Introgressive hybridisation and incipient ecological speciation amongst saltmarsh *Aphrodes* leafhoppers

By

Joanna Katherine Bluemel

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Organisms and the Environment research group,
School of Biosciences, Cardiff University

Supervisor: Prof. W. O. C. Symondson
Co-supervisor: Dr. M. Virant-Doberlet
Additional external supervisor: Dr. A. Stewart

September 2011

DECLARATION

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed …… (candidate) Date ……02/03/2012……

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of ……PhD……(insert MCh, MD, MPhil, PhD etc, as appropriate)

Signed …… (candidate) Date ……02/03/2012……

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by explicit references.

Signed …… (candidate) Date ……02/03/2012……

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed …… (candidate) Date ……02/03/2012……

STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access previously approved by the Graduate Development Committee.**

Signed …… (candidate) Date ……02/03/2012……

Introgressive hybridisation and incipient ecological speciation amongst saltmarsh *Aphrodes* leafhoppers

By

Joanna Katherine Bluemel*

Supervisors: Prof. W.O.C. Symondson
Cardiff School of Biosciences
University of Cardiff
Biomedical Sciences Building
Museum Avenue
Cardiff, CF10 3AX
Wales, UK
E-mail: Symondson@cardiff.ac.uk

Dr. M. Virant-Doberlet
Department of Entomology
National Institute of Biology
Vecna pot 111
SI-1000 Ljubljana
Slovenia
E-mail: meta.virant@nib.si

* Present address:
Organisms and the Environment research group
Cardiff School of Biosciences
University of Cardiff
Biomedical Sciences Building
Museum Avenue
Cardiff, CF10 3AX
Wales, UK
E-mail: BluemelJK1@cardiff.ac.uk

SUMMARY

Ecological and host adapted races provide evidence that evolutionary divergence and sympatric speciation can occur through divergent natural selection in the face of continued gene flow. Likewise, hybridisation and introgression (interspecific gene flow) are commonly identified in natural populations, between what are described as distinct taxa. These processes have implications for how we define species and the processes necessary for the persistence and initiation of species and speciation, above and below the species level. The main focus of the present study was elucidation of the nature and extent of differentiation, and processes involved in shaping diversity within and between, species of the *Aphrodes* leafhopper genus, Curtis 1833, particularly from UK saltmarshes. A multidisciplinary approach was taken, combining the use of morphological, behavioural (vibrational mating signals), mitochondrial DNA (cytochrome oxidase subunit I gene sequencing) and multiple genome-wide nuclear marker (amplified fragment length polymorphism) analyses to test hypotheses relating to taxonomy, ecological speciation and hybridisation among *Aphrodes* leafhoppers. Of primary interest were: 1) identification of *Aphrodes* inhabiting saltmarshes, and first confirmation that two species (*A. makarovi* and *A. aestuarina*) exist there; 2) comparison of divergent ecological lineages of inland and estuarine *A. makarovi*, showing possible incipient speciation and evidence of convergent morphological evolution of estuarine *A. makarovi* and *A. aestuarina*; 3) exploration of the evolutionary significance of an introgressed hybrid population of *A. aestuarina*, found only in the Medway estuary, showing complete mitochondrial capture and some nuclear introgression.

To Wiz, Plug, Buffalo, Boss, Mr Bump and Indy-boy

Thank you for all your love, support and inspirational adventures

ACKNOWLEDGEMENTS

Firstly, thank you to my supervisors for the support and advice given throughout the project. Bill, thank you for making this project feasible and for giving me this highly rewarding opportunity. Meta, thank you for bringing this exciting but challenging project to Cardiff, which seems like a lifetime ago. I am very grateful for all the hours you invested in our manuscript, for the use of your *Aphrodes* specimens, vibrational mating signal data and bioacoustics training. Thank you for all the delicious Slovenian chocolate! I really appreciate the fast turn-around time on the thesis chapter edits that you both made, the frequent meetings we had and your useful comments throughout. You both gave me enthusiasm to persist even during the lowest times due to your passion for the subject, for which I am very grateful.

I would like to pay a very special thanks to Andrew King for the invaluable training (and patience) in the lab, scientific expertise in the subject areas, and your inability to say 'no' to someone in need. Your comments on chapters and advice with the data analyses were an enormous help. I would also like to acknowledge the mitochondrial DNA sequencing you carried out which helped to fill in some of the gaps in my data set and alleviated the technical problems I was experiencing with the AFLP analyses. I thoroughly enjoyed the jokes and laughs we shared in the lab and the fact that you made many difficult tasks seem that little more manageable. You are an exceptional scientist who I have always looked up to and aspire to emulate one day. Thank you for everything, I don't think I could have done it without you.

To Mike Wilson (National Museum of Wales), thank you very much for all the encouragement, morphological and taxonomic expertise, loan of *Aphrodes* museum specimens, use of photographic imaging equipment, and your continued interest in the *Aphrodes* story. Thank you for the many cups of tea and the amusing chats that kept me sane during tough times over the years. To both Mike and Roland, thank you for the fun times we shared at the conference in Germany and the confidence that I gained from it. I would also like to thank Alan Stewart (University of Sussex) for help in the field and for advice on *Aphrodes* during the early stages of this project. I am grateful to Tony Irwin, Norfolk Museum Service, for the kind loan of the *Aphrodes* archive specimens and confidence in my abilities to carry out molecular work on the *A. aestuarina* syntype specimens.

I would like to thank Isa-Rita Russo (Cardiff University) for the constructive comments on both earlier drafts of the manuscript and thesis chapters, advice and training on genetic data analyses and most of all your highly valued friendship. You have been a great companion during my final years and your support is greatly appreciated. Thank you for all the laughs, tears, motivating discussions and your genuine kindness.

Thank you to Ruaridh McFadyen (undergraduate Cardiff University project student) for the high quality imaging and morphological banding pattern and greyscale intensity data collections, which made a significant contribution to the work carried out in Chapter 4 of this thesis. Thank you for all your hard work. To Eleanor Sherrard-Smith, thank you for the use of your mitochondrial DNA sequences obtained during you undergraduate studies at Cardiff. To my sampling volunteers and very good friends, Pete, Mitch, Mawgan and Jeb, thank you for your hard work on some difficult field excursions and the entertaining times that we shared, some of which will never be forgotten!

To Jo Cable, thank you for the mentor role that you provided during the last two years and so kindly gave up your time for. Your kind words and valuable advice gave me confidence and encouraged me to persist. To Mike Bruford, thank you for the constructive comments over the years, the use of your lab and allowing me to contribute to it. The lab meetings were exceptionally helpful for solving technical issues and gave me valuable experience working as part of a hardworking and highly motivated team. Thank you to Carsten Müller, Paul Kemp and Rob Thomas; at some points over the last four years you have all given me advice, encouragement and support, for which I am very grateful.

To the Portuguese: Joana, Maf C, Rui, Tania and Maf B, thank you for invading G10, your continued support and cherished friendships. I will miss the many coffee and cigarette breaks that we all shared, come rain or shine. To all other members of G10, Milena, Jeff, Lucija, Pier, Renata, Gabi, Mireille, Big Dave, Little Dave, Xiang Zhan, Rob, Geoff, Tim, Dan, Adam, Vicky, Azniza, Catherine, Sigrid, Luke, Niall, James, Eddie, Pati, Loys, AnnaMaria and anyone else I have missed out, thank you for all the support and great times shared over the years. It has been enjoyable working with you all. Thank you for putting up with my ‘I hate DNA days’ and all the help and advice you gave, especially concerning technical issues.

I would like to thank all the staff at Cardiff University who have helped in some way or another over the years; particularly Steve and Steve from Cardiff University Molecular Support Unit, James Osborne and Steff Adams for help with data analyses and use of Condor computing facility, Joan, Ten, Steve, Chris and Allun. Much of the computational work performed in this study was carried out using the computational facilities of the Advanced Research Computing at Cardiff (ARCCA) Division, Cardiff University.

Finally, to my dearest friends and family, your love, support and friendships have helped me finally reach this important stage in my academic career, for which I thank you all. Ma and Steve, thank you for your unconditional love, support and inspirational adventures. To Pete, thank you for looking after me especially during the last six months, I couldn't have reached this far without you. You have stuck by me through good times and bad and given me all the love and encouragement anyone could ask for.

TABLE OF CONTENTS

Chapter 4: Ecological adaptation and early-stage sympatric divergence of *Aphrodes makarovi* (Hemiptera, Cicadellidae) into inland and estuarine lineages? Evidence from morphology, mtDNA and AFLP markers

Chapter 5: Introgressive hybridisation in *Aphrodes* leafhoppers (Hemiptera, Cicadellidae): exploring the mismatches between vibrational signals, mitochondrial DNA and AFLP genotypes

Chapter 6: General conclusions

Chapter 1: General Introduction

The main focus of the present study was elucidation of the nature and extent of differentiation and the evolutionary processes involved in shaping diversity within and between species of the *Aphrodes* genus, Curtis (1833), from the UK, with particular focus on saltmarsh inhabiting species. This chapter gives a general overview of speciation and the evolutionary processes explored in this study (1.1), the study system (1.2) and the molecular techniques employed (1.3).

1.1. Species and speciation

Inferences, regarding the evolutionary history of species, can be gained by examining the genetic relatedness and degree of differentiation between individuals and populations. Patterns of differentiation are the product of processes operating at two time scales: evolutionary time that incorporates broad-scale changes associated with environmental conditions, and ecological time over which population processes occur, such as demographic changes, migration and local extinction (Martin & Simon 1990). It is the relationships between these processes and the relative importance of these small scale and large-scale changes that explain the diversity of life that most evolutionary biologists are concerned with. It is, however, well acknowledged that significant evolutionary change can occur rapidly over short ecological time scales (tens of generations or less) (Carroll *et al.* 2007; Loxdale 2010). Population level comparisons involving a number of closely related taxa that vary in the level of differentiation among them form the basis of our understanding of species formation. Studying populations allows examination of variation among populations, historical associations and processes concerned with changes in their genetic structure that may have lead to speciation (Wright 1931; Knowles 2004; Knowles 2009).

Species concepts aim to divide biological diversity into meaningful discrete units i.e. species, and are numerous in the scientific literature (Coyne & Orr 2004; Mallet 2007a). It seems futile to attempt to debate further which is the most appropriate concept in general, as it is well documented that closely related species vary considerably in the

types (morphological, behavioural, ecological, genetic) and degree of differentiation among them. Thus, different species concepts (or a combination thereof) will be more or less appropriate for different taxa. Distinct species often show variation in morphological characters with little or no overlap between them (morphological species concept, Darwin 1859). The occurrence of morphological distinction in areas of sympatry is often used to infer reproductive isolation among taxa. However, when populations of morphological variants are found within a widespread species in different geographic regions, it is hard to discern whether they will remain distinct when they come in contact, and thus whether they represent distinct species or subspecies (Coyne & Orr 2004; Mallet 2007a).

Conversely, when sympatric species vary in traits other than morphology, they are commonly referred to as cryptic or sibling species (Mayr 1963; Quicke 1993). The biological species concept (BSC) or recognition species concept (Mayr 1963; Paterson 1985) highlights the importance of reproductive isolation and mate recognition, with no *a priori* assumptions of an association between speciation and morphology. These concepts have become increasingly more important due to the mounting evidence that overlapping morphological variability and cryptic species are a common phenomenon among invertebrates (Henry *et al.* 1999, 2002; Price *et al.* 2007; Joyce *et al.* 2010; Chapter 3).

The biological species concept (Mayr 1963) is the theory that forms our most widely accepted definition of species today (Coyne & Orr 2004) and states that reproductive isolation is the main driving force behind speciation and provides a definitive and rigid definition of species. At the species level, traits associated with mate choice and sexual selection are termed 'prezygotic' isolating mechanisms whereas hybrid inviability and sterility are termed 'postzygotic' isolating mechanisms. However, this concept is not completely fool proof due to increasing evidence that hybridisation and introgression between what are regarded as distinct species is more common than previously thought and can actually promote divergence (Mallet 2007b). Also, below the species level, there is increasing evidence showing differentiated populations classed as a single species that remain distinct in sympatry despite considerable gene flow (Drès & Mallet 2002; Mallet 2008).

Since the BSC, many other species concepts have been proposed, such as the recognition species concept (Paterson 1985), where sharing the same mate recognition system denotes species. This concept would define species similarly to the BSC, based on isolating mechanisms that keep species separate and a recognition system that ensures breeding occurs within a species. The ecological species concept (Van Valen 1976) relates to populations adapted to different niches and that divergent natural selection acting on traits between populations in different niches leads to the evolution of reproductive isolation (Schluter 2001). The phylogenetic species concept (Cracraft 1989) considers species as a cluster of organisms that is genetically diagnosable from other clusters, and within which there is a pattern of ancestry and descent. Identification of cryptic species has been aided with the use of DNA based taxonomy and phylogenetic analysis (Hebert *et al.* 2004; Pfenninger *et al.* 2006; Williams *et al.* 2006; King *et al.* 2008).

