eight individuals from each life history stage were used for the DNA extractions. All 24 individuals were sonicated for a total period of 2 minutes before being returned to 80 per cent ethanol and stored at -20ºC. The DNA was used directly for PCR, along with the DNA from *S. posticatum*.

A 658bp region of the coxI gene was amplified using PCR. The total reaction volume was 50µl, comprising 1µl DNA, the primers LCO1490 (5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3') and HCO2198 (5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3') (Folmer *et al.* 1994) at a concentration of 1µM each, 0.5µl dNTPs (40mM mix; Bioline), 0.5µl BSA (10mg ml-1; New England Biolabs), 5µl 10 x Taq buffer (Sigma-Aldrich) and 0.5µl Taq (Sigma-Aldrich)/PFU (Promega) (10:1 ratio) DNA polymerase mix. The thermal cycling was carried out on a PTC-225 Tetrad thermal cycler (MJ Research) as follows: 94°C for 2 minutes; 35 cycles of: 94°C for 30s, 49°C for 30s, 72°C for 2 minutes; 72°C for 10 minutes.

The presence of PCR product was determined by running samples on a 1 per cent agarose gel stained with ethidium bromide. Quantification of PCR product was carried out using a Bioanalyser and the product was purified using the Qiagen gel extraction kit (Qiagen Ltd). PCR products were cloned into the pGEM Easy T vector (Promega) and sequenced in one direction on an ABI 377 automated sequencer (Applied Biosystems) using the Big Dye v.3 sequencing kit. Sequences were edited and aligned using Sequencher 4.5 (Gene Codes Corp.) and then compared with coxI sequences previously obtained for *S. posticatum* samples from the River Stour, Spetisbury, Dorset (O.S. grid reference ST912030). These sequences were derived using a Ish-Horowicz destructive DNA extraction method (Ish-Horowicz 1982).

2.3 Results and discussion

The yield of DNA obtained from larval specimens of *S. posticatum* was clearly higher when sonication was employed, compared with the non-sonication control (Figure 2.1A). The mean yield rose from 497.2ng/mg insect in the control samples to 1326.3ng/mg insect after 30s of sonication. The optimum yield of DNA was observed after 60s of sonication (1580.3ng/mg insect), with the yield steadily decreasing for longer periods of sonication (to 1502.1ng/mg insect after 120s and 1414.7ng/mg insect after 240s).

These values, however, were not found to be significant using a standard two-tailed T-test assuming equal variance (P=0.813 and 0.619 respectively). It is suggested, therefore, that although more DNA is released for sonication times greater than 60s, the DNA already present in the samples will be more prone to sheering. As such, the total amount of DNA

available may not change significantly, due to variation in the absorbance calculations for different size ranges of DNA.

The purity of the DNA samples clearly decreased with increased sonication, although the non-sonication control proved to be an exception (Figure 2.1B). The Genequant spectrophotometer does not give a definitive purity value, but it allows comparisons between samples of a given type. The 260nm/280nm absorbance ratio has been used as an indication of the level of protein and other contaminants in the DNA samples. A value of 1.8 or above is considered to be free of contaminating protein. The purity of the DNA samples decreased from a value of 2.10 after 60s of sonication to 1.54 after 240s of sonication, further suggesting that a sonication time of 60s is optimal for this DNA extraction process. Increased sonication time allows more proteins to be released from the insect into the surrounding water, thereby decreasing the 260nm/280nm ratio.

PCR was used to amplify a 658bp region of coxI in all the *S. posticatum* DNA samples. Those samples obtained from larvae that had not been sonicated yielded no PCR product (Figure 2.1C). Thus, whilst sonication did appear to increase the likelihood of getting a PCR product, the duration of sonication did not appear to have any great effect. Indeed, although the greatest yield of DNA was obtained for the 60s treatment time, the same number of PCR products was observed for both the 30s and 240s sonication times (six out of a possible eight in all cases). Only samples obtained after 120s sonication yielded fewer PCR products (four out of eight). Sequencing indicated that the amplified fragments were indeed from Simuliidae DNA. Comparing these sequences with *S. posticatum* sequences obtained using a destructive DNA extraction method (data not shown) confirmed that the correct DNA barcode could be obtained with this non-destructive method of DNA extraction.

PCR product was obtained from sonication-extracted DNA for 60–75 per cent of the *S. erythrocephalum* specimens (Figure 3.1C). All life history stages were amplified, illustrating the versatility of this process in blackfly DNA barcoding projects. Three individuals, representing each life stage, were sequenced and were confirmed to be *S. erythrocephalum* based upon sequence identity to material collected from a second site in Oxfordshire (River Ock Bridge, Oxforfordshire, 10 April 2006; OS grid ref. SU399956). Not all specimens were amenable to this process of extraction, however, possibly due to the presence of inhibitors, the release of insufficient amounts of DNA or excessive shearing of the DNA during the sonication process.

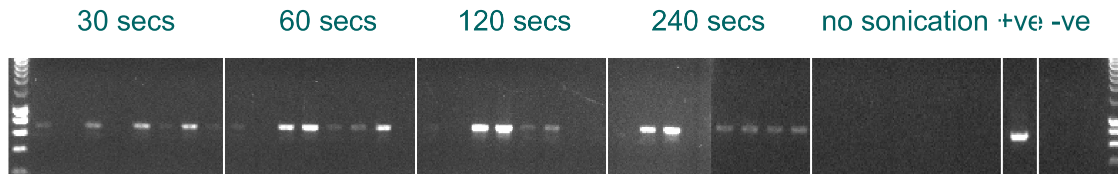
Success rates are variable with all DNA extraction protocols and we anticipate this method being employed as an initial screening for high-throughput barcoding or population genetics studies. The 70 per cent success rate for this extraction method is, we feel, offset by the speed and cost effectiveness of the process. With abundant specimens such as blackflies, this extraction method could be used as part of a rapid screen, with successful specimens selected for subsequent DNA barcode analysis.

A rapid non-destructive method for DNA extraction from small insects has been described. In essence, sonication in water allowed the release of sufficient DNA for use in barcoding exercises. No clean-up step is required prior to amplification and sequencing, as is required for another non-destructive method for extracting DNA from insects (Pons 2006), making it possible to extract DNA of sufficient quality in minutes. The post-sonication

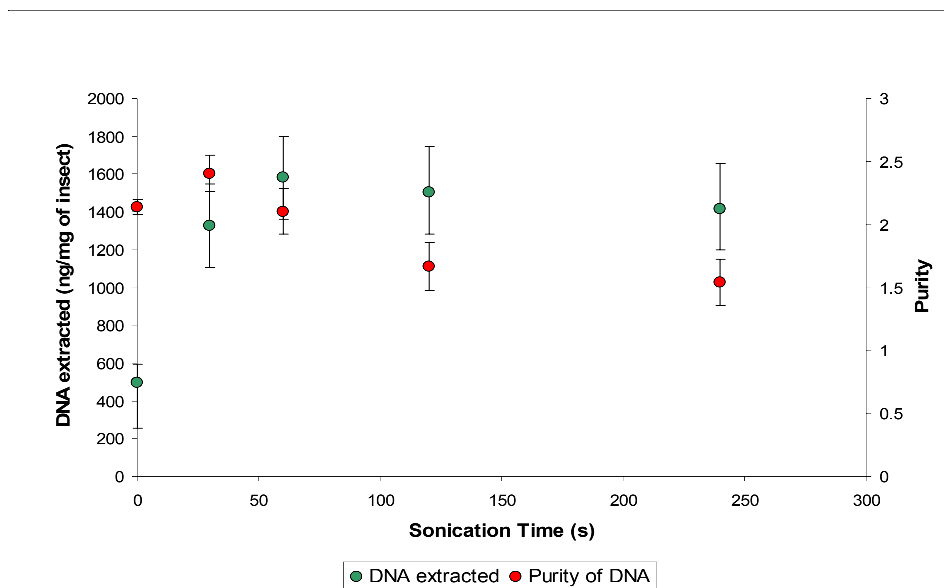
specimens are maintained intact, making it possible to identify diagnostic morphological characteristics.

It is proposed that this extraction method will allow rapid amplification and sequencing of molecular barcodes, and make it possible to associate individual barcodes with the voucher specimens from which they were obtained. It should also make it possible to confirm the classical morphological taxonomy currently in place and, indeed, determine whether cryptic or sibling species exist.

A



B



C

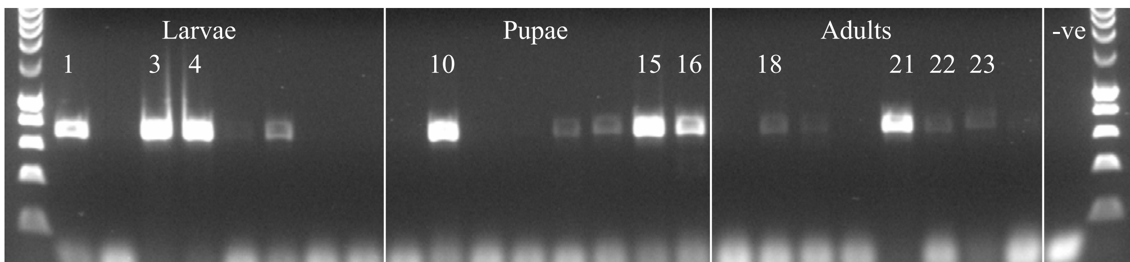


Figure 2.1 Non-destruction DNA extraction protocol
A. CoxI PCR amplification of 24 specimens of *S. posticatum* from sonication extracted DNA. Eight replicate specimens were used for five different sonication durations, from no-sonication through to 240s. B. Plot showing amount of DNA extracted for each specimen and the purity of the DNA obtained. C. CoxI PCR amplification of 24 specimens of *S. erythrocephalum* from sonication-extracted DNA. Numbers indicate specimens sequenced.