In the field of conservation biology, concepts have been proposed to identify distinct populations, such as Evolutionarily Significant Units (ESUs) and Management Units (MUs) (Moritz 1994; Fraser & Bernatchez 2001). Such biological classification of intraspecific diversity is useful for the conservation and management of endangered and exploited species (Moritz 1994) to prioritise taxa for conservation effort (Hammond *et al.* 2001; Chen *et al.* 2010).

Many other species concepts exist but generally speaking, sympatric groups of organisms that remain differentiated in sympatry will be accepted as distinct species (Cracraft 1989; Mallet 1995; Coyne & Orr 2004). Geographically distinct populations, however, may be either classed as distinct species or sub-species depending on the concept used (Mallet 2008). Arbitrary definitions are therefore unavoidable for geographically distinct populations (Mayr 1963). Biologists should therefore avoid the concept of discrete categories into which organisms should be allocated, but rather underline the level of variation within and among groups of organisms (Hendry *et al.* 2000). It is important to utilise a range of methods and to correlate a range of characters to delineate species (Sites & Marshall 2004), an approach that has been employed in studies involving morphologically cryptic taxa (Pfenninger *et al.* 2006; Price *et al.* 2007; Towes & Irwin 2008), and here within this thesis (Chapter 2 and 3). To gain understanding of the processes involved in how species form it is first important to examine the variation present using a number of traits to avoid erroneous judgements.

Speciation can occur rapidly and adaptive radiations are the extreme proof of this, whereby multiple colonisations lead to rapid diversification into multiple species each with different ecological adaptations. Examples of such huge diversity in nature comes from Hawaiian Drosophilidae (O'Grady & DeSalle 2008), *Anolis* lizards on Caribbean islands, Darwin's finches in the Galapagos islands, African lake cichlids (Schluter 2000). These unique biological systems have been studied extensively. A recent review indicates that a variety of genetic, ecological, developmental and historical geographical patterns influence the complex scenarios that exist and that much more data is needed to fully understand the processes concerned with adaptive radiations (Gavrilets & Losos 2009).

Geographic models of speciation have been described to explain how such diversity can arise and vary in the degree of geographic isolation between diverging populations and thus the extent of gene flow between them (Coyne & Orr 2004 for a review). Speciation in allopatry is a widely accepted model, whereby a physical barrier (due to climatic or geological events) separates populations, which are geographically isolated and thus divergence is not constrained by the homogenising effects of gene flow between populations. Populations accumulate genotypic and phenotypic differences as they undergo genetic drift, differing selection pressures (depending on the nature of the barrier), demographic events and random mutations. Reproductive isolation then follows as a consequence of the accumulation of genetic differentiation between populations. Peripatric speciation involves the formation of species in isolated, small peripheral populations that are prevented from exchanging genes with the main population either by colonisation (founder event) of an isolated habitat or by the isolation of a small population. This model is similar to allopatric speciation, but differs in the size of the populations involved (Coyne & Orr 2004).

Sympatric speciation requires the evolution of reproductive isolation in sympatry, despite gene flow between populations (Rice 1987; Dieckmann & Doebeli 1999; Kondrashov & Kondrashov 1999; Tregenza & Butlin 1999; Via 2001). Sympatric speciation remains controversial today due to the lack of substantial evidence in nature and simply because it is hard to rule out stages of allopatric divergence in the history of species. Divergence can be initiated when different populations evolve adaptations to different niches with varying ecological conditions (divergent natural selection), which provides a barrier to gene flow and allows long-term coexistence (Berlocher & Feeder

2002; Coyne & Orr 2004). Such models of speciation involving disruptive natural selection for specialisation (selection against intermediates) require a genetically determined niche preference (such as feeding, oviposition or mating), niche adaptation (trade-offs in fitness, alleles that increase ability to survive in one environment but reduce survival in others), and assortative mating (preferential mating with individuals inhabiting the same niche) (Rice 1987; Coyne & Orr 2004). Reproductive isolation can evolve in sympatry either due to direct disruptive/divergent selection affecting loci influencing habitat choice or indirectly due to pleiotropic effects (via disruptive selection on other traits) (see Via 2001 for a review).

African lake cichlids are the most diverse extant vertebrate adaptive radiations and provide unique systems for studying speciation and adaptive radiation (Seehausen 2006). Sympatric speciation models have been suggested that combine both natural and sexual selection to explain the coexistence of multiple species of cichlids in African lakes (Schliewen *et al.* 1994; Barluenga *et al.* 2006; Salzburger 2009). However, the relative roles of divergent sexual and natural selection, geography and hybridisation in promoting speciation of African lake Cichlids are still an extensively debated topic (Schliewen *et al.* 2001; Coyne & Orr 2004; Sparks 2004; Seehausen 2004, 2006; Gavrilets & Losos 2009; Salzburger 2009; Joyce *et al.* 2011). Sympatric speciation today is viewed as a likely and indeed confirmed process although we have little idea how common it might be (Via 2001; Berlocher & Feder 2002; Coyne & Orr 2004).

Parapatric speciation is intermediate between the allopatric and sympatric model extremes, where the zones of two diverging populations are separate but do overlap in a narrow zone of contact. Models of parapatric speciation involve clines, such as those found along ecological gradients. Subpopulations become adapted to their local habitat but this divergence is impeded by gene flow among adjacent and ecologically distinctive populations. Once differentiated populations arise, prezygotic isolation can evolve via reinforcement to reduce occurrence of intermediate hybrid genotypes with reduced fitness although other mechanisms do exist (Schluter 2001). Other modes of parapatric speciation include the stepping-stone model, where distinct populations undergo reduced gene flow, or they may include a combination of features from both types of parapatric speciation models (Coyne & Orr 2004).

Promoters of the BSC believe that for reproductive isolation (and thus speciation) to occur, some form of geographic isolation (allopatric divergence) is required. Under this concept hybrids are viewed as a breakdown of reproductive isolation as a consequence of secondary contact between isolated populations that had previously undergone allopatric divergence (Mallet 2008). Ecological or host races are likely to be explained by phenotypic plasticity or that differentiation between such populations evolved allopatrically, and that these races represent sibling species that became sympatric due to secondary contact (Mallet 2008). In light of the increasing amount of genetic data from natural populations, identification of numerous examples of hybridisation and introgression (interspecific gene flow) suggest that these phenomena are relatively common and can (and do) contribute to speciation (even in non-polyploid organisms) (Arnold 1992; Buerkle *et al* 2000; Coyne & Orr 2004; Mallet 2005, 2007b; Gompert *et al.* 2006a). Additionally, natural ecological races have been identified that coexist despite gene flow (Berlocher & Feder 2002; Drès & Mallet 2002; Mallet 2008).

Reproductive isolation is undoubtedly important for the formation and maintenance of the sexual populations we call species, but clearly is not the only factor that is important in the diversification for many taxa. In nature there is a whole continuum of variation, from a single panmictic population, to polymorphic populations, host races, sibling species to full species (Drès & Mallet 2002; Mallet 2008). Above the species level there are cases of hybridisation, introgression and hybrid speciation. Each stage of differentiation along this continuum may give different insights into processes of evolution above and below the species level. This understanding forms the basis of the work carried out in this thesis.

1.1.1. Ecological and host races

Insects contribute a major proportion of the world's biodiversity and are therefore suitable model organisms for studying ecology and evolution (Loxdale 2010). A recent increase in evolutionary studies on insects has provided additional supporting evidence for sympatric speciation, whereby ecologically driven selection leads to reproductive isolation (Via 2001; Drès & Mallet 2002; Mallet 2008). It is now acknowledged that evolutionary change, in an ecological context, can occur over relatively short time periods (tens of generations or less) in a wide range of taxa (Caroll *et al.* 2007) and in

insects is most commonly through host shifts or changes in chromosome number (Loxdale 2010).

Host shifts are thought to be a main reason for the high diversity of phytophagous insects observed in nature (Berlocher & Feder 2002; Funk *et al.* 2002), which is a known mechanism for finding new resources, thereby reducing inter- and intra specific competition and predation/parasitism pressures (Loxdale 2010). This permits closely related species to co-exist in the same habitat on different host plants (host races) and as a by-product, can facilitate reproductive isolation by providing a (partial) barrier to gene flow. A genetically determined host plant preference and assortative mating, with respect to host plant, are required for host-mediated speciation to occur, and host race formation is often viewed as the initial step towards sympatric speciation (Berlocher & Feder 2002).

A well-studied example of sympatric host race formation is the North American tethritid apple maggot fly, *Rhagoletis pomonella* (Walsh) (Feder *et al.* 1998; Feder *et al.* 1999; Berlocher & Feder 2002), which shifted from hawthorn (*Crataegus* sp.) to domesticated apple (*Malus pumila*) c. 150 years ago. Factors maintaining partial reproductive isolation include assortative mating (facilitated by host/habitat preference for oviposition and mating), pre- or postzygotic isolating barriers, including sex pheromones, host plant/fruit cues and incomplete allochronic isolation (differences in fruiting phenologies causing an offset in the life cycles of the two host races). Further empirical evidence for host races is the pea aphid, *Acyrthosiphon pisum*, found on two hosts, alfalfa (*Medicago sativae*) and red clover (*Trifolium pratense*) (Via 2001). These host races show adaptive genetic differences, strong preferences for native hosts, reduced fitness on non-native hosts (low fecundity and high mortality) and reduced hybrid fitness (on both hosts compared with individuals of the resident race) (Via 2001). Hawthorne & Via (2001) demonstrated a genetic trade-off in performance; through genetic mapping of F_2 hybrid performance they found a number of quantitative trait loci groups that had opposite effects on host performance.

Fitness trade-offs (i.e. adapted genotypes that are fitter in one environment but unfavourable in another (Schluter 2000)) are required for ecological speciation to occur, so that a generalist genotype cannot evolve as a product of recombination (Peccoud & Simon 2010). Generalist genotypes may evolve through recombination if particular

genomic regions affect fitness in one environment only. This may be caused by linkage disequilibrium between ecologically important alleles that had accumulated due to ancient geographical or reproductive isolation (Peccoud & Simon 2010). When a genetic trade-off in fitness across environments is seen then adaptations for both habitats cannot recombine in the same genome and divergent natural selection will prevent the evolution of generalist genotypes. Natural selection at the genetic level results from either antagonistic pleiotropy (alleles in two environments have antagonistic effects) or close linkage between alleles that have opposite fitness effects in different environments (Peccoud & Simon 2010). Furthermore, Via & West (2008) showed that 'divergence hitchhiking' around quantitative trait loci for traits that cause ecologically based reproductive isolation can reinforce divergence in sympatry. Genomic regions adjacent to quantitative trait loci under divergent selection can experience reduced recombination due to selection against hybrids (Via & west 2008).

Ecological specialisation can also be driven by adaptation along environmental and ecological gradients such as salinity or altitude, which could also explain the diversification of a widespread species (Bonin *et al.* 2006; Manel *et al.* 2009). It is very difficult to rule out previous allopatric divergence to explain patterns of genetic diversity and therefore studies examining very recently differentiated populations are as important as those looking at well-determined host races or sibling species to document all stages of divergence, from polymorphic populations to full species (Drès & Mallet 2002). However, many examples in nature suggest that speciation can be initiated through ecologically driven natural selection but often fails to complete speciation (Nosil *et al.* 2009). There is no way to determine if ecologically and genetically differentiated populations will continue to evolve into distinct species. However, once a population has reached a later stage of sympatric divergence, assortative mating and mate recognition make the recombination and fusion of diverged genotypes unlikely (Loxdale 2010).

1.1.2. Hybridisation, introgression and hybrid speciation

Hybridisation is a common phenomenon in natural plant and animal populations. However, hybridisation often leads to maladapted genotypes and hybrid sterility (Mallet 2008). The BSC states that species are reproductively isolated populations and since this concept was first mooted the idea that hybridisation and gene flow (introgression) between species could be important evolutionary mechanisms was discouraged (Mallet 2008). Hybridisation was only implied as an evolutionary process in contact zones between populations that had previously diverged in allopatry (secondary contact), where prezygotic barriers are reinforced as a response to selection against maladapted hybrids (Harrison 1993). The role of hybridisation as an evolutionary mechanism is becoming increasingly well acknowledged due to mounting evidence that it is more common than previously thought and can promote heterozygosity, adaptive potential and can even lead to speciation (Arnold 1992; Buerkle *et al* 2000; Coyne & Orr 2004; Seehausen 2004; Mallet 2005, 2007b; Gompert *et al.* 2006a).

If viable F_1 hybrids are formed between two genetically distinct parental taxa, backcrossing to parental types (introgression) can produce a mosaic of genotypes with differing proportions of parental contributions (clines in genetic proportions), which usually occurs in narrow contact zone, termed a hybrid zone (Barton & Hewitt 1989). Hybrid zones often coincide with ecotones or different habitat boundaries (Harrison 1993) and are characterised as unimodal, shallow clines (consisting of many intermediates, termed a hybrid swarm) or bimodal steep clines (mainly pure parental genotypes that are similar to each other with few intermediates) (Harrison 1993; Jiggins & Mallet 2000). Bimodality is thought to be associated with assortative mating or fertilisation, and to a lesser extent, total genetic divergence or intrinsic incompatibility (Jiggins & Mallet 2000). Bimodal hybrid zones can form due to ecological divergence (if the environment imposes selection on alleles) and stable contact zones are a common feature of host and ecological races (Drès & Mallet 2002; section 1.1.1), giving evidence for the sympatric route to speciation (Jiggins & Mallet 2000). Alternatively if selection against hybrids maintains a hybrid zone (due to secondary contact) the zone can move from place to place and is termed a tension zone (Barton & Hewitt 1985).