3 DNA barcoding the blackfly fauna of Britain (Diptera: Simuliidae)

3.1 Introduction

Here, we examine whether DNA barcoding can be used to discriminate the blackfly fauna of Britain. Approximately 1,800 simuliid species are listed as valid, based upon current knowledge, constituting about 2 per cent of all dipteran species. The British blackfly fauna is currently composed of 33 morphospecies and at least three cytospecies complexes (Table 3.1) (Bass 1998, Bass *et al.* 1995, Crosskey 1991, Post *et al.* 2007).

The majority of species recognised in the British Isles can be identified as adults (Davies 1966, Davies 1968), but immature stages are more problematic and morphological homoplasy is a frequent issue with these closely related species. This can impact in a number of ways. Grouping species into a higher taxonomic classification, such as species group or subgenera, can result in an underestimation of true biodiversity and, potentially, a failure to detect perturbations in community structure.

We present an evaluation of coxI profiles for 32 of the 33 species of British blackfly, most of which have been examined from multiple specimens, with the principle objective of ascertaining whether DNA barcoding using coxI gene can achieve unambiguous species identification in blackflies.

Table 3.1 Inventory of British blackflies (Diptera: Simuliidae)

	Tribe	Genus	Subgenus	Species
1	Prosimulini	*Metacnephia*	-	*Metacnephia amphora* Ladle & Bass, 1975
2	Prosimulini	*Prosimulium*	*Prosimulium*	*Prosimulium hirtipes* Fries, 1824
3	Prosimulini	*Prosimulium*	*Prosimulium*	*Prosimulium latimucro* Enderlein, 1925
4	Prosimulini	*Prosimulium*	*Prosimulium*	*Prosimulium tomosvaryi* Enderlein, 1921
5	Simulini	*Simulium*	*Hellichiella*	*Simulium latipes* Meigen, 1804
6	Simulini	*Simulium*	*Nevermannia*	*Simulium armoricanum* Doby & David, 1961
7	Simulini	*Simulium*	*Nevermannia*	*Simulium angustitarse* Lundström 1911
8	Simulini	*Simulium*	*Nevermannia*	*Simulium costatum* Friederichs, 1920
9	Simulini	*Simulium*	*Nevermannia*	*Simulium cryophilum* Rubtsov, 1959
10	Simulini	*Simulium*	*Nevermannia*	*Simulium dunfellense* Davies, 1966
11	Simulini	*Simulium*	*Nevermannia*	*Simulium juxtacrenobium* Bass & Brockhouse, 1990
12	Simulini	*Simulium*	*Nevermannia*	*Simulium lundstromi* Enderlein, 1921
13	Simulini	*Simulium*	*Nevermannia*	*Simulium urbanum* Davies, 1966
14	Simulini	*Simulium*	*Nevermannia*	*Simulium vernum* Macquart, 1826
15	Simulini	*Simulium*	*Eusimulium*	*Simulium angustipes* Edwards, 1915
16	Simulini	*Simulium*	*Eusimulium*	*Simulium aureum* Fries, 1824
17	Simulini	*Simulium*	*Eusimulium*	*Simulium velutinum* Santos Abreu, 1922
18	Simulini	*Simulium*	*Eusimulium*	*Simulium petricolum*Rivosecci 1963
19	Simulini	*Simulium*	*Wilhelmia*	*Simulium equinum* Linnaeus, 1758
20	Simulini	*Simulium*	*Wilhelmia*	*Simulium lineatum* Meigen, 1804
21	Simulini	*Simulium*	*Wilhelmia*	*Simulium pseudequinum* Séguy, 1921
22	Simulini	*Simulium*	*Boophthora*	*Simulium erythrocephalum* De Geer, 1776
23	Simulini	*Simulium*	*Simulium*	*Simulium argyreatum* Meigen, 1838
24	Simulini	*Simulium*	*Simulium*	*Simulium intermedium* Roubaud, 1906
25	Simulini	*Simulium*	*Simulium*	*Simulium ornatum* Meigen, 1818
26	Simulini	*Simulium*	*Simulium*	*Simulium posticatum* Meigen, 1838
27	Simulini	*Simulium*	*Simulium*	*Simulium morsitans* Edwards, 1915
28	Simulini	*Simulium*	*Simulium*	*Simulium noelleri* Friederichs, 1920
29	Simulini	*Simulium*	*Simulium*	*Simulium reptans* Linnaeus, 1758
30	Simulini	*Simulium*	*Simulium*	*Simulium rostratum* Lundström, 1911
31	Simulini	*Simulium*	*Simulium*	*Simulium trifasciatum* Curtis, 1839
32	Simulini	*Simulium*	*Simulium*	*Simulium tuberosum* Lundström, 1911
33	Simulini	*Simulium*	*Simulium*	*Simulium variegatum* Meigen, 1818

Compiled by Crosskey (1991), with the exception of a single revision (Bass *et al.* 1995) and the addition of a recently determined species, *S. petricolum* (Post *et al.* 2007).

3.2 Materials and methods

Blackflies were collected primarily as larvae and pupae from a range of substrates, including rocks, trailing grass or reeds and fallen leaves and branches, along rivers and streams. Larvae and pupae from the same site were pooled together as one population, fixed in ethanol and returned to the laboratory. Pupae and larvae were identified to species level where possible using the key to the immature stages (Bass 1998).

For each of the four *Eusimulium* species, an adult specimen was obtained by collecting pupae and rearing in tubes until the adults emerged and then identifying species based upon dissected genitalia (Bass 1985, Davies 1968). Members of the *S. ornatum* group (*S. intermedium*, *S. ornatum* and *S. trifasciatum*) were identified based upon pupal microtubules of the thoracic cuticle, the weave of the pupal cocoon and their geographical location. There are at least seven cytotypes within the *S. ornatum* group (Post 1980), but due to the time constraints of this study specimens were not examined chromosomally to establish cytotype and identification was only carried out to the species level. A single adult specimen of *Simulium tuberosum*, the 33rd species discovered in Britain, was found, upon analysis, to be misidentified and to be a member of the *Simulium vernum* complex.

DNA was extracted using the non-destructive, sonication method that provided whole specimen voucher material from individuals less that one month old (see Section 2). Older material was extracted using either the method of Ish-Horowicz (Ish-Horowicz 1982) or the high pure PCR template kit (Roche), depending on the number and condition of the specimens. A 658bp fragment of the mitochondrial coxI gene was amplified using the primers LCO1490 5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3' and HCO2198 5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3' (Folmer *et al.* 1994).

PCR was performed in a total volume of 25µl containing 1X reaction buffer, 2.0mM $MgCl_2$, 0.2mM of each dNTP, 0.4µM of each primer, two units of Taq DNA polymerase (Sigma) and 0.5µL of DNA sample. The thermal cycling was carried out on a PTC-225 Tetrad* thermal cycler (MJ Research) and conditions consisted of: 94°C for 2 minutes; 35 cycles of: 94°C for 30s, 49°C for 30s, 72°C for 2 minutes; 72°C for 10 minutes. The presence or absence of PCR product was determined by running samples on a 1 per cent agarose gel stained with ethidium bromide.

PCR products were cloned into the pGEM Easy T vector (Promega) and sequenced in one direction using M13F primer on an ABI 377 automated sequencer (Applied Biosystems) using the Big Dye v.3 sequencing kit. Sequences were edited and aligned using Sequencher 4.5 (Gene Codes Corp.). Sequence divergences were calculated using the Kimura two parameter (K2P) distance model (Kimura, 1980). Neighbor-joining (NJ) trees of K2P distances were created to provide a graphic representation of the patterning of divergence between species (Saitou and Nei 1987). In the three chosen subgenera of blackflies, bootstrapping was performed in MEGA3.1 *(Kumar et al.*, 2004) with 1000 replications.

3.3 Results

We present results for all 32 species, followed by more detailed examinations of three subgenera of blackflies. These subgenera are: *Nevermannia* (nine species containing the vernum-complex within which *S. naturale* has been synonymised); *Simulium* (eight species); and *Eusimulium*, which contains three members of the *S. aureum* group and the recently reported species *S. petricolum* (Post *et al.* 2007).

3.3.1 All species

A total of 32 species were analysed, and for each species up to nine specimens were examined, giving a total of 266 sequences. The full K2P/NJ tree is detailed in Appendix I and is presented in summary form as Figure 3.1.

The coxI sequences were all 658bp long with no insertions, deletions or stop codons observed in any sequence. A continuous open reading frame is consistent with all amplified coxI sequences being ascribed functional mitochondrial gene status. Together with the fact that all amplified sequences were 658bp in length, this suggests that NUMTs (nuclear DNA sequences originating from mitochondrial DNA sequences) were not a part of our dataset (NUMTS are typically smaller than 600bp) (Zhang and Hewitt 1996).

The average K2P distance of individuals within species was 0.86 per cent, compared with 15.1 per cent between species (Figure 3.1). Overall, therefore, there was around 17 times more variation among congeneric species than among conspecific individuals. Mean divergence among species within subgenera increased to 4.24 per cent.

Nucleotide composition averaged over all specimens showed an A–T bias (A = 27.7 per cent, T = 36.5 per cent, C = 18.8 per cent, G = 17.0 per cent). No significant variation was exhibited in the GC content of different subgenera (Tables 3.2 and 3.3).