Introgression can often be asymmetrical between hybridising taxa, the extent of which will depend on whether the introgressing alleles show fitness advantages in the genetic

background into which they have introgressed. Introgression may give clues to recent movements in a hybrid zone (Harrison 1993). When alleles are favoured in different environments or genetic backgrounds then selection maintains differences, despite random mixing, and if alleles are universally favoured they may spread through the whole population (Barton & Gale 1993).

Selection can result in differences between gene genealogies in a population, because each gene has a distinctive history, which is determined by selection and mutation (Ballard & Whitlock 2004). It is therefore important to use a number of marker types to make inference about the history of species as conflicting genealogies can be obtained depending on the marker type. For example mitochondrial DNA (mtDNA) is known to introgress more readily than nuclear DNA (nDNA) (Ballard and Whitlock 2004) and there are many examples of animals that show mitochondrial genome introgression but no introgression of nuclear genes. Ferris *et al.* (1983) found the mtDNA from *Mus musculus domesticus* in populations of *M. m. musculus*. However, morphology, isoenzymes and nDNA sequences changed concordantly across the hybrid zone, suggesting that mitochondrial introgression had occurred due to an ancient founder event rather than persistent introgression across a current hybrid zone.

Fixation of mtDNA haplotypes from arctic charr (*Salvelnus alpinus*) was found in a population of lake trout (*S. namaycush*), whereas morphological and nDNA signatures in the introgressed lake trout population were typical (Wilson & Bernatchez 1998). The authors suggest that the most plausible explanation is that hybridisation occurred soon after deglaciation when populations came into contact, and repeated back-crossing of hybrids with the lake trout occurred (Wilson & Bernatchez 1998). Two possible reasons exist for the fixation of the mtDNA haplotype in the lake trout population, either due to chance (drift) or due to selection favouring the *S. alpinus* mtDNA type and/or associated nuclear genes (Bernatchez *et al.* 1995; Wilson & Bernatchez 1998). Gompert *et al.* (2006b) identified extensive mtDNA introgression in *Lycaeides* butterflies. Indirect selection for interspecific mtDNA in numerous nuclear genetic backgrounds was likely facilitated by a *Wolbachia* infection causing a selective sweep of a single mtDNA haplotype (Gompert *et al.* 2008). This demonstrates the potential for introgressive hybridization to have substantial and possible long-term effects on the genetic configuration of species and can produce considerable discrepancies among speciation histories based on nuclear and mitochondrial markers.

Hybridisation can lead to hybrid speciation (see Mallet 2007b for a review) and can occur due to genome duplication (allopolyploidy), which is a common mechanism in plants but rarer in animals (although it has been documented for some insects characterised by parthenogenic reproduction, Loxdale 2010). Doubling of chromosome numbers in polyploid hybrids facilitates reproductive isolation from both parent species (Mallet 2007b). Typically, when polyploid species mate with their diploid parent species, they produce sterile triploid progeny (Mallet 2007b).

Homoploid hybrid speciation, however, is harder to describe (Mallet 2007b). Two parent species hybridize and contribute genes to a daughter species without a change in chromosome number (genome remains diploid) (Buerkle *et al.* 2000) and can be termed recombinational hybrid speciation. Homoploid hybrid species are typically only partially reproductively isolated from parent species and it is hard to determine whether genetic patterns are due to hybridisation, introgression or from retention of an ancestral polymorphism (Mallet 2007b). Homoploid hybrid speciation is well known in flowering plants (Gross & Rieseberg 2005) and a few cases in animals have been documented including *Lycaeidies* butterflies (Gompert *et al.* 2006a), *Rhagoletis* fruit flies (Schwarz *et al.* 2005), *Heliconius* butterflies (Mavárez *et al.* 2006) and *Xiphophorus* fish (Meyer *et al.* 2006), and very recently two case in birds, yellow-rumped warbler (Brelsford *et al.* 2011) and sparrows (Hermansen *et al.* 2011).

Hybridisation can increase heterozygosity and while initially most homoploid hybrid recombinants are likely to be unfit, certain extreme hybrid genotypes may allow hybrids to exploit niches that are not available to the parents (Buerkle *et al* 2000; Mallet 2007b). For a distinct homoploid hybrid species to form, hybrid recombinants must be partially reproductively isolated from their parent species to prevent further backcrossing and introgression of parental genes, as well as being fit and competitively successful in their environment (Gross & Rieseberg 2005; Schwarz *et al.* 2005; Gompert *et al.* 2006a). For example, the alpine adapted *Lycaeides* butterfly is a hybrid species originating from the admixture of two distinct parent species *Lycaeides melissa* and *L. idas* (Gompert *et al.* 2006a). Based on two mtDNA genes, three nuclear genes, AFLP and microsatellite markers, the alpine form shows a mosaic genome that is distinctive from, and younger than that of both parent species. The high elevation habitat and different host plant species occupied by the homoploid hybrid species provides a barrier to gene flow with either parent species (Gompert *et al.* 2006a).

1.2. Study system

1.2.1. *Aphrodes* leafhoppers

Leafhoppers of the genus *Aphrodes* Curtis, 1833 (Hemiptera, Auchenorrhyncha, Cicadellidae, Aphrodinae) represent the model system chosen to explore the evolutionary processes discussed in section 1.1.1 and 1.1.2.

Aphrodes leafhoppers (Fig. 1.1) are abundant, widely distributed over the Palaearctic and have also been introduced to North America. They are univoltine, with egg development occurring overwinter (Nickel & Remane 2002; Nickel 2003). While they are important species in grassland leafhopper communities (Nickel & Achtziger 2005), they are also vectors of phytoplasmas that cause plant diseases (Weintraub & Beanland 2006; Wilson & Weintraub 2007). The four currently recognised species, *A. makarovi* (Zakhvatkin 1948), *A. bicincta* (Schrank 1776), *A. diminuta* (Ribaut 1952) and *A. aestuarina* (Edwards 1908) are very similar morphologically, making identification problematic (Hamilton 1975; Le Quesne 1988; Tishechkin 1998). Each species presents different ecological preferences, such as habitat type and host plant (Table 1.1). However, the degree of niche specialisation is unknown as in many cases their ranges overlap and more than one species can be found in the same habitat.

Table 1.1 Current *Aphrodes* species and their host plant preferences, modified from Edwards (1908), Kirby (1992), Tishechkin (1998) and Nickel (2003) and Biedermann & Niedringhaus (2004).

Species	**Host plant preference**	**Height found – meters above sea level (m a.s.l.)**	**Habitat**
A. makarovi (Zakhvatkin, 1948)	*Urtica, Taraxacum* and *Cirsium* sp.	< 800m a.s.l.	Ubiquitous, due to nature of host plants
A. bicincta (Schrank, 1776)	Fabaceae sp. - Meadows/abandoned fields	< 500m above sea level	Warm, humid or dry areas, with low productivity
A. diminuta (Ribaut, 1952)	Fabaceae sp. - Meadows	> 1000m a.s.l.	Cool and wet areas with more neutral soils
A. aestuarina (Edwards, 1908)	Salt-marsh - Shrubby seablite, *Suaeda vera* Sea Purslane, *Atriplex portulacoides*	~ 0m a.s.l.	Cold, wet, salty habitats, undergoing daily inundation

Sexual communication (conspecific mate recognition) in leafhoppers (and all Auchenorrhyncha, except cicadas) is facilitated exclusively by species-specific vibrational mating signals (Claridge 1985; Čokl & Virant-Doberlet 2003). In the *Aphrodes* genus, substrate borne vibrational mating signals have been shown to discriminate among species (Tishechkin 1998; Virant-Doberlet *et al.* 2005, 2006), as found in other phytophagous insect taxa (Virant-Doberlet & Čokl 2004).

Figure 1.1. Left, female and right, male *Aphrodes* leafhopper. Images sourced from T. and D. Pendleton (http://www.eakringbirds.com/eakringbirds6/insectinfocusaphrodes makarovi.htm) and T. Bantock (http://www.britishbugs.org.uk/Homoptera/ Cicadellidae/Aphrodes_makarovi.html), respectively. Accessed on 27/09/2011.

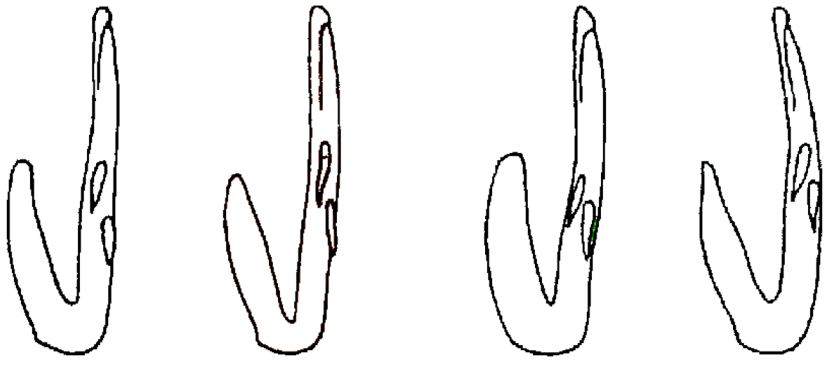
1.2.2. Taxonomy

There are divergent opinions on the taxonomic classification of the *Aphrodes* genus (Le Quesne & Payne 1981; Hamilton 1983; Kirby 1992; Tishechkin 1998; Nickel 2003), due to problems caused by the fact that morphological characters to distinguish males of the four currently recognized species are not reliable. Because of high morphological variability, these leafhoppers have often been designated the *Aphrodes bicincta* species group or complex (Le Quesne 1988; Tishechkin 1998). Irrespective of early species descriptions (Table 1.1), it was not until the late 1970's that *Aphrodes* were accepted as being a heterogeneous genus, and were simply divided into *A. bicincta* (smaller type) and *A. makarovi* (larger mesophilous type) (Hamilton 1983; Tishechkin 1998). Recently, the two other forms have been given species rank, the halophilous *A. aestuarina,* and the mesophilous *A. diminuta.* A study carried out by Tishechkin (1998) examining male species-specific vibrational calling signals (section 1.2.5) provides the most robust evidence for *Aphrodes* species designation. Relatively recently, the genus *Aphrodes* itself has also been split into several other genera (e.g. *Planaphrodes* and *Anoscopus*, Nickel 2003).

Because of inconsistencies in identifying species and several taxonomic revisions, there are many unresolved synonyms (Hamilton 1983; Tishechkin 1998). This has resulted in the unsatisfactory situation in which natural history collections often group all *Aphrodes* specimens under the generic name *A. bicincta* (Schrank 1776) or specimens are archived under separate names that are considered synonyms (for example *A. makarovi* (Zachvatkin 1948) and *A. costatus* (Panzer 1799) in the Fauna Europaea database). Furthermore, the problem is compounded by the fact that many archived specimens are females or nymphs for which accurate identification based on morphological characters is not possible.

1.2.3. Morphology

Previous morphological identification of *Aphrodes* relied on the differences in the positions of the spines on the male aedeagus (Fig. 1.2, Nast 1976; Tishechkin 1998). The use of male genitalia for identifying species is common in many insect groups (Quicke 1993). Other characteristics used to determine species status of males include the proportions of fore wings and vertex, shape of the penis shaft, body size and the shape of the sternal apodemes of abdominal segment II (Tishechkin 1998). These morphological characteristics are unreliable for classification because the coloration, size and aedeagal form are highly variable, and there is a considerable overlap in these characters between the separate species (Hamilton 1975; Le Quesne 1988; Tishechkin 1998). Additionally, descriptions were based on small sample sizes.



Aphrodes bicincta ***Aphrodes diminuta*** ***Aphrodes makarovi*** ***Aphrodes aestuarina***

Figure 1.2. Positions of the spines on the male's aedeagus and the shape of the penis shaft for the four *Aphrodes* species, *Aphrodes bicincta, Aphrodes diminuta, Aphrodes makarovi* and *Aphrodes aestuarina* (modified from Tishechkin 1998).

Aphrodes females are morphologically cryptic to the extent that no consistent differences have been found between them (Tishechkin 1998; Biedermann & Niedringhaus 2004). A possible reason for no morphological differentiation reported in females, may be that within mixed communities of *Aphrodes,* females may have been attributed incorrectly to a particular species, when identified only by the species identification of males present at the same locality. Thus, no consideration in previous studies was given to the possibility that several species may be present at a single locality.

1.2.4. Host plant / habitat preferences

Ecological differences between species have been observed, such as host plant preference, height found above sea level and habitat type (Kirby 1992; Tishechkin 1998; Nickel 2003) (Table 1.1). However, ranges of some species overlap and to what degree these preferences are fixed is debatable. *Aphrodes* males also signal on non-host plants (Virant-Doberlet, personal communication) and other studies have reported the polyphagous nature of some species (e.g. *A. makarovi*, Biedermann & Niedringhaus 2004).

While the species status of *A. makarovi, A. bicincta* and *A. diminuta* is well acknowledged by taxonomists (following Tishechkin 1998), classification of the halophilous *A. aestuarina* has been questioned due to no consistent morphological differences found from that of *A. makarovi* (Biedermann & Niedringhaus 2004) and vibrational mating signal data has been lacking for populations of *Aphrodes* inhabiting coastal saltmarsh habitats. *Aphrodes aestuarina* was classified as a distinct species due to its specific niche and slight differences in size (longer and thinner) and colouration (Edwards 1908), but with much caution as it is also documented as a possible synonym of *A. makarovi* (Kirby 1992; Nickel 2003). *Aphrodes aestuarina* is thought to be endemic to the UK (Tishechkin 1998), but specimens thought to be this species have been reported in saltmarsh habitats in the North and Baltic Sea, in Germany and Poland (Kirby 1992; Biedermann & Niedringhaus 2004).

Aphrodes aestuarina is described as being habitat specific, occurring only in saline meadows and coastal saltmarshes (Nickel 2003), adapted to survive daily inundation on Shrubby Seablite, *Suaeda vera* (Amaranthaceae, Subfamily Suaedoideae; previously known as *S. Fruticosa*), and occasionally on Annual Seablite, *S. maritima* (Edwards 1908; Kirby 1992). Sampling data from Virant-Doberlet (personal communication) suggests that *A. aestuarina* was collected from locations where little or no *S. vera* was present but mainly Sea Purslane, *Atriplex portulacoides* (Amaranthaceae, Subfamily Chenopodioideae; previously known as *Halimione portulacoides*, see also Kirby 1992). So the degree of host plant specificity is questionable.

1.2.5. Vibrational signals

Insect vibrational mating signals are mediated as vibrations through a substrate, such as their host plant, a mode of communication seen in other plant-feeding insects (Claridge 1985; Čokl & Virant-Doberlet 2003) and predatory insects such as lacewings (Henry *et al.* 2002). Vibratory signals are crucial for finding a mate and during courtship, providing a mechanism for species recognition (Claridge 1985). Different mating signals are likely to evolve due to adaptations to the signal transmission properties of their respective hosts (McNett & Cocroft 2008). Recognition of species based solely on vibrational mating signals is problematic for taxonomists because living specimens need to be examined and training by researchers to identify song phenotypes, to use specialised equipment, and to carry out laboratory based playback experiments is often required.

Adult *Aphrodes* males produce species-specific vibrational mating signals (Tishechkin 1998, Fig 1.3) to which con-specific females respond. This is thought to be associated with the prevention of inter-specific matings, as a form of reproductive isolation, as female responses show strong preference for male calling signals of their own species (Virant-Doberlet unpublished). Due to the morphological similarities among *Aphrodes* species, the male species-specific mating signals have been used for identification, which currently represents the most robust method for distinguishing among *Aphrodes* (Tishechkin 1998; Virant-Doberlet *et al.* 2005). Females of all four *Aphrodes* species emit similar vibrational mating signals that differ in their duration and in the click repetition time (Virant-Doberlet, unpublished data). However, playback experiments are

technically demanding and time consuming and therefore not practical. In addition, mated females won't respond to male signals (Virant-Doberlet, personal communication).

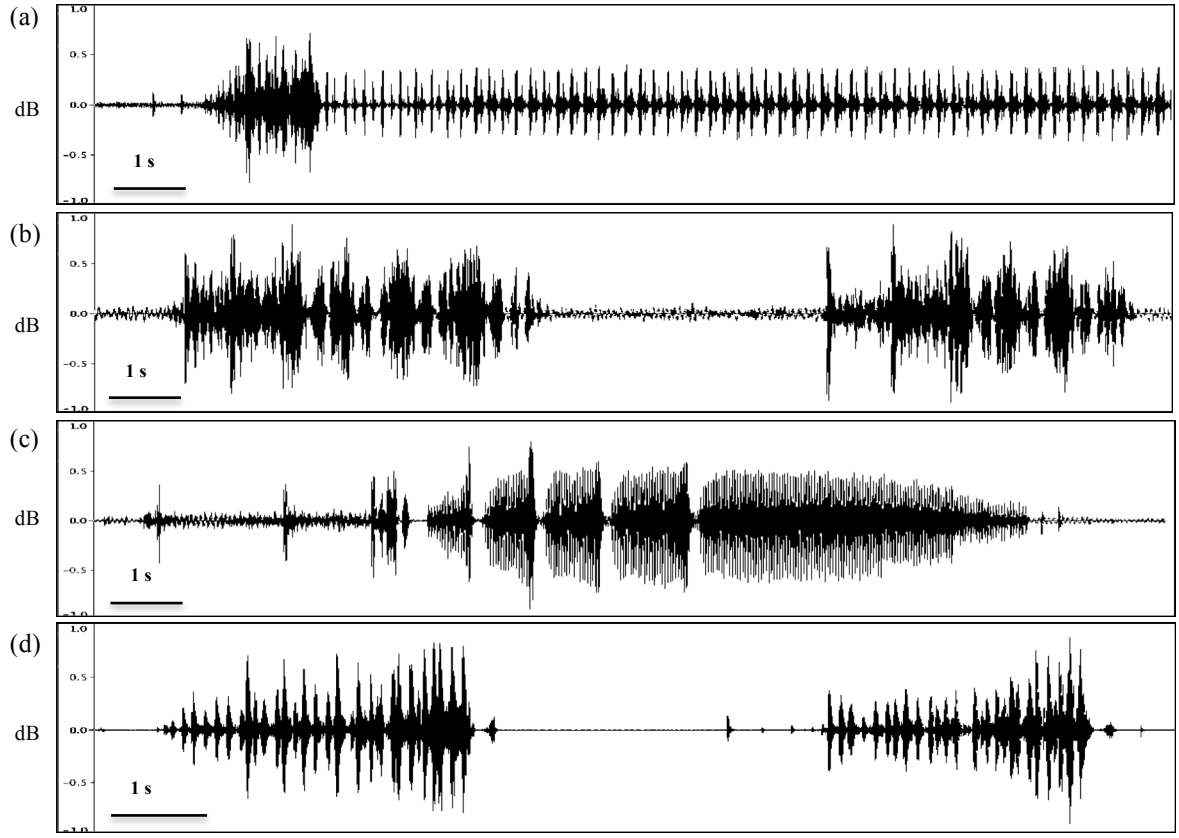


Figure 1.3. Acoustic signal patterns illustrating the composition of typical male calling songs of *Aphrodes* species (modified from Virant-Doberlet *et al.* 2005), a) *Aphrodes bicincta,* b) *A. diminuta,* c) *A. makarovi* and d) new signal recorded by M. Virant-Doberlet, later confirmed to be that of *A. aestuarina* (Chapter 2, Chapter 3).

1.2.6. Preliminary genetic data

In a series of unpublished undergraduate projects (Bluemel 2006; Sherrard-Smith 2007, unpublished data), genetic analysis of *Aphrodes* species was carried out after analyses of estuarine *Aphrodes* vibrational mating signals revealed two mating signals, *A. makarovi* (previously documented at inland sites on *Urtica* sp.) and a new mating signal, different to any signal previously described for this genus (Fig. 1.3d, Virant-Doberlet *et al.* 2005, 2006). It was unknown at the time whether this new signal belonged to that of the estuarine species *A. aestuarina* (as the mating signal has not been

previously recorded), a hybrid (as the new signal contains song elements similar to that found in *A. bicincta*) or possibly a new species. Preliminary unpublished molecular data using both amplified fragment length polymorphism (AFLP) and mtDNA cytochrome oxidase subunit I gene (COI) sequencing techniques (section 1.3) identified populations of estuarine *Aphrodes* in the Medway estuary (Kent, UK), that possessed the mtDNA sequence of *A. makarovi* but emitted the new recorded mating signal.

Preliminary AFLP data analyses (Bluemel 2006, unpublished data) suggested that hybridisation and introgression between *A. makarovi* and *A. bicincta* might explain the intermediate AFLP profiles, *A. makarovi* mtDNA and mating signal elements similar to those of *A. bicincta* found in these mismatched estuarine specimens (Bluemel 2006, unpublished data). Additionally, due to mating signal differences from both parent species, and thus the possibility of reproductive isolation based on mate choice, hypotheses were also proposed as to whether these mismatched specimens are in the initial stage of hybrid speciation. The evolution of a novel mating signal due to recombination could have then led to reproductive isolation of the hybrid population from the parent species, coupled with the strong selection pressure to survive daily tidal inundation and high salt concentrations in salt-marsh habitats, thus allowing the formation of a novel hybrid species. Host shift to an extreme environment and production of a novel vibrational communication signal would sufficiently facilitate reproductive isolation from parent species, and is likely to give a robust scenario for the formation of a hybrid species (Mallet 2007b). Examples of homoploid hybrid speciation often involve a habitat shift to an extreme environment or host plant shift where there is little competition from parental species, thus preventing back-crossing and facilitating reproductive isolation (Schwarz *et al.* 2005; Gompert *et al.* 2006a).

Further sampling of UK saltmarshes revealed a population of estuarine *Aphrodes* that emitted the same mating signal as identified in the Medway estuary mismatched specimens, but also distinct AFLP genotypes and mtDNA sequences, suggesting that these specimens were probably the originally described estuarine species *A. aestuarina* (Sherrard-Smith 2007, unpublished data). This was later confirmed in Chapter 2 (Bluemel *et al.* 2011), by comparing mtDNA sequences obtained from these specimens to those found in the *A. aestuarina* museum syntype series (Edwards 1908). The mismatched specimens from the Medway estuary also clustered with *A. aestuarina* specimens in preliminary AFLP analyses but clearly grouped with *A. makarovi* in

mtDNA phylogenetic analyses (Sherrard-Smith 2007, unpublished data). Because the mismatched specimens share the same host-plant and habitat preference as both of their parent species, isolation from *A. aestuarina* (to which there seems to be no behavioural barrier to gene flow as they share the same mating signal) for a long enough period of time to allow genetic divergence to occur is unlikely. This result led to the current hypothesis that hybridisation and mtDNA introgression may have occurred between *A. aestuarina* and *A. makarovi* to produce the genetic pattern observed, the likelihood of which is assessed in Chapter 5.

Additionally, identification of *A. makarovi* inhabiting estuarine as well as previously documented inland habitats (Chapter 2, Chapter 3), with very different host plant types and ecological factors and selection pressures led to the hypothesis that *A. makarovi* occupying different habitat types may represent two divergent ecological races. Varying selection pressures associated with each habitat type may promote adaptive divergence in *A. makarovi* populations, the likelihood of which is assessed in Chapter 4.

1.3. Molecular approaches

A combination of single gene (mtDNA COI sequencing) and genome-wide (AFLP marker) approaches were chosen to study the genus *Aphrodes*. The molecular techniques are discussed below, outlining the advantages and limitation of these methods.

1.3.1. Single gene approaches

1.3.1.1. Mitochondrial DNA sequencing

Mitochondrial DNA has proven to be a powerful tool in evolutionary biology, providing insights into population structure, gene flow, phylogenetic associations, hybridisation and biogeography (Wilson *et al.* 1985). Due to the specific qualities of the

molecule (discussed below), mtDNA has been heralded for providing a bridge between population genetics and systematics (Avise *et al.* 1987).

Mitochondrial DNA is a unique molecule that differs significantly from nDNA and represents only a small fraction of the genome (the mtDNA genome is c. 0.00055% of the total human genome) (Ballard & Whitlock 2004). Animal mtDNA is a duplex, covalently closed circular molecule that replicates itself and transcribes protein-coding genes within the organelle (Moritz *et al.* 1987). It consists of 15,000 – 17,000 bases, with two ribosomal RNA (rRNA), 22 transfer RNA (tRNA) and 13 protein coding genes, which code for subunits in the electron transport chain responsible for ATP production (Moritz *et al.* 1987; Ballard & Whitlock 2004). A 'control' region that contains sequences that initiate replication and transcription is also present (Moritz *et al.* 1987).

Mitochondrial DNA is maternally inherited in the majority of animal species, where all copies present in an individual are normally identical (homoplasmic) but can show high levels of intraspecific polymorphism (Wilson *et al.* 1985). High rates of evolution and easy amplification (via polymerase chain reaction, PCR) make this molecule particularly useful for studying closely related taxa (as mtDNA phylogenies evolve faster than nDNA). The mean rate of base substitution is higher in mtDNA than in nDNA (c. 10x faster initial rate of sequence divergence) (Wilson *et al.* 1985; Ballard & Whitlock 2004). Although different parts of the molecule evolve at different rates (Moritz *et al.* 1987) and this rate of nucleotide substitution cannot be assumed to be the same for all taxa (Ballard & Whitlock 2004). The rate of base substitution levels out after some time because many bases are conserved, particularly in protein coding regions at the first two codon sites (Moritz *et al.* 1987).