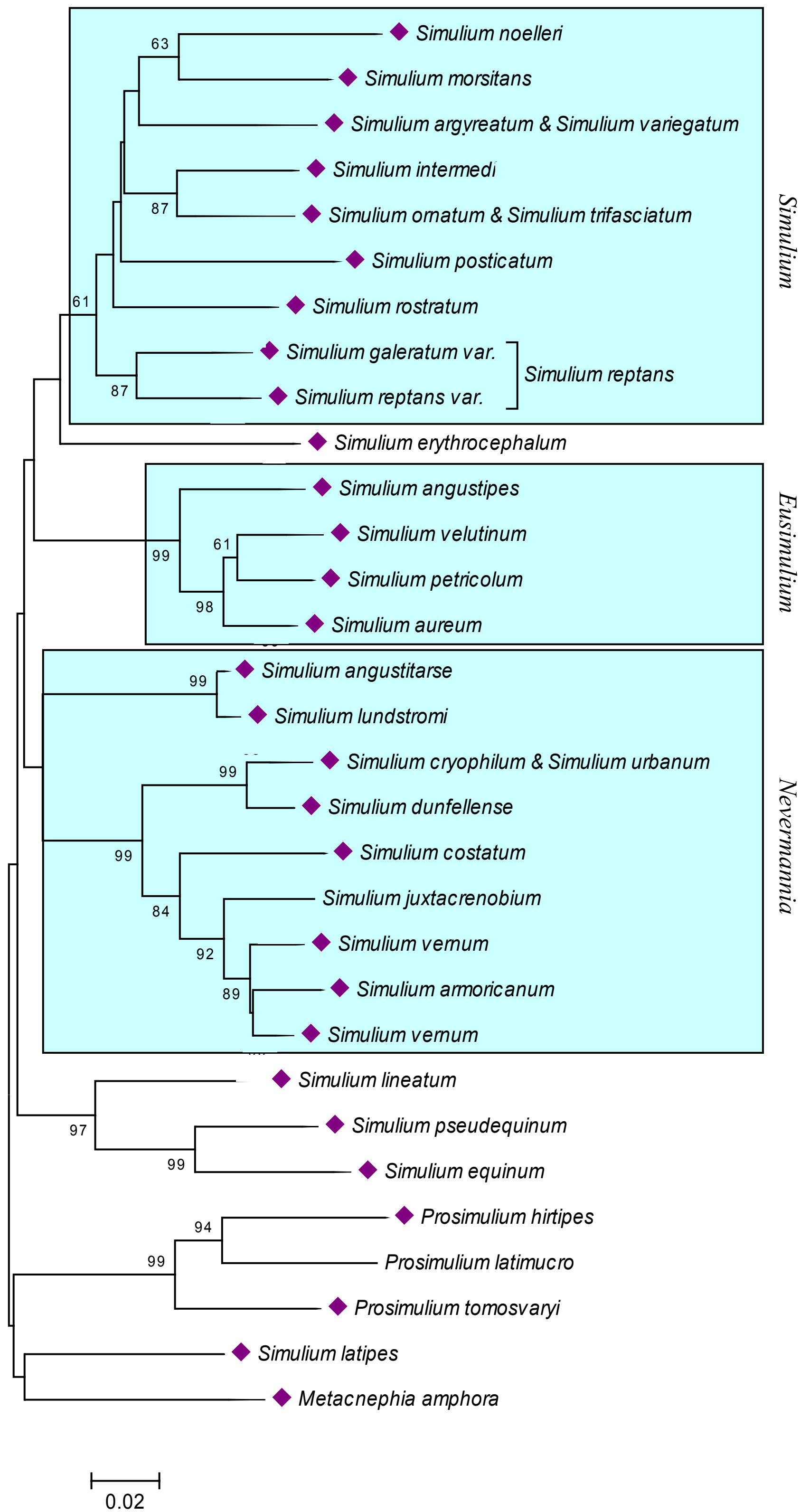


Figure 3.1 Neighbor-joining tree of 256 coxI sequences from 32 blackfly species, using K2P distances

Collapsed clades of multiple specimens of individual species are marked by a purple diamond. Bootstrap values greater than 50 shown. The three subgenera examined in more detail are highlighted in blue boxes.

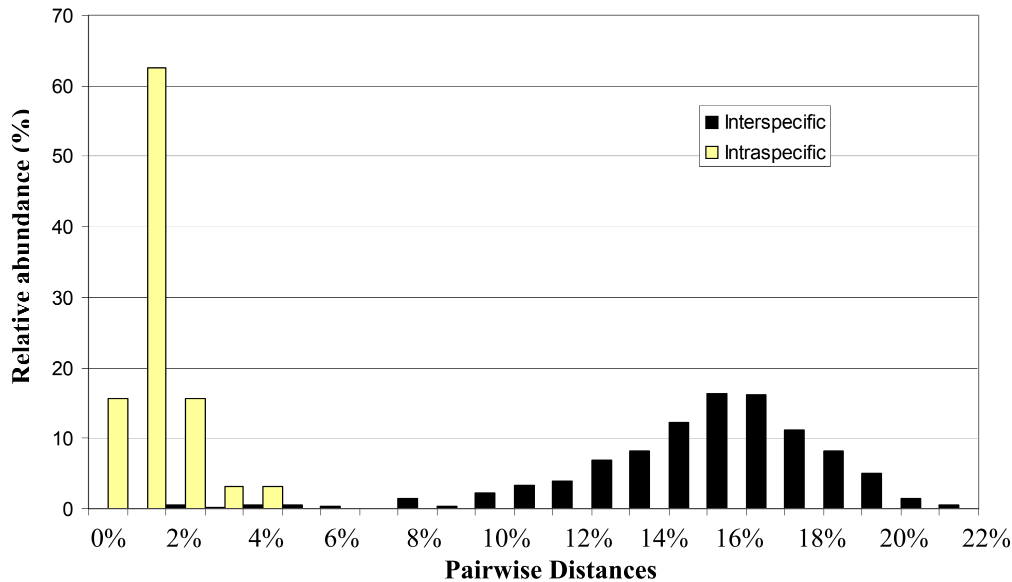


Figure 3.2 Overlap between intra- and interspecific genetic variability for congruent sequences

Table 3.2 Mean percentage base composition (with standard error), comparing *coxI* sequences of five simuliid subgenera

Group	Number of species	G	% base composition C	A	T
Wilhelmia	3	17.66 ± 0.03	19.78 ± 0.10	28.00 ± 0.05	34.57 ± 0.09
Prosimulium	3	17.39 ± 0.01	20.16 ± 0.03	27.14 ± 0.03	35.31 ± 0.06
Eusimulium	4	16.99 ± 0.07	17.96 ± 0.10	28.89 ± 0.06	36.16 ± 0.08
Nevermannia	9	16.93 ± 0.04	19.05 ± 0.19	27.36 ± 0.10	36.67 ± 0.12
Simulium	9	16.61 ± 0.04	18.20 ± 0.09	27.25 ± 0.06	37.94 ± 0.08
Simuliidae	32	17.03 ± 0.03	18.80 ± 0.07	27.71 ± 0.05	36.47 ± 0.09

Table 3.3 GC content of the 1st, 2nd and 3rd codon positions (with standard error) in five simuliid subgenera

Group	Number of species	GC% per codon position 1st	2nd	3rd
Wilhelmia	3	46.47 ± 0.06	43.31 ± 0.01	22.52 ± 0.30
Prosimulium	3	49.36 ± 0.02	43.25 ± 0.02	20.08 ± 0.14
Eusimulium	4	48.93 ± 0.09	43.32 ± 0.03	12.64 ± 0.21
Nevermannia	9	47.23 ± 0.16	43.38 ± 0.03	17.30 ± 0.50
Simulium	9	47.82 ± 0.13	43.34 ± 0.02	13.30 ± 0.27
Simuliidae	32	47.55 ± 0.07	43.33 ± 0.01	16.60 ± 0.29

3.3.2 Subgenera *Simulium*

We examined 10 of the 11 members of the subgenera *Simulium*, with one species, *S. tuberosum*, remaining unobtainable for the duration of the project. The majority of species were represented by multiple specimens (Figure 3.3). One species, *S. trifasciatum*, was represented by only one individual and was found to be a part of a *S. ornatum* subclade. *Simulium reptans* material was composed of two variants, and these produced monophyletic subclades with high bootstrap support and were therefore ascribed conspecific status.

All assemblages of conspecific individuals had bootstrap values of around 99 per cent with the exception of the *S. ornatum* clade, which had a bootstrap value of 50 per cent. K2P nucleotide diversity within species was limited, ranging from 0.1 per cent to 3.9 per cent. For most species, individuals from two or more populations were examined. There was no obvious correlation between higher intraspecies diversity and greater geographical variation. Divergence between species was high, with an average K2P distance of 10.87 per cent. K2P nucleotide diversity within species exhibited a broad range (1.5 per cent to 13.1 per cent), with the lower value being found for the comparison between *S. argyreatum* and *S. variegatum*, which produced an unresolved assemblage.

3.3.3 Subgenera *Eusimulium*

Members of the subgenus *Eusimulium* attack birds and are major vectors of the protozoan parasites *Leucocytozoon* (bird malaria) and *Trypanosoma* (bird trypanosomiasis) (Crosskey 1990). Four species of *Eusimulium* were collected from 20 different sites in Britain. The coxI region was sequenced from one verified adult of *S. aureum* and *S. petricolum* and two *S. velutinum* adults. All *S. angustipes* material was either larvae or pupae. In these stages, *S. angustipes* is indistinguishable from *S. velutinum* or *S. petricolum*, but was putatively ascribed species status based upon habitat. Immature specimens of *Simulium aureum* were ascribed to species-level based upon the morphological features reported by Bass (1998).

A neighbor-joining tree analysis produced four major clades with high bootstrap values (99 per cent). Each clade contained the identified adult specimen and a further clade contained all the putatively described *S. angustipes* specimens. One specimen, which was found in a small stream with *S. angustipes* and *S. velutinum*, was discovered to have an almost identical coxI barcode. Within distances ranging from 0.2–1.7 per cent, the latter resulted from what would appear to be two subclades in the *S. velutinum* assemblage. Divergence between species was relatively high, with an average K2P distance of 6.3 per cent and a range of 4.7–7.9 per cent. Comparative analysis of these sequences showed that the *S. aureum* group is genetically very similar to *S. velutinum*, with a between species average K2P distance of 4.7 per cent. *S. angustipes* material formed the most distant clade, with a between species average K2P distance of 7.2–7.9 per cent for all three species.

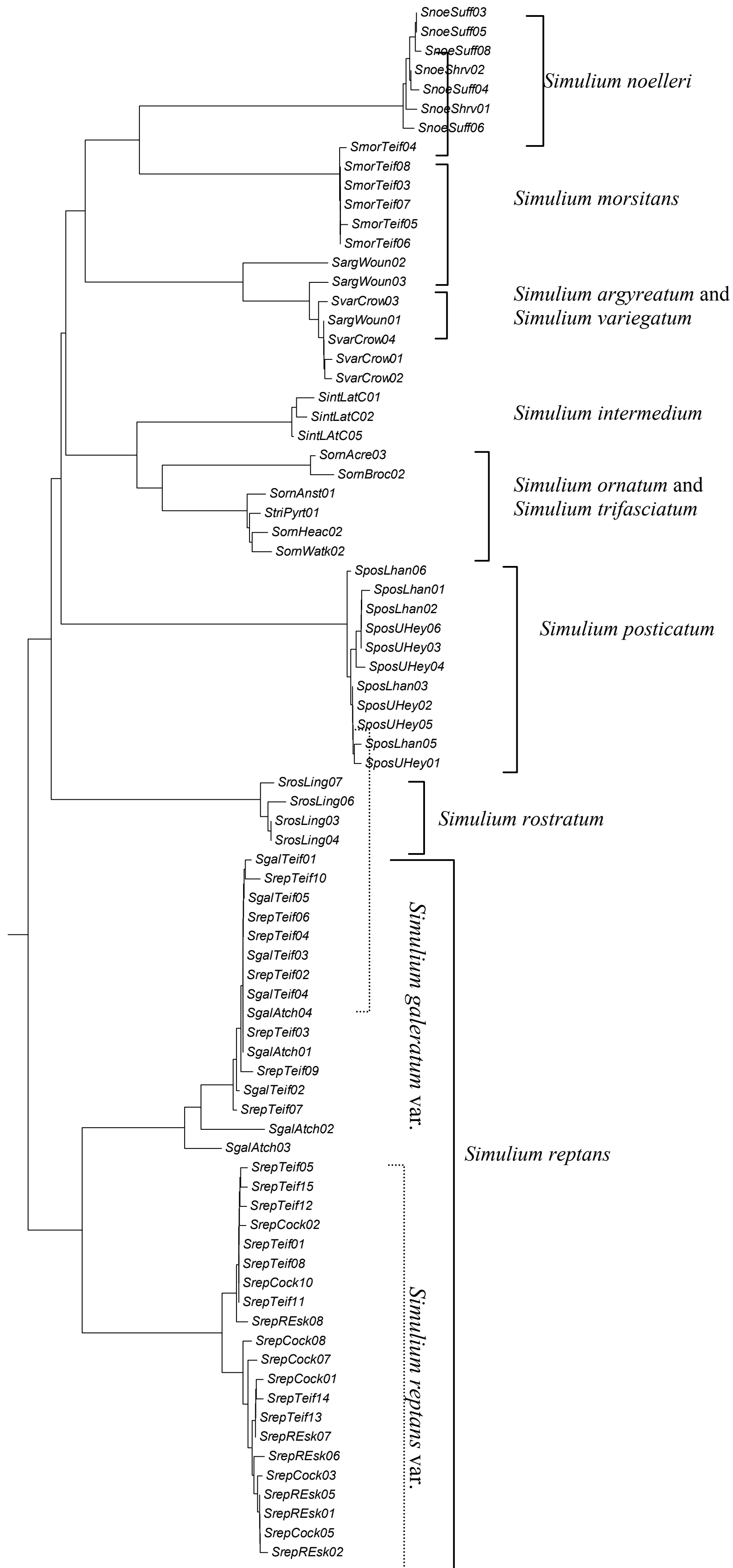


Figure 3.3 K2P distance neighbor-joining tree of 45 coxI sequences from 10 species belonging to the subgenus *Simulium*

Bootstrap values greater than 50 shown for 1000 replicates. Species codes are followed by population codes and then by an individual identifier.

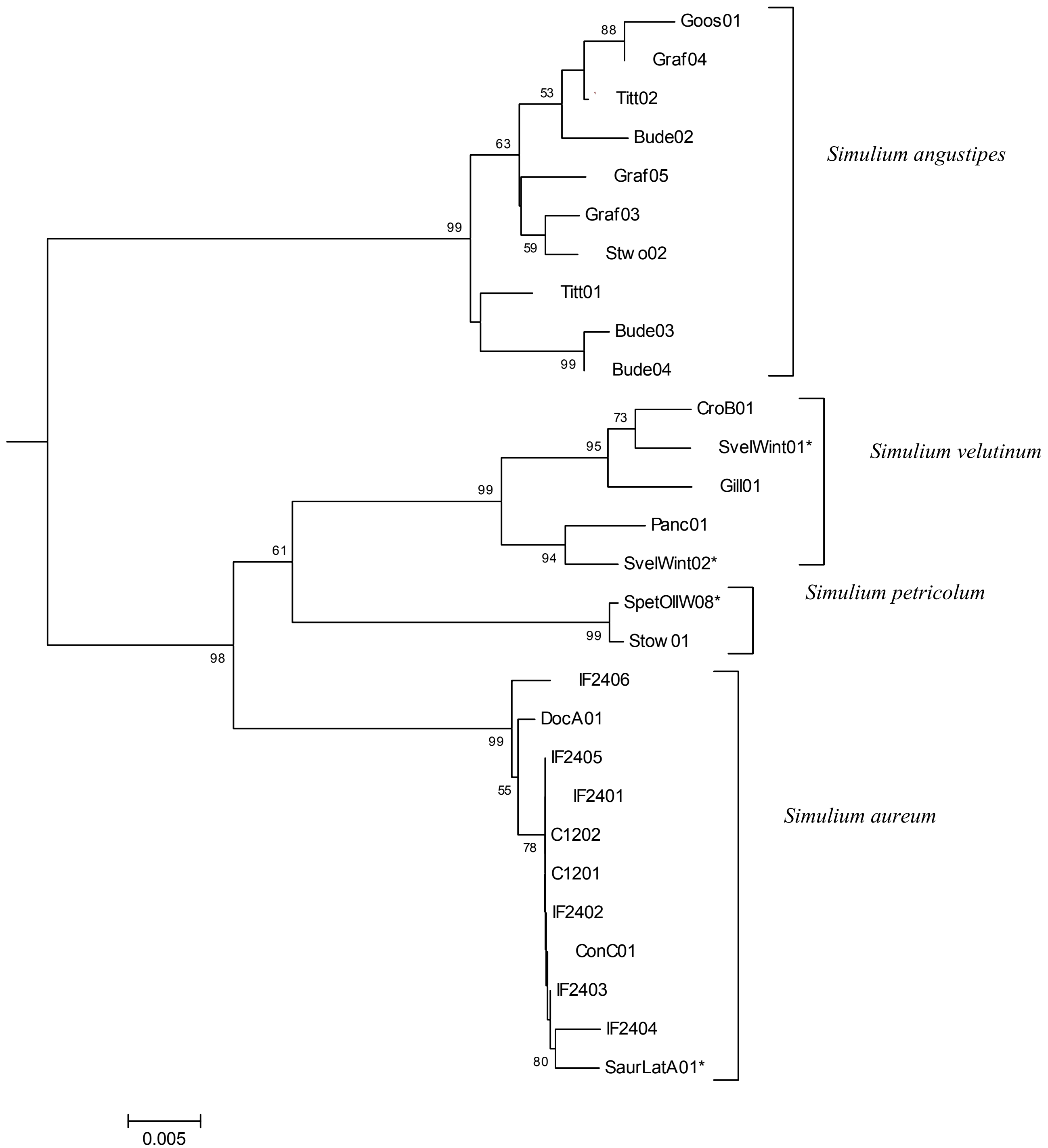


Figure 3.4 K2P distance neighbor-joining tree of 28 coxI sequences from four members of the *Simulium aureum* group belonging to the subgenera *Eusimulium*
Bootstrap values greater than 50 shown for 1000 replicates. Species codes are followed by population codes and then by an individual identifier for specimens identified as adults (marked with an asterisk). All other material came from immature stages and could not be confidently ascribed to a specific species.