The protein-coding gene COI is relatively conserved and is known as a 'species describer'. Isolation and amplification can be easily achieved using 'universal' primers, such as those of Folmer *et al.* (1994), as used in this study. It has been used for evolutionary, phylogenetic, DNA barcoding studies and ancient DNA analysis (Jiggins 2003; Sparks 2004; Hebert *et al.* 2004; Hajibabaei *et al.* 2006; Williams *et al.* 2006; Gompert *et al.* 2006a, 2006b, 2008; Rowley *et al.* 2007; King *et al.* 2008).

The genetic transmission of the mtDNA occurs through the maternal line, typically without recombination (Wilson *et al.* 1985; Avise *et al.* 1987, although see Ballard & Whitlock 2004). This preserves information relating to ancestry and can provide insights concerning geographical structuring of populations where often nDNA cannot (Wilson *et al.* 1985; Moritz *et al.* 1987). However, a few cases are known where low levels of paternal leakage of mtDNA has occurred in a number of taxa, resulting in individuals possessing more than one mtDNA haplotype (heteroplasmic) due to biparental inheritance, which can affect the phylogenetic patterns of mtDNA (Wilson *et al.* 1985; Ballard & Whitlock 2004).

Due to its maternal inheritance, mtDNA alone is insufficient to identify the origins of hybrids or to highlight cases of introgression, where a mismatch of mitochondrial and nuclear genealogies is often seen (Wilson *et al.* 1985; Moritz *et al.* 1987; Bernatchez *et al.* 1995; Wilson & Bernatchez 1998; Shaw 2002; Seehausen 2004; Mallet 2005). Mitochondrial DNA can often be misleading when used to examine speciation histories in cryptic or young species radiations as time since speciation may be insufficient for differences to accumulate (incomplete lineage sorting) (Moritz *et al.* 1987). The genetic pattern produced by the retention of an ancestral polymorphism or incomplete lineage sorting is similar to that produced by introgressive hybridisation (Ballard & Whitlock 2004) and thus both nuclear and mitochondrial markers should be incorporated into studies where hybridisation is possible Chapter 5).

The fact that mtDNA is a maternally inherited haploid genome means that the effective population size is lower on average than that of nDNA, suggesting that mtDNA should fix new alleles faster relative to nDNA (Ballard & Whitlock 2004). Under neutral evolution the effective population size does not affect the substitution rate. However mtDNA cannot be assumed to be a neutral marker (Ballard & Whitlock 2004). If positive selection is acting on some nucleotide substitutions then the substitution rates will be much slower in genomes with a smaller effective population size (Ballard & Whitlock 2004). Important factors when using mtDNA to infer the rate of genetic drift should be considered. If certain mtDNA element/s result in fitness advantages and improve survival, then a selective sweep can drive the fixation of particular haplotypes associated with higher fitness. Additionally, background selection (elimination of low-fitness variants) will also reduce the effective population size of the mitochondria (Ballard & Whitlock 2004). Demographic events such as population expansion are also

expected to produce similar mitochondrial patterns to those expected from selective sweeps. Reduced mitochondrial DNA variation and loss of geographical structure through selective sweeps of a single mtDNA variant have been linked to *Wolbachia* infections (Jiggins 2003). Thus it is important to employ the use of multiple gene markers when inferring the phylogeographic history of species.

There are limitations that need to be considered when interpreting mtDNA sequence divergence as a standard for species identification (Meyer & Paulay 2005), and incorrect inferences can be caused by sequencing errors, through amplification of pseudogenes (nuclear copies of mtDNA known as numts, Benasson *et al.* 2001) or genetic sequence alignment errors (Löytynoja & Goldman 2008). This illustrates the importance of combining both nuclear and mitochondrial markers for evolutionary studies (Ballard & Whitlock 2004).

Mitochondrial DNA has an important role in animal biology and due to the interesting qualities of the molecule outlined above and the variety of processes that can affect mtDNA patterns, suggests that this marker should not be seen as only an additional tool in phylogenetic and population genetic studies. Mitochondrial DNA provides interesting avenues of research and further attempts should be made to explore the ecology and evolution of mtDNA and to understand the nature of selection acting on mtDNA itself (Ballard & Whitlock 2004).

1.3.1.2. Ancient DNA analysis of museum specimens using mitochondrial markers

Molecular analysis of ancient museum specimens provides a valuable tool for determining which morphologically similar species is equivalent to the first described type specimens and the phylogenetic relationships of those type specimens to other species (Austin & Arnold 2001; Austin & Melville 2006; Chapter 2 – Bluemel *et al.* 2011). Austin & Arnold (2001) amplified cytochrome b and tRNA-Glu gene sequences of type specimens of extinct Mascarene Island giant tortoise shells (*Cylindraspis*) of unknown island origin, to compare these with the remains of subfossil species of known origins. Primers for short (100-130 base pair (bp)), overlapping fragments were amplified with two rounds of PCR (where 1µl of the first PCR product is added to a second PCR). Roughly 400 bp of mitochondrial sequence was obtained for all museum

specimens analysed and their origins were identified from phylogenetic analysis of this sequence with high bootstrap support.

The large copy number of mitochondrial DNA occurring in each cell means that mitochondrial genes are favourable targets for museum specimen analysis (Hajibabaei *et al.* 2005). However, due to the degraded nature of ancient DNA it is time consuming and costly to amplify large sequences from specimens generally more than 10 years old (Rohland *et al.* 2004; Hajibabaei *et al.* 2005). Factors affecting DNA degradation include hydrolysis and oxidation, the presence of heat and time since death (Lindahl 1993) and the maximum amplifiable sequence or DNA survival depends on the degree of degradation. DNA degradation will be different for each specific sample and may depend on the particular collection method used (killing agent often depends on the preference of the collector) and the storage conditions since the time of collection. For these reasons it may not always be possible to successfully extract and amplify DNA from all specimens of this nature (Gilbert *et al.* 2007).

Hajibabaei *et al.* (2006) carried out a study to determine if a minimalist barcode sequence could accurately identify specimens that possess degraded DNA to species-level. Short sequences (~100 bp amplicons) were recovered for more than 90% of wasp and moth museum specimens analysed (ranging from 1-21 years in age), which proved highly efficient for species identification when comparisons were restricted to a closely related taxonomic group. Tests were also carried out for barcode data sets of Australian fish and Lepidopteran species, whereby full-length barcode data sets were cut down into shorter regions and the measures of sequence divergence and variability were compared between full and shorter length sequences (Hajibabaei *et al.* 2006). Results suggest that shorter length barcode sequences would provide as accurate species-level relationships as with the full-length sequences, although species resolution was lower when discriminating between species in large assemblages (~200 species of Australian fish) compared to the empirical data sets carried out within small assemblages (moth and wasp museum specimen tests). It was also noted that the position within the longer 'standard' fragment and the length of the shorter fragments was important for species discrimination in cases where very short sequences were analysed (135 bp in this case).

Using primers designed specifically for a taxonomic group may improve chances of amplifying degraded DNA, rather than using general primers, which also eliminates the

possibility of amplifying unwanted DNA from other taxonomic groups or human DNA, which would be the most likely source of contamination (Austin & Arnold 2001). It may also be important to use primers amplifying the full-length sequence as well as shorter sequences to rule out contamination from fresh DNA of those species analysed, because generally when using old material high molecular weight DNA would most likely be of recent origin. The use of extraction negatives and PCR controls containing no DNA would also help to rule this out as a possible source of contamination.

Extraction methods usually require at least partial dissection of specimens and can often affect the external integrity, which may be important if further analyses need to be carried out and particularly for museum specimens of important value, such as type specimens or rare species (Gilbert *et al.* 2007). It is therefore essential to choose a method causing the least damage and if possible to keep the external integrity of the specimens intact so that they can be put back into museum collections as voucher specimens (documented specimens used for research).

A number of studies using ancient and modern day DNA from terrestrial arthropods have addressed this issue, in which extractions were carried out using the intact specimen (Gilbert *et al.* 2007; Rowley *et al.* 2007) or even using small amounts of material for extraction, such as a single leg (Harper *et al.* 2006). The latter reduces external damage to specimens and enables specimens to be put back into museum collections. Of the studies that extracted DNA from specimens as a whole, both yielded DNA suitable for sequencing and effects of extraction buffers on the specimens ranged from no significant external damage/change (Gilbert *et al.* 2007) to slight discolouration to slight-to-moderate distortion of external features (Rowley *et al.* 2007), but not to the extent that morphological identification to the species level was affected and specimens were of suitable integrity to be put back into museum collections to act as morphological vouchers. However, for particularly important specimens, such as type specimens, use of a single appendage (if not important for species identification) may be more suitable rather than the whole specimen (Gilbert *et al.* 2007; Rowley *et al.* 2007), as only a small amount of tissue is needed for DNA isolation and sequencing (Hajibabaei *et al.* 2005).

1.3.2. Genomic approaches

1.3.2.1. Amplified Fragment Length Polymorphism (AFLP)

AFLP has proven to be a valuable genetic marker technique for population genetics, ecological and evolutionary studies (Bonin *et al.* 2007; Meudt & Clarke 2007) since first developed in the 1990's (Vos *et al.* 1995). AFLP has mainly been used for studies investigating economically important crop species, fungi and bacteria (Bensch & Åkesson 2005). Additionally, this technique has proven to be useful when distinguishing genetic boundaries, especially between cryptic taxa (Parsons & Shaw 2001; King *et al.* 2008; Toews & Irwin 2008). It has been increasingly used for a variety of organisms to examine genetic diversity, population structure, identify hybrids and to detect markers associated with phenotypes (Bonin *et al.* 2007).

AFLP is a highly sensitive technique that produces a barcoding pattern of high-resolution genome-wide DNA fragments (AFLP loci) (Bensch & Åkesson 2005). A large number of dominant loci can be assayed using this technique without the need for designing specific primers (Ajmone-Marsan *et al.* 1997; Bensch & Åkesson 2005), and is therefore particularly useful for studies based on non-model organisms (Meudt & Clarke 2007), as is the focus of this study. AFLP was applied here because it can be used to detect bi-parental inheritance, hybridization, and both inter-specific and intra-specific variation (Bensch & Åkesson 2005). AFLP is a technically demanding and fairly expensive technique, but it is more reliable than RFLPs (restriction fragment length polymorphism) (Ajmone-Marsan *et al.* 1997). Microsatellite techniques show higher levels of polymorphism but are initially more costly to produce than AFLPs (Ajmone-Marsan *et al.* 1997; Bensch & Åkesson 2005).

AFLP loci are generally treated as dominant markers and therefore have lower information content when compared with bi-allelic markers such as microsatellites (Vekemans *et al.* 2002). In a diploid individual, three genotypic classes can be obtained with multi-allelic markers (X_1X_1, X_1X_2, X_2X_2) but due to the dominance of AFLP markers only two character states, band presence (scored as 1, X_1X_1 and X_1X_2) and band absence (scored as 0, X_2X_2) are recorded (Hollingsworth & Ennos 2004). Because of the lower information content of AFLP markers, it is difficult to identify heterozygote individuals from individual homozygotes for the band presence allele

(Bonin *et al.* 2007). This disadvantage is counteracted by the large number of genome wide markers that are recovered using the AFLP technique compared to the relatively low number (5-20) of highly informative microsatellite markers typically used (Campbell *et al.* 2003).

The performance of AFLP has been assessed when compared to codominant markers such as microsatellites (Campbell *et al.* 2003; Gaudeul *et al.* 2004), which illustrated the relative effectiveness of both marker types. However, the majority of statistical methods have been designed specifically for codominant markers, which are often applied directly to AFLPs without detailed discussions or assessments of their suitability (Hollingsworth & Ennos 2004). It is therefore advised to use statistical approaches that are specifically designed for dominant markers or binary data and for the type of biological questions under investigation (Bonin *et al.* 2007).

Methods used to analyse AFLP data are either band-based approaches (distance measures based on the pattern of band presence or absence) or allele frequency approaches (estimates the allele frequency at each locus) (Bonin *et al.* 2007). The allele frequency approach is population orientated, and many methods require prior assumptions such as Hardy-Weinberg equilibrium (HWE), because the inbreeding coefficient is rarely known and cannot be reliably estimated for AFLP (although see Foll *et al.* 2008). The Bayesian approach (Zhivotovsky 1999) is robust against moderate departure from HWE and gives satisfactory estimates of null allele frequencies and is now routinely used (Bonin *et al.* 2007). Recently developed Bayesian methods designed specifically for dominant markers, that are particularly relevant for this study, include those for identifying hybrid individuals and genetic admixture proportions such as NEWHYBRIDS (Anderson & Thompson 2002) and STRUCTURE (Falush *et al.* 2007) and identifying genomic regions that may be associated with adaptive divergence, such as BAYESCAN (Foll & Gagoitti 2008). These approaches (and other band-based approaches) are discussed in more detail in Chapters 4 and Chapter 5.

Another drawback of AFLP (and microsatellites) is the effects of fragment-size homoplasy (Vekemans *et al.* 2002; Caballero *et al.* 2008), due to the lack of homology of co-migrating fragments. Here, fragments of a particular size may involve more than one locus and may lead to incorrect conclusions (Caballero *et al.* 2008). Other sources of error include technical errors, data handling and scoring errors and human errors that

can affect genotyping results (Pompanon *et al.* 2005; Meudt & Clarke 2007). Therefore it is important to report error rates calculated from a number of repeated genotypes (Bonin *et al.*2004; Whitlock *et al.* 2008).