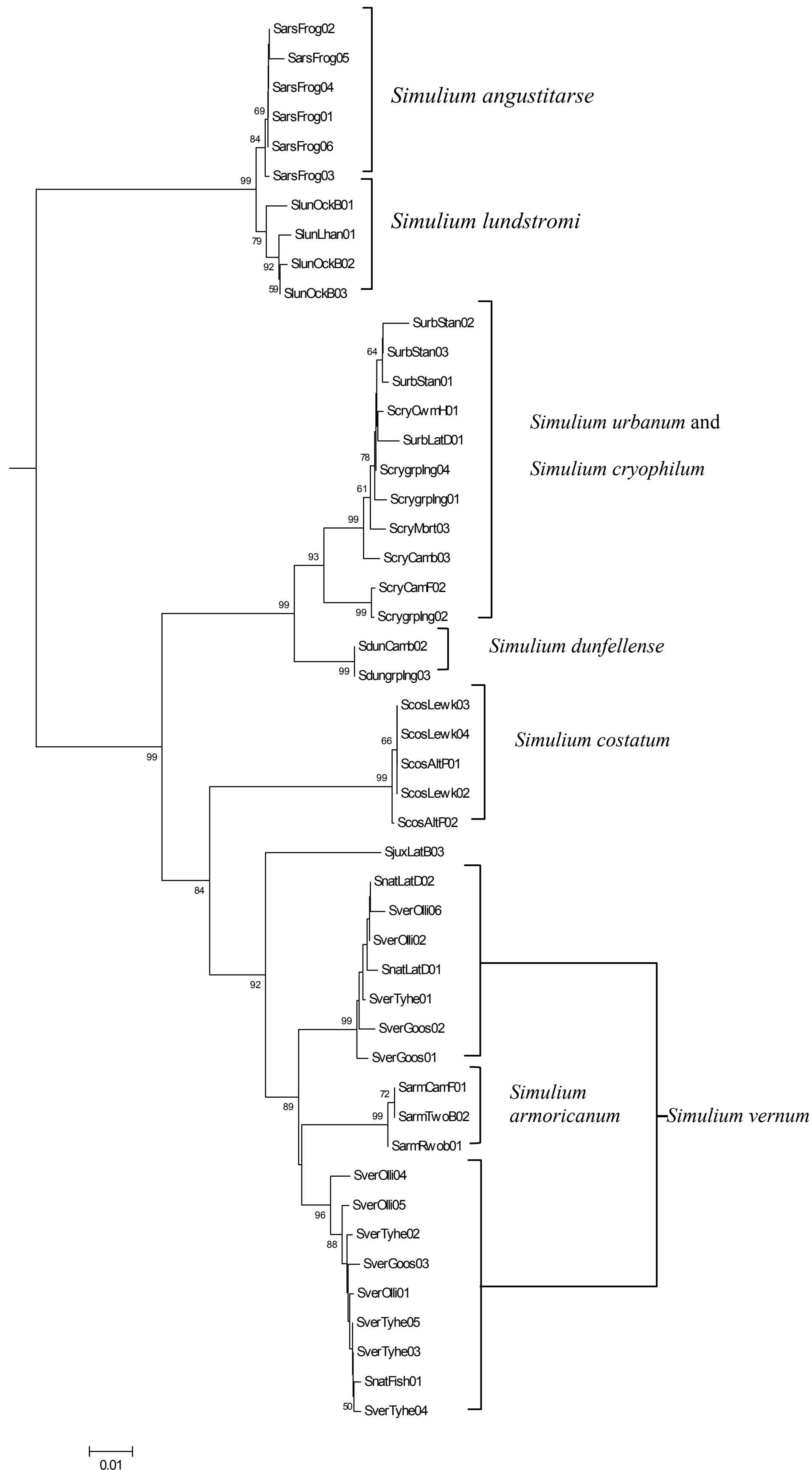


Figure 3.5 K2P distance neighbor-joining tree of 48 coxI sequences from eight species belonging to the subgenera *Nevermannia*

Bootstrap values greater than 50 shown for 1000 replicates. Species codes are followed by population codes and then by an individual identifier.

3.3.4 Subgenera *Nevermannia*

Members of the subgenus *Nevermannia* are predominantly found in small streams and are rarely found in large rivers. There are 48 species of *Nevermannia* in Europe (http://www.faunaeur.org; Crosskey and Howard 2004). We examined all nine species of *Nervermannia* found in Britain; all except *S. juxtacrenobium* were represented by multiple sequences (Figure 3.5). Four principle clades were evident, supported by bootstrap values ranging from 79–100 per cent. A *S. vernum* clade was supported by a bootstrap value of 79 per cent and was composed of two *S. vernum* subclades and a *S. armoricanum* subclade. Fourteen verified *S. vernum* specimens from four geographically-diverse populations were present in both *S. vernum* subclades. A *S. costatum* clade composed of sequences from two populations was supported by a bootstrap value of 100 per cent. A *S. cryophilum* clade was composed of sequences derived from material that was morphologically identified as *S. urbanum*, *S. dunfelense*, *S. cryophilum* and *S. armoricanum* but was not resolved as separate lineages with coxI. Three subclades of the *S. cryophilum* group assemblage were designated I–III. A fourth clade was composed of *S. lundstromi* and *S. angustitarse* sequences, which, although separated, did not form monophyletic lineages. K2P nucleotide diversity within species ranged from 0 per cent to 1.7 per cent, due to not all the species clustering as separate assemblages. This was most evident in the *S. cryophilum* and *S. vernum* assemblages. Divergence between species covered a wide range, with average interspecies K2P distances ranging from 1.00 per cent to 14.8 per cent.

3.4 Discussion

Numerous biologists have argued that the future of descriptive taxonomy will ultimately depend upon the consolidation of traditional morphotaxonomy with new technology. Many ideas have been discussed (Godfray 2002), with attention recently focusing on the contribution of molecular taxonomy using DNA barcoding. Despite the acceptance that DNA sequences can uncontrovertibly assist in investigating disparate life history stages, the association of male and female specimens and cryptic diversity within a species, many biologists have questioned the resolution of DNA barcoding and its ability to augment traditional taxonomy.

This study has strongly confirmed the ability of coxI barcodes to identify the majority of simuliid species in Britain. We sequenced (predominantly multiple) specimens of three genera: three *Prosimulium* species, one *Metacnephia* species and 18 *Simulium* species. Of the 32 species sequenced, 26 could subsequently be discriminated (81 per cent). All problematic taxa were found in the subgenera *Simulium* and *Nevermannia*. All the species were amplified with the standard barcoding primer pair and there was no need to optimise or redesign the primer sequence.

Despite the frequency with which mtDNA loci were detected in nuclear DNA (NUMTs), we found no evidence of coxI sequences originating from the nuclear genome. A review of the occurrence of NUMTs in plant and animal genomes found no evidence of NUMTs in the genome of the mosquito *Anopheles gambiae*, which is the closest dipteran genome to the Simuliidae (Richly and Leister 2004). However, three NUMTs have been reported in

the genome of *Drosophila melanogaster* and this highlights the need for vigilance in examining insect amplicons for potential pseudogene status (Bensasson *et al.* 2001).

Although DNA barcode analysis seeks only to delineate species boundaries, there is clearly some phylogenetic information in coxI sequence data. For instance, the four major subgenera were present as monophyletic clades with high bootstrap values. The members of the *Prosimulium* and *Metacnephia,* along with one species of the genus *Simulium* (*S. latipes*)*,* formed an assemblage basal to the *Simulium* clades. *S. latipes* is classified as a member of the subgenus *Hellichiella* and is the only species of this subgenus found in Britain. Only two sequences were available at the time of analysis; further phylogenetic studies will require not only more material from this species but also more material from other species in the *Hellichiella*, in order to resolve its relationship within the Simuliidae.

Nevertheless, it is not possible to recover the true phylogeny of Simuliidae by conducting K2P and neighbor-joining analysis on a 658bp fragment of mitochondrial DNA. Extensive phylogenetic studies have been conducted on simuliid species employing loci such as 28S rDNA, EF-1alpha, DDC, PEPCK, and 12S rDNA (Moulton 2000, Moulton 2003). Considerably more information from other gene regions (including nuclear genes) should be included in an analysis. Additional analytic methods should also be deployed, including maximum parsimony, maximum likelihood and minimum evolution (Moulton 2000, Moulton 2003)

Barcoding discriminated all the blackfly species from the subgenera *Boopthora*, *Eusimulium*, *Helichelia, Prosimulium* and *Wilhelmia*, as well as *Metacnephia amphora*, and would clearly be capable of unambiguously identifying all life stages: larvae, pupae and adults, and, potentially, egg masses. However, we were unable to resolve unambiguously all species from the subgenus *Nevermannia* and *Simulium*. Theoretical arguments for and against DNA barcodes often debate the overlap between intraspecific and interspecific genetic variabilities (Stoeckle 2003, Ward *et al.* 2005a, Will and Rubinoff 2004). In Diptera, extensive overlap (up to 15.5 per cent) has been shown and many (99 per cent) of the pairwise distances for congeneric sequences fall into the area of overlap (Meier *et al.* 2006).

We found a much less extensive overlap in this study. Although some overlap does exist, it is clearly only a small proportion of the total pairwise comparisons. This overlap can be associated with the closely related taxa that fail to resolve using the discrimination criteria employed in this study. One such species, *S. trifasciatum*, was represented by only one individual, which was found to be part of a *S. ornatum* subclade. As the distance between these two subclades is quite extensive, it may be possible that *S. trifasciatum* is represented by more than one individual and that the *S. ornatum* subclade that it belongs to contains other *S. trifasciatum* specimens.

S. ornatum and *S. trifasciatum* can only be separated at the adult and pupal stage by the presence of either round or pointed micro-tubercles on the thoracic cuticle. The *S. ornatum* specimens were represented by larvae and designated *S. ornatum* through association with habitat and rivers. *S. trifasciatum* is only found in small streams and is much rarer than *S. ornatum*. This means that it cannot be discounted that the specimens found in the *S. trifasciatum* subclade are from the same species. It is also possible,

however, that introgression occurred between *S. trifasciatum* and *S. ornatum* (*S. ornatum* was also found at the Pyrton site).

The sequences from the *S. ornatum* clade may therefore be all *S. ornatum* and are divided into two subclades of the same morphospecies. The two subclades are well supported by bootstrap values of 99 per cent, but extensive K2P distances between the two subclades (5.2 per cent) support the theory that these two subclades are separate species. It is, however, still unclear as to whether these are in fact *S. ornatum* and *S. trifasciatum* or whether they are both *S. ornatum*, which has been shown to comprise at least four different cytotypes. If these specimens are all true *S. ornatum* then the subclades may, in fact, associate with two of the four cytotypes. Due to the problems associated with chromosome analysis (redundancy of specimens, length of time and expertise required), it was not possible to analyse cytotyped material within the limited timeframe of the project. We anticipate in the future evaluating a larger dataset of *S. ornatum* specimens to investigate the correlation between cytotype and DNA barcode.

Two further species, *S. argyreatum* and *S. variegatum*, were unresolved within the *Simulium* subgenes. Controversy surrounds these two species. The principle method of defining them is pupal variation, in the form of the presence or absence of patagia on the thorax, along with adult characteristics (Davies 1966). Up until recently, larvae were discriminated based upon the contrast between apotome markings and the base colour of the head capsule. This was later revised by Bass, who concluded that it was an unreliable taxonomic character state for these two species (Bass 1998). In other species, the presence of patagia has been questioned as a reliable taxonomic character (R.J. Post personal communication). A much larger number of specimens, comprising all life history stages, need to be studied in order to establish the status of these two closely related species.

The most positive support for DNA barcoding came from the subgenus *Eusimulium*. Here, all the species of the *S. aureum* group are predominantly morphologically identical and can only truly be resolved using adult morphology and cytotaxonomy of the larvae. We have shown clear demarcations of these species using DNA barcodes and have also produced a preliminary population survey of sites in Britain. As a result, a second site for *S. petricolum*, which is a recently determined species and new to Britain, has been identified. This new site will provide more information about the ecology, morphology and species range of this species in Britain.

The subgenera *Nevermannia* is inherently problematic to resolve morphologically. Most of the specimens studied were larvae and head capsules were retained as voucher specimens for the majority of individuals. Re-examining some of this material caused some reassessment of species status. We believe that a number of unresolved species were a result of intraspecies morphological variation overlapping with diagnostic characters for other species, in particular *S. cryophilum* and *S. dunfellense*. Material initially regarded as one species was often revised as a result of molecular analysis and then morphological reassessed.

S. dunfellense remains to be resolved successfully and the two sites where *S. dunfellense* was identified will probably be revisited in order to provide more material for future analysis. Also problematic within the *Nevermannia* were the specimens of *S. urbanum*.

Three individual specimens of *S. urbanum* were collected from the type locality (Stanmore Common, North London) by the author. This material was identified by the author as *S. urbanum* and this identification verified by Roger Crosskey (Natural History Museum). *S. urbanum* was insufficiently resolved from the other specimens of *S. cryophilum*, and one individual from Latchmore brook in the New Forest appeared to be more closely related to *S. cryophilum* than to the material from Stanmore Common. It is possible that *S. urbanum* from Latchmore Brook may have been misidentified or that coxI is unable to resolve these closely related species. Certainly, *S. urbanum* is only found in a few locations in southern England. It may therefore be a morphological variant of *S. cryophilum* or there may be introgression between *S. cryophilum* and *S. urbanum*. This is a possibility, as *S. cryophilum* has also been reported at Stanmore Common (R. Crosskey personal communication). However, further work on additional material from a number of sites is required in order to verify this claim.

It was interesting to note that *S. vernum* formed two distinct monophyletic subclades, divided by the *S. armoricanum* assemblage. Based upon chromosome studies, *S. vernum* is known to be a species complex, composed of at least four cytospecies in Britain (Brockhouse 1985). The two *S. vernum* subclades could very well correspond to cytospecies groupings. For the *Nevermannia* specimens, cytological verification of material was not carried out. This was because of problems associated with preservation and the limited application to life stages (chromosomal preparations can only be conducted on late instar larvae), as well as the cytotaxonomic expertise required to validate specimens. Nevertheless, the coxI results are not congruent with cytological studies.

Specimens from all three populations – two of which, ‘Olli’ and ‘Goos’, were from southern England – were divided into the two subclades. Cytological studies found that only one cytotype, ‘Knebworth’, predominated in southern England, with the exception of a few locations that contained different cytotypes (‘Dorset IIs-1’, ‘Lymington’ and ‘Dorset IIs-2+3’). The specimens that were morphologically described as *S. armoricanum* were collected from Dartmoor and may represent a new form of *S. vernum*. Morphological, cytological and genetic studies of these populations will be required to validate the identity of these specimens.

The strength of a molecular taxonomic system depends on the resolving power of the identification. Despite the limitations of DNA barcodes illustrated herein, it is also clear that such sequences can provide strong support to resolve the unsubstantiated taxanomic status of certain species. In our analysis, *Simulium reptans* (Linnaeus 1758) formed two monophyletic clades, which were founded upon the presence of two variants of this species in our dataset. Since 1920, *Simulium reptans* in Britain has been reported as exhibiting two different larval morphotypes: a typical *S. reptans* and an atypical *S. reptans* var. *galeratum*, which differ in the markings of the larval head capsule (Edwards 1920). The sequences form two principle clades: one composed of all typical *S. reptans* larvae (plus three undefined pupae from the River Teifi) and a second clade composed of all var. *galeratum* larvae (plus seven unknown pupae from the River Teifi).

This shallow within-clade and deep between-clade divergence suggests the existence of significant genetic differentiation between two distinct populations, which correspond precisely with the two larval morphotypes. This pattern very strongly supports the

existence of two distinct species, corresponding to *S. reptans* and *S. galeratum* (both *sensu* Edwards). The average intraspecific K2P distance for *S. galeratum* and *S. reptans* was very low (0.67 per cent and 0.78 per cent respectively), with the K2P distances between species (mean 7.06 per cent) being nine to 10-fold greater. The intra- and interspecific genetic variability between *S. reptans* and *S. galeratum* were similar to the normal 10-fold difference established by Hebert *et al.* (2004). This observation has been examined in more detail elsewhere (Day *et al.* 2006).

The various unresolved questions about specimen identification that are briefly presented in this report (for the species *S. cryophilum*, *S. urbanum*, *S. variegatum*, *S. argyreatum*, *S. trifasciatum* and *S. ornatum*) indicate the need for extensive morphological and molecular studies on all life stages of these species. They also highlight the importance of retaining voucher specimens wherever possible, or at least retaining an e-voucher photograph. While we retained a single voucher specimen from the same population for the majority of species discussed herein, along with the head capsules of larvae and in some instances the gill filaments of the pupae, other samples were destroyed in the extraction process. The size and age of some of the specimens necessitated the destruction of the entire specimen to maximise the amount of recovered DNA. Retaining large numbers of specimens as vouchers will require significant curation and bioinformatic facilities, which will be implemented for future studies of blackfly DNA barcodes.

This study presents a representative DNA barcoding study of the blackfly fauna of Britain and highlights the areas necessary for future study, in order to establish the extent to which coxI can resolve all blackfly species. Certainly, at present, DNA barcoding can justifiably be employed to resolve around 80 per cent of the British blackfly fauna. Examination of the subgenus *Nevermannia* and *Simulium* has shown that more taxonomic, as well as molecular analysis is required to resolve these unresolved species. A paucity of taxonomic information is available on this group of species and cytological data is only available for larva. It is not therefore possible to correlate cytotype with the pupal or adult stages. Nevertheless, DNA barcodes can be made available for all life stages. With broader collaboration and expert morphological analysis, DNA barcodes will greatly benefit biodiversity surveys of blackfly taxa throughout the world.

4 Conclusion

At present it is not possible to advocate a DNA barcoding system that can accurately identify all members within a blackfly community *en masse*. It is, however, possible to identify many species in certain life history stages that cannot be resolved based on morphology. The strength of a molecular taxonomic system depends on the resolving power of the identification. Despite the limitations of DNA barcodes illustrated in this report, it is clear that such sequences can provide strong support to resolve the unsubstantiated taxonomic status of certain species. Furthermore, such an approach does not require extensive taxonomist training and is a useful addition to existing tools available for species identification.

DNA barcoding is one of many molecular methods that may be useful to improve or complement existing methods used within the Environment Agency and the choice would be dependent on the subsequent application. Other methods such as DNA arrays (micro and macro), Quantitative Real-time PCR and DNA profiling techniques (see Environment Agency (2003) for an overview of methods) have the potential to fulfil the Environment Agency's commitments to biodiversity and monitoring in both aquatic and terrestrial systems. The pace at which these methods are being refined is great as is their throughput. The only way we can test the feasibility and resolution of these methods is to test them in small proof-of-concept studies.

References

AVISE, J.C., 1994. Molecular markers, natural history and evolution. New York: Chapman & Hall.

BAKER, S.C., DALEBOUT, M.L., LAVERY, S. AND ROSS, H.A., 2003. www.DNA-surveillance: applied molecular taxonomy for species conservation and discovery. *Trends in Ecology & Evolution*, 18 (6): 271–272.

BÁLDI, A. 1999. Biodiversity in Hungary: advantages and limitations of taxonomically complete faunal inventories. *Natural Area Journal* 19, 73-78.

BÁLDI, A. 2003. Using higher taxa as surrogates of species richness: a study based on 3700 Coleoptera, Diptera, and Acari species in Central-Hungarian reserves. *Basic and Applied Ecology* 4, 589-593.

BASS, J., 1998. *Last-instar larvae and pupae of the Simuliidae of Britain and Ireland*. Freshwater Biological Association no. 55

BASS, J.A.B., 1985. New keys to the immatures of the British simuliids: plans, pitfalls and progress so far with the *Eusimuliums*. *Newsletter of the British Simulium Group*, 11, 4–6.

BASS, J.A.B., CROSSKEY, R.W. AND WERNER, D., 1995. On the European blackfly *Simulium lundstromi* and inclusion within this species of *S. latigonium* as a new synonym. *Bulletin of the British Simulium Group*, 5, 7–19.

BENSASSON, D., ZHANG, D.-X., HARTL, D.L. AND HEWITT, G.M., 2001. Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution*, 16 (6), 314-321.

BLAXTER, M. AND FLOYD, R., 2003. Molecular taxonomics for biodiversity surveys: already a reality. *Trends in Ecology & Evolution*, 18 (6), 268-269.

BROCKHOUSE, C., 1985. Sibling species and sex chromosomes in *Eusimulium vernum* (Diptera: Simuliidae). *Canadian Journal of Zoology*, 63, 2145–2161.