1.4. Research questions

The main research questions explored in Chapters of this thesis are outlined below. Questions explored in Chapter 2 relate to ancient mtDNA analysis of *Aphrodes* museum specimens:

1. Can short, taxonomically informative mtDNA COI sequences be obtained and used to identify *Aphrodes* museum specimens, including c. 100 year old specimens from the syntype series of *A. aestuarina* (Edwards 1908) when compared to sequences obtained for freshly collected specimens that had been unequivocally identified as a member of a particular species?
2. Are the type specimens of *A. aestuarina* correct; do they identify a single, distinct taxonomic unit, or do they contain a mix of species?

Questions explored in Chapter 3 concern the degree of differentiation among freshly collected specimens of each *Aphrodes* species, with particular attention to those inhabiting estuarine habitats. Chapter 3 is closely linked with Chapter 2 and provided the validated data set used for ancient DNA primer design for *Aphrodes* museum specimens:

3. Are the four currently recognised *Aphrodes* species morphologically, behaviourally and genetically distinguishable from each other?
4. What is the extent of mtDNA variation and the phylogenetic relationships among *Aphrodes* species sampled across the UK?
5. Does *A. aestuarina* represent a distinct saltmarsh species or an ecological variant of *A. makarovi*?
6. What is the distribution of *Aphrodes* species inhabiting saltmarsh sites in the UK?

Questions addressed in Chapter 4 relate to the degree of intraspecific divergence among *A. makarovi* populations inhabiting inland and estuarine habitats:

7. Are populations of *A. makarovi* found in inland and estuarine habitats, occurring on different host plants, genetically and morphologically distinguishable from each other?
8. Are *A. makarovi* populations genetically structured with respect to habitat type or geographic locality?
9. Do populations inhabiting alternative habitat types possess nuclear AFLP loci showing high levels of genetic differentiation (high F_{ST} values) based on neutrality, to indicate that divergent natural selection may be acting on sympatric habitat associated populations of *A. makarovi*?

Questions addressed in Chapter 5 relating to the likelihood of hybridisation and introgression between *A. makarovi* and *A. aestuarina*:

10. What is the likelihood of hybridisation and mtDNA introgression, between *A. makarovi* and *A. aestuarina,* in specimens collected from the Medway estuary (Kent) that show a mismatch between vibrational mating signal and mtDNA sequence?
11. Is there any evidence of recent hybridisation (intermediate AFLP genotypes) or nuclear introgression (backcross hybrids) in these mismatched specimens?

1.5. Thesis outline

All chapters (except 1 and 6) are written as manuscripts that have been or will be submitted, in a slightly modified form, for publication in peer-reviewed journals. Chapter 2 has been published in *Molecular Ecology Resources* and is therefore presented in a different format to the following chapters and this work is referred to as Bluemel *et al.* 2011 throughout. A reference list is given at the end of each chapter rather than at the end of the thesis. Due to this format, there may be instances of duplication across chapters. Each chapter contains its own appendices where required.

CHAPTER 2: Primers for identification of type and other archived specimens of *Aphrodes* leafhoppers (Hemiptera, Cicadellidae)

This chapter outlines the methods used to design *Aphrodes* specific sets of PCR primers for amplifying and sequencing of short amplicons of the mtDNA COI gene to examine *Aphrodes* museum specimens of varying ages. A high number of misclassified specimens were identified using a relatively small sample of *Aphrodes* museum specimens, including a sample from the syntype series of *A. aestuarina* (Edwards 1908). It suggests that the species description for *A. aestuarina* was erroneous in that it was unknowingly based upon a mix of species. This work clearly underlines the need to validate museum specimens using molecular methods where identity is in doubt, based on reliable standards for species discrimination.

CHAPTER 3: Differentiation among species within the *Aphrodes* leafhopper genus (Hemiptera, Cicadellidae), comparing morphological, bioacoustic and DNA-based taxonomy.

A multi-disciplinary approach was taken to explore the variation among species within the *Aphrodes* leafhopper genus. This chapter is closely linked to Chapter 2 in that it provided the reliable standards for *Aphrodes* species that were subsequently used for the validation of museum specimens. This was carried out so that reliable identification of *Aphrodes* specimens could be achieved (particularly those found in estuarine habitats) and to provide a basis to explore further evolutionary questions discussed in the remainder of this thesis. Despite considerable morphological similarity, a combination of male vibrational mating signal/female preference and phylogenetic analyses of mtDNA COI gene sequence data provided good support for the existence of four *Aphrodes* species, which remain distinct in sympatry. This study highlights the value of taking a holistic approach when examining biological diversity and provided insights into exciting evolutionary processes (such as hybridisation, introgression and ecological adaptation) that may explain the patterns of morphological, behavioural and genetic variation reported.

CHAPTER 4: Ecological adaptation and early-stage sympatric divergence of *Aphrodes makarovi* (Hemiptera, Cicadellidae) into inland and estuarine lineages? Evidence from morphology, mtDNA and AFLP markers.

This chapter assesses the likelihood of divergent ecologically-driven natural selection driving divergence of *A. makarovi* populations inhabiting inland and estuarine habitats on two different primary host plants (*Urtica* sp. and *Atriplex portulacoides* respectively). A genome scan of 495 AFLP markers was carried out to examine the genetic structuring of nine UK populations of adult *A. makarovi,* collected from inland and estuarine habitats (including, at two sites, adjacent inland/estuarine populations). Significant morphological variation and genetic (AFLP) population structuring associated with habitat type was identified, relative to that explained by geographic locality. However, mtDNA sequence data revealed no structure relating to habitat or geographic locality. The lack of fixed divergent AFLP loci or significant mtDNA structure suggests that *A. makarovi* populations have diverged very recently and are in the early stages of sympatric ecotype formation. Initial divergence of inland and estuarine *A. makarovi* populations in allopatry cannot be ruled out. However, the observed pattern does not suggest that populations have experienced a long period of vicariance. Further work exploring the historical genetic structuring of *A. makarovi* (including populations from mainland Europe), the degree of host/habitat fidelity, fitness costs associated with each habitat/host plant, intrinsic genetic incompatibilities and hybrid fitness are needed.

CHAPTER 5: Introgressive hybridisation in *Aphrodes* leafhoppers: exploring the mismatches between vibrational signals, mitochondrial DNA and AFLP genotypes.

This chapter extends the research reported in Chapter 3 by addressing the likely cause of the mismatch between mating signal and mtDNA sequence data identified for specimens in the Medway estuary (Kent). A genome scan of 554 AFLP loci was carried out to explore the likelihood of hybridisation and introgression between *A. makarovi* and *A. aestuarina.* Unambiguous distinction between *A. makarovi* and *A. aestuarina* was recovered in nuclear AFLP Bayesian clustering analyses, concordant with mating signal data for all populations including the Medway estuary. Complete fixation of mtDNA of the *A. makarovi* lineage was observed in the mismatched Medway estuary *A. aestuarina* population and 95.6% of mismatched specimens were found to possess the most common *A. makarovi* haplotype (mH1), also present among sympatric Medway *A. makarovi*. Together these results reported suggest that interspecific mtDNA exchange is

likely to explain the reticulate evolutionary pathway generated for this mismatched population, rather than retention of an ancient ancestral polymorphism or convergence. Low levels of uni-directional nuclear introgression were observed, suggesting that the hybridisation event is likely to be of historical origin, followed by repeated backcrossing of hybrids with *A. aestuarina.* Reasons as to why *A. aestuarina* specimens were fixed for *A. makarovi* mtDNA in this population remains unclear. It is possible that a selective sweep of *A. makarovi* mtDNA occurred within *A. aestuarina* populations in the Medway estuary following introgressive hybridisation, possibly due to: a) an unknown direct fitness advantage to possession of *A. makarovi* mtDNA in this genetic background or in this region; b) due to indirect selection on the mtDNA genome due to cytoplasmic infections commonly identified in many invertebrate taxa; c) by chance (genetic drift). Unravelling the possible cause for mtDNA introgression in this population requires further work.

CHAPTER 6: General conclusions

The thesis concludes with a general synthesis of the major findings of this multidisciplinary study of inter- and intra-specific variation among *Aphrodes* leafhoppers. In addition, some recommendations for future research are made.

1.6. Relevance of the study

Apart from providing significant contributions to insect systematics and biodiversity present in the UK, this study provides an example of morphological crypsis within a closely related genus of behaviourally and genetically distinct invertebrate species.

The findings also have important evolutionary implications, adding to the growing body of literature showing that hybridisation and introgression (interspecific gene flow) above the species level is commonly identified in natural populations between what are described as distinct species. Additionally, ecological adaptation (due to divergent natural selection) is likely to be a major driving force in structuring the interspecific genetic variation in the early-stage divergence of currently sympatric populations of the phytophagous leafhopper, *A. makarovi.* These findings have important implications based on how we define species and the processes important for the persistence and initiation of species and speciation, above and below the species level.

1.7. References

Ajmone-Marsan, P., Valentini, A., Cassandro, M., Vecchiotti-Antaldi, G., Bertoni, G., & Kuiper, M. (1997). AFLP™ markers for DNA fingerprinting in cattle. *Animal Genetics, 28,* 418-426.

Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, *160*, 1217-1229.

Arnold, M. L. (1992). Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics*, *23*, 237-261.

Austin, J. J., & Arnold, E. N. (2001). Ancient mitochondrial DNA and morphology elucidate an extinct island radiation of Indian Ocean giant tortoises (*Cylindraspis*). *Proceedings of the Royal Society B: Biological Sciences*, *268*, 2515-2523.

Austin, J. J., & Melville, J. (2006). Incorporating historical museum specimens into molecular systematic and conservation genetics research. *Molecular Ecology Notes*, *6*, 1089-1092.

Avise, J. C., Arnold, R. M., Ball, E., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., & Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review Ecology and Systematics*, *18*, 489-522.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*, 729-744.

Barluenga, M., Stölting, K. N., Salzburger, W., Muschick, M., & Meyer, A. (2006). Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*, *439*, 719-723.

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*, 113-148.

Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*, 497-503.

Barton, N. H., Gale, K. S. (1993). Genetic analysis of hybrid zones. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), Oxford University Press, New York. (pp. 13–45).

Bensasson, D., Zhang, D.-X., Hartl, D. L., & Hewitt, G. M. (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution*, *16*, 314-321.

Bensch, S., & Åkesson, M. (2005). Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology*, *14,* 2899-2914.

Berlocher, S. H., & Feder, J. L. (2002). Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Review of Entomology*, *47*, 773-815.

Bernatchez, L., Glémet, H., Wilson, C. C., & Danzmann, R. G. (1995). Introgression and fixation of Arctic char (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of brook trout (*Salvelinus fontinalis*). *Canadian Journal of Fisheries and Aquatic Sciences*, *52*, 179-185.

Biedermann, R., & Niedringhaus, R. (2004). *Die Zikaden Deutschlands*. WABV, Scheeßel, (pp. 214-215).

Bluemel, J. (2006). Phylogeny and taxonomic status of cryptic European leafhoppers of the genus *Aphrodes* (Hemiptera: Auchenorrhyncha: Cicadellidae) investigated using AFLP and mitochondrial COI sequencing. *Unpublished Thesis.* Cardiff University.

Bluemel, J. K., King, R. A., Virant-Doberlet, M., & Symondson, W. O. C. (2011). Primers for identification of type and other archived specimens of *Aphrodes* leafhoppers (Hemiptera, Cicadellidae). *Molecular Ecology Resources, 11*, 770-774.

Bonin, A., Bellemain, E., Eidesen, P. B., Pompanon, F., Brochmann, C., & Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, *13*, 3261-3273.

Bonin, A., Taberlet, P., Miaud, C., & Pompanon, F. (2006). Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution, 23*, 773-783.

Bonin, A., Ehrich, D., & Manel, S. (2007). Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, *16*, 3737-3758.

Brelsford, A., Milá, B., & Irwin, D. E. (2011). Hybrid origin of Audubon's warbler. *Molecular Ecology*, *20*, 2380-2389.

Buerkle, C. A., Morris, R. J., Asmussen, M. A., & Rieseberg, L. H. (2000). The likelihood of homoploid hybrid speciation. *Heredity*, *84*, 441-451.

Caballero, A., Quesada, H., & Rolán-Alvarez, E. (2008). Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. *Genetics*, *179*, 539-554.

Campbell, D., Duchesne, P., & Bernatchez, L. (2003). AFLP utility for population assignment studies: analytical investigation and empirical comparison with microsatellites. *Molecular Ecology*, *12*, 1979-1991.

Carroll, S. P., Hendry, A. P., Reznick, D. N., & Fox, C. W. (2007). Evolution on ecological time-scales. *Functional Ecology*, *21*, 387-393.

Chen, L., Bruford, M. W., Xu, S., Zhou, K., & Yang, G. (2010). Microsatellite variation and significant population genetic structure of endangered finless porpoises (*Neophocaena phocaenoides*) in Chinese coastal waters and the Yangtze River. *Trends in Ecology & Evolution*, *157*, 1453-1462.