CROSSKEY, R.W., 1990. *The Natural History of Blackflies*. Chichester: John Wiley & Sons.

CROSSKEY, R.W., 1991. A new checklist of the blackflies of Britain and Ireland, with geographical and type information (Diptera: Simuliidae). *Entomologist's Gazette*, 42, 206–217.

CROSSKEY, R.W. AND HOWARD, T.M., 2004. A revised taxonomic and geographical inventory of world blackflies (Diptera: Simuliidae) Department of Entomology, The Natural History Museum, London. Available from: http://www.nhm.ac.uk/entomology/projects/blackflies/index.html [Accessed 21 November 2006].

DAVIES, L., 1966. The taxonomy of British black-flies (Diptera: Simuliidae). *Transactions of the Royal Entomological Society of London* 118 (14), 413–511.

DAVIES, L., 1968. A key to the British species of Simuliidae (Diptera) in the larval, pupal and adult stages. *Scient. Publs Freshwater Biol. Ass.*, 24, 1–126.

DAY, J.C., GOODALL, T.I. AND POST, R.J., 2006. *Confirmation of the species status of Simulium galeratum Edwards, 1920 (Diptera: Simuliidae) in Britain using DNA barcoding*. In preparation.

DE BIASI, A.M., BIANCHI, C.N AND Morri, C. 2003. Analysis of macrobenthic communities at different taxonomic levels: an example from an estuarine environment in the Ligurian Sea (NW Mediterranean). *Estuarine, Coastal and Shelf Science* 58, 99–106.

DOBZHANSKY, T., 1937. *Genetics and the origins of the species*. New York: Columbia University Press.

EBACH, M.C., and HOLDREGE, C. 2005. DNA barcoding is no substitute for taxonomy. *Nature* 434 697.

EDWARDS, F.W., 1920. On the British species of *Simulium*. II. The early stages; with corrections and additions to Part I. *Bulletin of Entomological Research*, 11, 211–246.

ELLIS, D. 1985. Taxonomic sufficiency in pollution assessment, *Marine Pollution Bulletin* 16, 459.

ENVIRONMENT AGENCY, 2003. Environmental Genomics – An Introduction. Environment Agency: Bristol.
http://publications.environment-agency.gov.uk/pdf/SGENOMICS-e-p.pdf

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.

GIANGRANDE, A. 2003. Biodiversity, conservation, and the ‘taxonomic impediment.’ *Aquatic Conservation: Marine and Freshwater Ecosystems* 13 (5), 451–459.

GODFRAY, H.C.J., 2002. Challenges for taxonomy. *Nature*, 417, 17–19.

HAJIBABAEI, M., SINGER, G.A.C., HEBERT, P.D.N., AND HICKEY, D.A. (2007) DNA barcoding: how it comlpements taxonomy, molecular phylogenetics and population genetics. *TRENDS in genetics* 23 (4) 167-172.

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 B: Biological Sciences*, 270, S96–S99.

HEBERT, P.D.N., PENTON, E.H., BURNS, J.M., JANZEN, D.H. AND HALLWACHS, W., 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (41), 14812–14817.

HEBERT, P.D.N., STOECKLE, M.Y., ZEMLAK, T.S. AND FRANCIS, C.M., 2004b. Identification of birds through DNA barcodes. *PLoS Biology*, 2 (10), e312.

HEINO, J. AND SOININEN, J. 2007. Are higher taxa adequate surrogates for species-level assemblage patterns and species richness in stream organisms? *Biological Conservation* 137, 78-89.

ISH-HOROWICZ, D., 1982. Rate of turnover of structural variants in the rDNA gene family of *Drosophila melanogaster*. *Nature*, 295: 564–568.

KIMURA, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, 16 (2), 111–120.

KUMAR, S., TAMURA, K. AND NEI, M., 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics*, 5, 150–163.

LIPSCOMB, D., PLATNICK, N. AND WHEELER, Q. 2003. The intellectual content of taxonomy: a comment on DNA taxonomy. *Trends in Ecology and Evolution* 18, 65-66.

MEIER, R., SHIYANG, K., VAIDYA, G. AND NG, P., 2006. DNA barcoding and taxonomy in *Diptera*: a tale of high intraspecific variability and low identification success. *Systematic Biology*, 55, 715–728.

MORITZ, C. AND CICERO, C. 2004. DNA barcoding: promise and pitfalls. *PLoS Biology* 2 (10), e354.

MOULTON, J.K., 2000. Molecular sequence data resolves basal divergences within Simuliidae (Diptera). *Systematic Entomology*, 25 (1), 95–113.

MOULTON, J.K., 2003. Can the current molecular arsenal adequately track rapid divergence events within Simuliidae (Diptera)? *Molecular Phylogenetics and Evolution*, 27 (1), 45–57.

PONS, J., 2006. DNA-based identification of preys from non-destructive, total DNA extractions of predators using arthropod universal primers. *Molecular Ecology Notes*, 6 (3), 623–626.

POST, R.J., 1980. Cytotaxonomy of the *Simulium ornatum* species-group in Britain. *Newsletter of the British Simulium Group*, 3, 3–5.

POST, R.J., MUSTAPHA, M. AND DAY, J.C., 2007. *The subgenus Eusimulium (Diptera: Simuliidae) in Britain. Aquatic Insects*. In preparation.

PRAMUAL, P., KUVANGKADILOK, C., BAIMAI, V. AND WALTON, C., 2005. Phylogeography of the black fly *Simulium tani* (Diptera: Simuliidae) from Thailand as inferred from mtDNA sequences. *Molecular Ecology*, 14 (13), 3989–4001.

PROUDLOVE, G. AND WOOD, P.J., 2003. The blind leading the blind: cryptic subterranean species and DNA taxonomy. *Trends in Ecology & Evolution*, 18 (6), 272–273.

RICHLY, E. AND LEISTER, D., 2004. NUMTs in sequenced eukaryotic genomes. *Molecular Biology & Evolution*, 21 (6), 1081–1084.

SAITOU, N. AND NEI, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology & Evolution*, 4, 406–425.

SEBERG, O., HUMPHRIES, C.J., KNAPP, S., STEVENSON, D.W., PETERSEN, G., SCHARFF, N., ANDERSEN, N.M. 2003. Shortcuts in systematics? A comment on DNA-based taxonomy. *Trends in Ecology and Evolution* 18, 63-65.

SMITH, M., FISHER, B.L. AND HEBERT, P.D., 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Philosophical Transcactions of the Royal Society London B: Biological Sciences*, 360 (1462), 1825-1834.

STOECKLE, M., 2003. Taxonomy, DNA, and the bar code of life. *BioScience*, 53, 796–797.

TAUTZ, D., ARCTANDER, P., MINELLI, A., THOMAS, R.H. AND VOGLER, A.P., 2002. DNA points the way ahead in taxonomy. *Nature*, 418 (6897), 479.

TERLIZZI, A., BEVILACQUA, S., FRASCHETTI, S. AND BOERO, F. 2003. Taxonomic sufficiency and the increasing insufficiency of taxonomic expertise. *Marine Pollution Bulletin* 46, 544-560.

WARD, R.D., ZEMLAK, T.S., INNES, B.H., LAST, P.R. AND HEBERT, P.D., 2005a. DNA barcoding Australia's fish species. *Philosophical Transcactions of the Royal Society London B: Biological Sciences*, 360 (1462), 1847–1857.

WARD, R.D., ZEMLAK, T.S., INNES, B.H., LAST, P.R. AND HEBERT, P.D., 2005b. DNA barcoding Australia's fish species. *Philosophical Transcactions of the Royal Society London B: Biological Sciences*, 360 (1462), 1847–1857.

WILL, K.W. AND RUBINOFF, D., 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics*, 20 (1), 47–55.

ZHANG, D.-X. AND HEWITT, G.M., 1996. Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology & Evolution*, 11, 247–251.

Glossary

Amplicon: A piece of DNA that has been synthesised using an amplification technique, such as the polymerase chain reaction (see below).

Amplification: an increase in the number of copies of a specific DNA fragment.

Bootstrap: A statistical procedure for achieving a better estimate of the parametric variance of a distribution. It involves sampling with replacement from the original sample to produce a pseudoreplicate with the same dimensions as the original.

Clade: See monophyly.

Conspecific: A term used to describe two or more individual organisms, populations or taxa that are thought to belong to the same species.

Cryptic species complex: In biology, a group of species that satisfy the scientific definition of species – they are reproductively isolated from each other – but are not morphologically distinguishable. The individual species within the complex can only be separated using non-morphological data, such as provided by DNA sequence analysis and chromosome banding, or thorough life history studies.

Cytochrome c oxidase: A large transmembrane protein complex that is the terminal electron acceptor in the electron transport chain used by mitochondria and bacteria.

Cytology: The study of cells and their functions.

DNA (Deoxyribonucleic Acid): the molecule that encodes genetic information. DNA is a double-stranded helix held together by bonds between pairs of nucleotides.

DNA barcoding: A taxonomic method that uses a short genetic marker in an organism's DNA to quickly and easily identify it as belonging to a particular species. The specific locus selected for barcoding is a region of the gene coding for the cytochrome oxidase I subunit.

Gene: a length of DNA which codes for a particular protein, or in certain cases a functional or structural RNA molecule.

Genome: all the genetic material in the chromosomes of a particular organism; its size is generally given as its total number of base pairs.

Kimura 2-parameter: This is a measure of the distance (difference) between two genetic sequences. It was established by Kimura (1980) and corrects for differences in the rates of transition and transversion.

Locus: A fixed position on a chromosome, such as a specific gene.

Macroarray: A low-density array of DNA molecules used for parallel hybridisation analysis (see microarray).

Microarray: a component of a device for screening genomic or cDNA. The array is a small glass slide or other solid surface on which thousands of immobilized oligodeoxynucleotide probes have been synthesized or robotically deposited in a

predetermined array, so that automated recording of fluorescence from each of the spots may score successful hybridizations. An array may be designed for the detection of all known genes of a species or selected specific sequences.