Claridge, M. F. (1985). Acoustic signals in the Homoptera: behaviour, taxonomy, and evolution. *Annual Review of Entomology, 30*, 297-317.

Čokl, A., & Virant-Doberlet, M. (2003). Communication with substrate-borne signals in small plant-dwelling insects. *Annual Review of Entomology, 48*, 29-50.

Coyne, J. A., & Orr, H. A. (2004). *Speciation.* Sinauer Associates, Sunderland, USA.

Cracraft, J., 1989. Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: Otte, D., Endler, J.A. (Eds.), *Speciation and its consequences*. Sinauer Associates, Sunderland, Massachusetts, (pp. 3-87).

Darwin, C. (1859) *The origin of species by means of natural selection*. Murray, London, U.K.

Dieckmann, U., & Doebeli, M. (1999). On the origin of species by sympatric speciation. *Nature*, *400*, 354-357.

Drès, M., & Mallet, J. (2002). Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *357*, 471-492.

Edwards, J. (1908). On some British Homoptera hitherto undescribed or unrecorded. *Entomology Monthly Magazine, 44*, 56-59.

Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, *7*, 574-578.

Feder, J. L., Berlocher, S. H. & Opp, S. B. (1998) Sympatric host-race formation and speciation in *Rhagoletis* (Diptera: Tephritidae): a tale of two species for Charles D. Genetic Structure and Local Adaptation in Natural Insect Populations (ed. by S. Mopper and S. Y. Strauss). Chapman & Hall, New York. (pp. 408–441).

Feder, J. L., & Filchak, K. E. (1999). It's about time: the evidence for host plant-mediated selection in the apple maggot fly, *Rhagoletis pomonella*, and its implications for fitness trade-offs in phytophagous insects. *Entomologia Experimentalis et Applicata*, *91*, 211-225.

Ferris, S. D., Sage, R. D., Huang, C. M., Nielsen, J. T., Ritte, U., & Wilson, A. C. (1983). Flow of mitochondrial DNA across a species boundary. *Proceedings of the National Academy of Sciences of the United States of America*, *80*, 2290-2294.

Foll, M., Beaumont, M. A., & Gaggiotti, O. (2008). An approximate Bayesian computation approach to overcome biases that arise when using amplified fragment length polymorphism markers to study population structure. *Genetics*, *179*, 927-939.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, *180*, 977-993.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & 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.

Fraser, D. J., & Bernatchez, L. (2001). Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, *10*, 2741-2752.

Funk, D. J., Filchak, K. E., & Feder, J. L. (2002). Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica*, *116*, 251-267.

Gaudeul, M., Till-Bottraud, I., Barjon, F., & Manel, S. (2004). Genetic diversity and differentiation in *Eryngium alpinum* L. (Apiaceae): comparison of AFLP and microsatellite markers. *Heredity*, *92*, 508-518.

Gavrilets, S., & Losos, J. B. (2009). Adaptive radiation: contrasting theory with data. *Science*, *323*, 732-737.

Gilbert, M. T. P., Moore, W., Melchior, L., & Worobey, M. (2007). DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS ONE*, *e272*, 1-4.

Gompert, Z., Fordyce, J. A., Forister, M. L., Shapiro, A. M., & Nice, C. C. (2006a). Homoploid hybrid speciation in an extreme habitat. *Science*, *314*, 1923-1925.

Gompert, Z., Nice, C. C., Fordyce, J. A., Forister, M. L., & Shapiro, A. M. (2006b). Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Molecular Ecology*, *15*, 1759-1768.

Gompert, Z., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2008). Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. *Molecular Ecology*, *17*, 5231-5244.

Gross, B. L., & Rieseberg, L. H. (2005). The ecological genetics of homoploid hybrid speciation. *Journal of Heredity*, *96*, 241-252.

Hajibabaei, M., DeWaard, R. J., Ivanova, N. V., Ratnasingham, S., Dooh, R. T., Kirk, S. L., Mackie, P. M., et al. (2005). Critical factors for assembling a high volume of DNA barcodes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*, 1959-1967.

Hajibabaei, M., Smith, M. A., Janzen, D. H., Rodriguez, J., Whitfield, J. B., & Hebert, P. D. N. (2006). A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, *6*, 959-964.

Hamilton, K. G. A. (1975). A review of the Northern hemisphere *Aphrodina* (Rhynchota: Homoptera: Cicadellidae) with special reference to the Nearctic fauna. *The Canadian Entomologist*, *107*, 1009–1027.

Hamilton, K. G. A. (1983). Introduced and native leafhoppers common to the old and new worlds (Rhynchota: Homoptera: Cicadellidae). *Canadian Entomologist, 115*, 473-511.

Hammond, R. L., W, M., Flores, B., Mohammed, O. B., Wacher, T., & Bruford, M. W. (2001). Phylogenetic reanalysis of the Saudi Gazelle and its implications for conservation. Conservation Biology, *15*, 1123-1133.

Harper, G. L., Maclean, N., & Goulson, D. (2006). Analysis of museum specimens suggests extreme genetic drift in the adonis blue butterfly (*Polyommatus bellargus*). *Biological Journal of the Linnean Society*, *88*, 447-452.

Harrison, R. G. (1993). Hybrids and hybrid zones: historical perspective. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), Oxford University Press, Oxford. (pp. 3–12).

Hawthorne, D. J., & Via, S. (2001). Genetic linkage of ecological specialization and reproductive isolation in pea aphids. *Nature*, *412*, 904-907.

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical

skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences, 101*, 14812-14817.

Hendry, A. P., Vamosi, S. M., Latham, S. J., Heilbuth, J. C., & Day, T. (2000). Questioning species realities. *Conservation Genetics*, *1*, 67-76.

Henry, C. S., Brooks, S. J., Johnson, J. B., & Duelli, P. (1999). Revised concept of *Chrysoperla mediterranea* (Hölzel), a green lacewing associated with conifers: courtship songs across 2800 kilometres of Europe (Neuroptera: Chrysopidae). *Systematic Entomology*, *24*, 335-350.

Henry, C. S., Brooks, S. J., Duelli, P., & Johnson, J. B. (2002). Discovering the true *Chrysoperla carnea* (Insecta: Neuroptera: Chrysopidae) using song analysis, morphology, and ecology. *Annals of the Entomological Society of America*, *95*, 172-191.

Hermansen, J. S., Saether, S. A., Elgvin, T. O., Borge, T., Hjelle, E., & Saetre, G-P. (2011). Hybrid speciation in sparrows I: phenotypic intermediacy, genetic admixture and barriers to gene flow. *Molecular Ecology*, *20*, 3812-3822.

Hollingsworth, P. M., & Ennos, R. A. (2004). Neighbour joining trees, dominant markers and population genetic structure. *Heredity*, *92*, 490-498.

Jiggins, C., & Mallet, J. (2000). Bimodal hybrid zones and speciation. *Trends in Ecology & Evolution*, *15*, 250-255.

Jiggins, F. M. (2003). Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics*, *164*, 5-12.

Joyce, A. L., Bernal, J. S., Vinson, S. B., Hunt, R. E., Schulthess, F., & Medina, R. F. (2010). Geographic variation in male courtship acoustics and genetic divergence of populations of the *Cotesia flavipes* species complex. *Entomologia Experimentalis et Applicata*, *137*, 153-164.

Joyce, D. A, Lunt, D. H., Genner, M. J., Turner, G. F., Bills, R., & Seehausen, O. (2011). Repeated colonization and hybridization in Lake Malawi cichlids. *Current Biology*, *21*, R108-109.

King, R. A., Tibble, A. L., & Symondson, W. O. C. (2008). Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. *Molecular Ecology*, *17*, 4684-4698.

Kirby, P. (1992). A review of the scarce and threatened Hemiptera of Great Britain. No. 2. *The Joint Nature Conservation Committee.* (pp. 154).

Knowles, L. L. (2004). The burgeoning field of statistical phylogeography. *Journal of Evolutionary Biology*, *17*, 1-10.

Knowles, L. L. (2009). Statistical phylogeography. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 593-612.

Kondrashov, A. S., & Kondrashov, F. A. (1999). Interactions among quantitative traits in the course of sympatric speciation. *Nature*, *400*, 351-354.

Le Quesne, W. J., & Payne, K. R. (1981). *Cicadellidae (Typhlocybinae) with a checklist of the British Auchenorrhyncha (Hemiptera, Homoptera).* Vol. II, Part 2(c). Fitton, M.G. (ed.). Royal Entomological Society of London. (pp. 45).

Le Quesne, W. J. (1988). The *Aphrodes bicinctus* complex. *Proceedings of the 6th Auchenorryncha meeting, Turin, Italy.* (pp. 413–415).

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, *362*, 709-715.

Löytynoja, A., & Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*, *320*, 1632-1635.

Loxdale, H. D. (2010). Rapid genetic changes in natural insect populations. *Ecological Entomology*, *35*, 155-164.

Mallet, J. (1995). A species definition for the modern synthesis. *Trends in Ecology & Evolution*, *10*, 294-299.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*, 229-237.

Mallet, J (2007a) *Species, concepts of.* In: Levin, SEA, (ed.) *Encyclopedia of Biodiversity*. Elsevier: Oxford. (pp. 427-440).

Mallet, J. (2007b). Hybrid speciation. *Nature*, *446*, 279-283.

Mallet, J. (2008). Hybridisation, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363,* 2971-2986.

Manel, S., Conord, C., & Després, L. (2009). Genome scan to assess the respective role of host-plant and environmental constraints on the adaptation of a widespread insect. *BMC Evolutionary Biology*, *9*, 1-10.

Martin, A., & Simon, C. (1990). Differing levels of among-population divergence in the mitochondrial DNA of periodical cicadas related to historical biogeography. *Evolution, 44*, 1066-1080.

Mavárez, J., Salazar, C. A., Bermingham, E., Salcedo, C., Jiggins, C. D., & Linares, M. (2006). Speciation by hybridization in Heliconius butterflies. *Nature*, *441*, 868-871.

Mayr, E. (1963). *Animal species and evolution.* Harvard University Press, Cambridge.

McNett, G. D., & Cocroft, R. B. (2008). Host shifts favor vibrational signal divergence in *Enchenopa binotata* treehoppers. *Behavioral Ecology*, *19*, 650-656.

Meudt, H. M., & Clarke, A. C. (2007). Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science*, *12*, 106-117.

Meyer, C. P., & Paulay, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, *3*, 1-10.

Meyer, A., Salzburger, W., & Schartl, M. (2006). Hybrid origin of a swordtail species (Teleostei: Xiphophorus clemenciae) driven by sexual selection. *Molecular Ecology*, *15*, 721- 730.

Moritz, C., Dowling, T. E., & Brown, W. M. (1987). Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics*, *18*, 269-292.

Moritz, C. (1994). Defining "Evolutionarily Significant Units" for conservation. *Trends in Ecology & Evolution*, *9*, 373-375.

Nast, J. (1976). Peiwiki (*Homoptera, Auchenorrhyncha*) Peinin. *Fragmenta Faunistica*. XXI, 157-160.

Nickel, H., & Remane, R. (2002). *Check list of the planthoppers and leafhoppers of Germany, with notes on food plants, diet width, life cycles, geographic range and conservation status (Hemiptera, Fulgoromorpha and Cicadomorpha)*. Artenliste der Zikaden Deutschlands.

Nickel, H. (2003). *The leafhoppers and planthoppers of Germany (Hemiptera, Auchenorrhyncha): patterns and strategies in a highly diverse group of phytophagous insects.* Pensoft Publishers, Sofia-Moscow. (pp. 103-105).

Nickel, H., & Achtziger, R. (2005). Do they ever come back? Responses of leaf-hopper communities to extensification of land use. *Journal of Insect Conservation, 9,* 319–333.

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, *24*, 145-156.

O'Grady, P., & Desalle, R. (2008). Out of Hawaii: the origin and biogeography of the genus *Scaptomyza* (Diptera: Drosophilidae). *Biology Letters*, *4*, 195-199.

Parsons, Y. M., & Shaw, K. L. (2001). Species boundaries and genetic diversity among Hawaiian crickets of the genus *Laupala* identified using amplified fragment length polymorphism. *Molecular Ecology*, *10*, 1765 -1772.

Paterson, H. E. H. (1985). The recognition concept of species. In *Species and Speciation* (E.S. Vrba, Ed.), Transvaal Museum Monograph No. 4. Transvaal Museum, Pretoria. (pp. 21-29).

Peccoud, J., & Simon, J-C. (2010). The pea aphid complex as a model of ecological speciation. *Ecological Entomology*, *35*, 119-130.

Pfenninger, M., Cordellier, M., & Streit, B. (2006). Comparing the efficacy of morphological and DNA-based taxonomy in the freshwater gastropod genus *Radix* (Basommatophora, Pulmonata). *BMC Evolutionary Biology, 6*, 1-14.

Pompanon, F., Bonin, A., Bellemain, E., & Taberlet, P. (2005). Genotyping errors: causes, consequences and solutions. *Nature*, *6*, 847-859.