Mitochondria: A membrane-enclosed organelle found in most eukaryotic cells. They are described as ‘cellular power plants’, because they convert food molecules into energy in the form of ATP (adenosine triphosphate) via the process of oxidative phosphorylation. A eukaryotic cell contains about 2,000 mitochondria. Mitochondria contain DNA that is independent of the DNA located in the cell nucleus.

Molecular taxonomy: The use of proteins and genes to determine evolutionary relationships. The percentage of DNA that is similar in species is taken into consideration.

Monophyly (monophyletic group): A group composed of a most recent common ancestor and all its descendants, also known as a clade.

Morphology: The study of an organism’s shape and form.

Neighbor-joining: A taxonomic method developed by Saitou and Nei (1987) that does not require that all lineages have diverged by equal amounts. The method is especially suited for datasets that comprise lineages with widely varying rates of evolution. It can be used in combination with methods that allow correction for superimposed substitutions.

Polymerase chain reaction (PCR): A molecular biological technique for enzymatically replicating a target region of DNA, thereby allowing a small amount of DNA to be amplified exponentially. PCR, on the whole, is an in vitro technique and is not limited by the form of DNA used as a target. It can be extensively modified to conduct a wide variety of genetic manipulations. The region of interest is targeted using short homologous pieces of DNA, known as primers, that anneal to the region of interest. An enzyme, DNA polymerase, then reproduces the region of DNA between the primers.

Primer: Short pre-existing polynucleotide chain to which new deoxyribonucleotides can be added by DNA polymerase.

Real-time PCR (or Quantitative PCR): is an enormously powerful tool for the specific logarithmic amplification of known sequences of DNA. If the amplification is monitored continuously or in real time, a linear relationship can be established between the log of the starting concentration and the point at which product reaches a critical threshold (within the logarithmic section of amplification). Using appropriate standards, this allows a calibration curve to be generated over 6 orders of magnitude of template. Many different chemistries and platforms have been developed to allow real time monitoring of the PCR reaction.

Simuliidae: The family of dipteran insects commonly known as blackflies (also known as buffalo gnats or turkey gnats). Like mosquitoes, female blackflies gain nourishment for their egg development by sucking the blood of other animals. They are usually small, black or grey, with short legs and antennae. They are a common nuisance for humans and spread several diseases, including river blindness, in Africa and America.

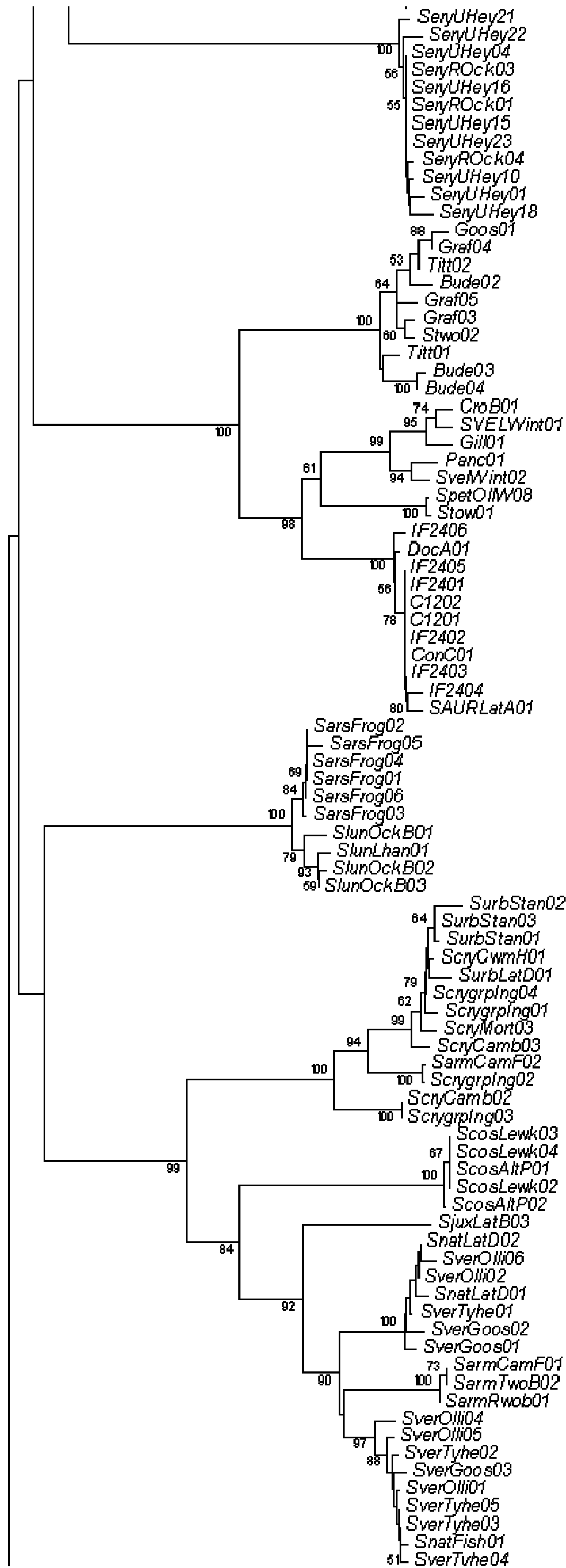
Species: Groups of populations (which are groups of individuals living together that are separated from other such groups) which can potentially interbreed or are actually interbreeding, that can successfully produce viable, fertile offspring (without the help of human technology).

Taq polymerase: A heat-stable DNA polymerase that is normally used in the polymerase chain reaction. It was isolated from *Thermus aquaticus*.

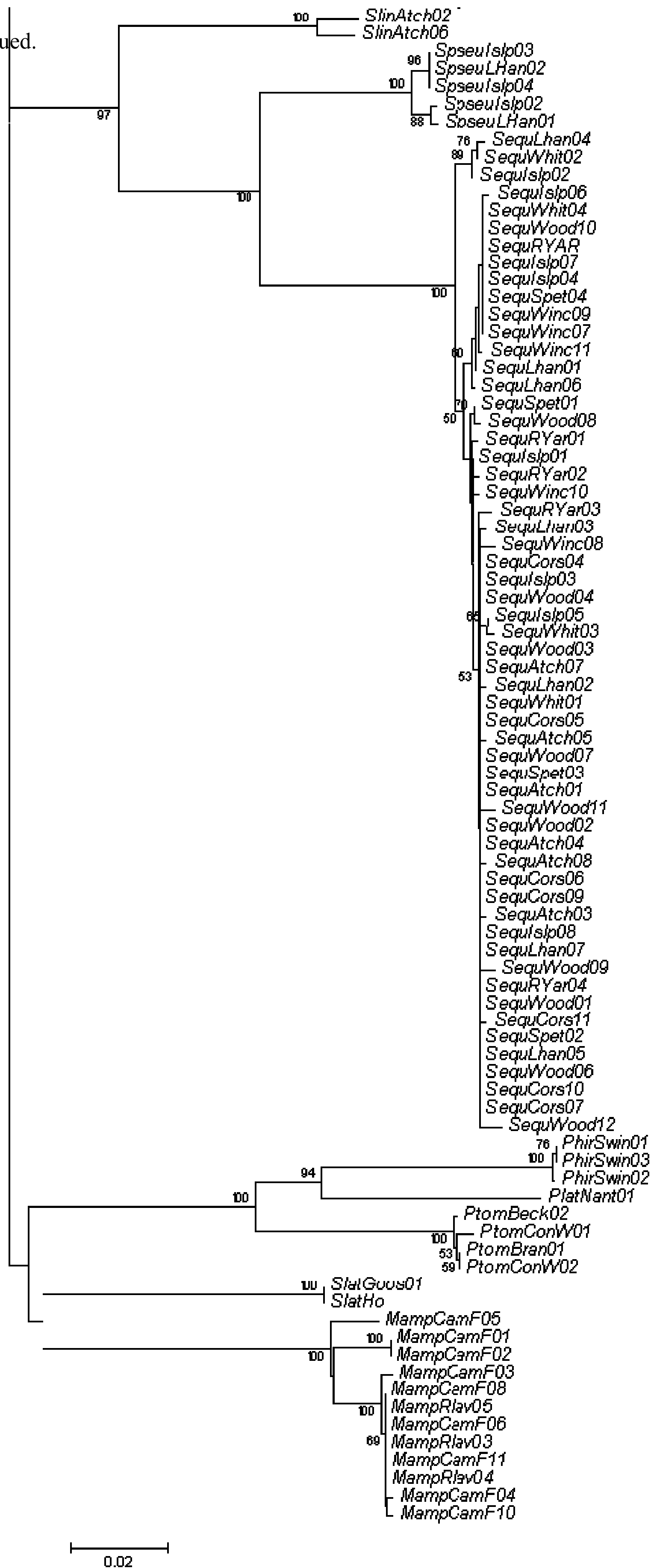
Taxonomy: The science of naming, describing and classifying organisms. Taxonomy forms part of the field of systematics, which establishes the evolutionary relationships between organisms.

Appendix

Appendix I continued.



Appendix I continued.



We are The Environment Agency. It's our job to look after your environment and make it **a better place** – for you, and for future generations.

Your environment is the air you breathe, the water you drink and the ground you walk on. Working with business, Government and society as a whole, we are making your environment cleaner and healthier.

The Environment Agency. Out there, making your environment a better place.

Published by:

Environment Agency
Rio House
Waterside Drive, Aztec West
Almondsbury, Bristol BS32 4UD
Tel: 0870 8506506
Email: enquiries@environment-agency.gov.uk
www.environment-agency.gov.uk

© Environment Agency

All rights reserved. This document may be reproduced with prior permission of the Environment Agency.