Price, B. W., Barker, N. P., & Villet, M. H. (2007). Patterns and processes underlying evolutionary significant units in the *Platypleura stridula* L. species complex (Hemiptera: Cicadidae) in the Cape Floristic Region, South Africa. *Molecular Ecology, 16,* 2574-2588.

Quicke, D. L. J. (1993). *Principles and techniques of contemporary taxonomy*. Blackie Academic & Proffessional, London.

Rice, W. R. (1987). Speciation via habitat specialization: the evolution of reproductive isolation as a correlated character. *Evolutionary Ecology*, *1*, 301-314.

Rohland, N., Siedel, H., & Hofreiter, M. (2004). Nondestructive DNA extraction method for mitochondrial DNA analyses of museum specimens. *BioTechniques*, *36*, 814-821.

Rowley, D. L., Coddington, J. A., Gates, M. W., Norrbom, A. L., Ochoa, R. A., Vandenberg, N. J., & Greenstone, M. H. (2007). Vouchering DNA-barcoded specimens: test of a nondestructive extraction protocol for terrestrial arthropods. *Molecular Ecology Notes*, *7*, 915-924.

Salzburger, W. (2009). The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. *Molecular Ecology*, *18*, 169-185.

Schliewen, U., Tautz, D., & Pääbo, S. (1994). Sympatric speciation by monophyly of crater lake cichlids. *Nature*, *368*, 629-632.

Schliewen, U., Rassmann, K., Markmann, M., Markert, J., Kocher, T., & Tautz, D. (2001). Genetic and ecological divergence of a monophyletic cichlid species pair under fully sympatric conditions in Lake Ejagham, Cameroon. *Molecular Ecology*, *10*, 1471-1488.

Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford University Press.

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology and Evolution*, *16*, 372-380.

Schwarz, D., Matta, B. M., Shakir-Botteri, N. L., & McPheron, B. A. (2005). Host shift to an invasive plant triggers rapid animal hybrid speciation. *Nature*, *436*, 546-549.

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, *19*, 199-207.

Seehausen, O. (2006). African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences*, *273*, 1987-1998.

Shaw, K. L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 16122-16127.

Sherrard-Smith, E. (2007). Investigating possible speciation through hybridisation in the leafhopper genus *Aphrodes* (Hemiptera: Auchenorrhyncha: Cicadellidae) using AFLP analysis and mitochondrial COI sequencing. *Unpublished Thesis.* Cardiff University.

Sites, J. W., & Marshall, J. C. (2004). Operational criteria for delimiting species. *Annual Review of Ecology, Evolution, and Systematics*, *35*, 199-227.

Sparks, J. S. (2004). Molecular phylogeny and biogeography of the Malagasy and South Asian cichlids (Teleostei: Perciformes: Cichlidae). *Molecular Phylogenetics and Evolution*, *30*, 599-614.

Tishechkin, D. Y. (1998). Acoustic signals and morphological characters of leafhoppers belonging to *Aphrodes bicinctus* group (Homoptera, Cicadellidae) from European Russia. *Entomological Review, 78*, 370-377.

Towes, D. P., & Irwin, D. E. (2008). Cryptic speciation in a Holarctic passerine revealed by genetic and bioacoustics analyses. *Molecular Ecology, 17*, 2691-2705.

Tregenza, T., & Butlin, R. K. (1999). Speciation without isolation. *Nature*, *400*, 311-312.

Van Valen, L. (1976). Ecological species, multispecies, and oaks. *Taxon 25,* 233-239.

Vekemans, X., Beauwens, T., Lemaire, M., & Roldán-Ruiz, I. (2002). Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, *11,* 139-151.

Via, S. (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution*, *16*, 381-390.

Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, *17*, 4334-4345.

Virant-Doberlet, M., & Čokl, A. (2004). Vibrational communication in insects. *Neotropical Entomology, 33*, 121–134.

Virant-Doberlet, M., Prešern, J., Stewart, A. J. A., & Wilson, M. R. (2005). The genus *Aphrodes*: bioacoustic approach to solve taxonomic problems. *Congress of International BioAcoustic Council (IBAC).* (pp. 35).

Virant-Doberlet, M., Prešern, J., Bluemel, J., King, A. R., Wilson, M. R., Symondson, W. O. C., & Stewart, A. J. A. (2006). Vibrational signals and reproductive isolation in the genus *Aphrodes* (Hemiptera: Cicadellidae). *11th International Meeting on Invertebrate Sound and Vibration*, University of Toronto at Scarborough.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Fritjers, A., Pot, J., Peleman, J., Kuiper, M., & Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, *23*, 4407-4414.

Weintraub, P. G., & Beanland, L. A. (2006). Insect vectors of phytoplasmas. *Annual Review of Entomology*, *51*, 91–111.

Whitlock, R., Hipperson, H., Mannarelli, M., Butlin, R. K., & Burke, T. (2008). An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Molecular Ecology Resources*, *8,* 725-735.

Williams, H. C., Ormerod, S. J., & Bruford, M. W. (2006). Molecular systematics and phylogeography of the cryptic species complex *Baetis rhodani* (Ephemeroptera, Baetidae). *Molecular Phylogenetics and Evolution*, *40*, 370-382.

Wilson, A. C., Cann, R., Carr, S. M., George, M., Gyllensten, U., Helm-Bychowski, K., Higuchi, R., Palumbi, S., Prager, E., Sage, R., & Stoneking, M. (1985). Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnean Society*, *26*, 375-400.

Wilson, C. C., & Bernatchez, L. (1998). The ghost of hybrids past: fixation of arctic charr (*Salvelinus alpinus*) mitochondrial DNA in an introgressed population of lake trout (*S. namaycush*). *Molecular Ecology*, *7*, 127-132.

Wilson, M. R., & Weintraub, P. G. (2007). An introduction to Auchenorrhyncha phytoplasma vectors. *Bulletin of Insectology*, *60*, 177-178.

Wright, S. (1931). Evolution in mendelian populations. *Genetics, 16*, 97-159.

Zhivotovsky, L. A. (1999). Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, *8,* 907-913.

Chapter 2:

Primers for identification of type and other archived specimens of *Aphrodes* leafhoppers (Hemiptera, Cicadellidae)

Reference:

Bluemel, J. K., King, R. A., Virant-Doberlet, M., & Symondson, W. O. C. (2011). Primers for identification of type and other archived specimens of *Aphrodes* leafhoppers (Hemiptera, Cicadellidae). *Molecular Ecology Resources, 11*, 770-774, (including supporting information, section 2.1).

(N.B. Page numbers have been modified for the purpose of this thesis)

MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2011) **11**, 770–774 doi: 10.1111/j.1755-0998.2011.03008.x

Primers for identification of type and other archived specimens of *Aphrodes* leafhoppers (Hemiptera, Cicadellidae)

JOANNA K. BLUEMEL,* R. ANDREW KING,*‡ META VIRANT-DOBERLET*† and WILLIAM O. C. SYMONDSON*
**Cardiff School of Biosciences, Cardiff University, Biomedical Sciences Building, Museum Avenue, Cardiff CF10 3AX, UK,*
†*Department of Entomology, National Institute of Biology, Večna pot 111, SI-1000 Ljubljana, Slovenia*

Abstract

Primers were developed for leafhoppers of the genus *Aphrodes* amplifying 84–244 bp fragments of the mitochondrial cytochrome oxidase subunit *I* gene. DNA was extracted from legs of over 100-year-old archived museum specimens, amplified and sequenced. The fragments contain sufficient variation to unequivocally identify the different species. The majority of the analysed museum specimens, including three specimens of the syntype series for the UK endemic species *A. aestuarina* (Edwards 1908), were found to have been assigned to the wrong species. This work clearly underlines the need to validate museum specimens using molecular methods where identity is in doubt, based on reliable standards for species discrimination.

Keywords: cytochrome oxidase subunit I, molecular diagnostics, mtDNA, museum specimens, taxonomy

Received 7 December 2010; revision received 11 February 2011; accepted 17 February 2011

Museum and other natural history collections represent an important source of genetic material (Austin & Melville 2006; Wandeler *et al.* 2007), and DNA obtained from museum specimens has been widely used in studies of phylogenetics and phylogeography (Stuart & Fritz 2008), populations genetics (Harper *et al.* 2006) and conservation genetics (Crandall *et al.* 2009). Museum collections also archive type specimens which were originally used to describe species based on morphological characters. For taxonomists, type specimens are essential in either verifying species status of fresh material (Austin & Melville 2006) or to validate the taxonomic status of type material according to updated knowledge (Graham *et al.* 2004).

Leafhoppers of the genus *Aphrodes* Curtis, 1833 (Hemiptera: Cicadellidae), are abundant, widely distributed over the Palaearctic and have also been introduced to North America. While they are important species in grassland leafhopper communities (Nickel & Achtziger 2005), they are also vectors of phytoplasmas that cause plant diseases (Weintraub & Beanland 2006). Problems have been caused by the fact that morphological characters to distinguish males of the four currently recognized species are not reliable. The coloration, size and aedeagal form are highly variable, and there is a considerable overlap in the morphological characters used to separate species (Hamilton 1975; Le Quesne 1988; Tishechkin 1998). Furthermore, more than one species can be found on the same site sharing the same habitat. Because of high variability, these leafhoppers have often been designated the *Aphrodes bicincta* species group or complex (Le Quesne 1988; Tishechkin 1998). In addition, because of inconsistencies in identifying species and several taxonomic revisions, there are many unresolved synonyms (Hamilton 1983; Tishechkin 1998). This has resulted in the unsatisfactory situation in which natural history collections often group all *Aphrodes* specimens under the generic name *A. bicincta* (Schrank 1776) or specimens are archived under separate names that are considered synonyms (for example *A. makarovi* Zachvatkin 1948 and *A. costatus* (Panzer 1799) in the Fauna Europaea database). Furthermore, the problem is compounded by the fact that many archived specimens are females or nymphs for which accurate identification based on morphological characters is currently not possible. In our work, we followed species determination after Tishechkin (1998) based on species-specific vibrational signals (further details in Methods S1, Supporting Information). Recorded vibrational signals were compared with previously described signals for *A. makarovi*, *A. bicincta* and *A. diminuta* Ribaut 1952 (synonymous to *A. centrorossica* Zachvatkin 1953) (Tishechkin 1998). To avoid creating additional confusion, we adopted species identification based on the male calling vibrational signals described by Tishechkin (1998) throughout.

Correspondence: Joanna Bluemel, Fax (0)29 20 874116; E-mail: BluemelJK1@cardiff.ac.uk
‡Present address: College of Life and Environmental Sciences, Geoffrey Pope Building, Stocker Road, Exeter, EX4 4QD, UK.

General invertebrate primers that can amplify the barcoding region of the cytochrome oxidase subunit I (COI) gene (Folmer *et al.* 1994) cannot be used on older museum specimens because the DNA becomes too degraded (Hajibabaei *et al.* 2006). The aim of our work was therefore to (i) design primers for amplification of shorter, but taxonomically informative, fragments of the mitochondrial (mtDNA) COI gene; and (ii) to evaluate primer suitability for amplifying degraded DNA extracted from museum specimens.

Primers were developed from the full 658-bp fragment of the COI gene amplified from a subset of fresh *Aphrodes* specimens using the general invertebrate primers LCO1490 and HCO2198 (Folmer *et al.* 1994). DNA was extracted from legs of multiple individuals of all four *Aphrodes* species using DNeasy Blood and Tissue kit (Qiagen, Crawley, UK) following the manufacturer's spin-column protocol. Species identity was previously determined by bioacoustic methods (see Methods S1, Supporting Information). PCR amplifications and sequencing of mtDNA were carried out as described in King *et al.* (2008). Sequences were aligned using CLUSTALX (Thompson *et al.* 1997), and primers were designed using AMPLICON (Jarman 2004). To incorporate the extent of variation within the genus, all species and haplotypes of *Aphrodes* found were included in the alignments. We first designed the primer pair APH-COI-F1/APH-COI-R1 (Table 1), which amplify a 244-bp amplicon, within which are 34 polymorphic sites and 22 species-specific sites (4–11 sites for each species). Two forward and two reverse primers were designed internally to this 244-bp fragment in order to obtain shorter overlapping amplicons covering the whole length of the fragment. Primers were tested on 10 recently collected (2–4 years old) specimens of each *Aphrodes* species identified by bioacoustic methods and stored in 100% ethanol. DNA was extracted from legs as described earlier and a 658-bp fragment of the COI gene was sequenced using the general invertebrate primers LCO1490 and HCO2198 (Folmer *et al.* 1994). Genbank accession numbers for each haplotype are provided as full 658-bp sequences (FR727167–FR727179). We then extracted DNA from 14 museum specimens collected 25 and 103 years ago (Table 2), using the DNeasy Blood and Tissue kit (QIAGEN) (see Methods S1, Supporting Information). Either whole specimens, body without legs or only unhomogenized legs were incubated for 12 or 24 h (Table 2) and a 100 μl elution with AE buffer was carried out in the final step, to increase the DNA concentration. In cases when the body (intact specimens or body without legs) was used for nondestructive DNA extraction, specimens were photographed before and after extraction (Rowley *et al.* 2007) (see Methods S1, Supporting Information) to see how the integrity of archived specimens was affected

Table 1 Primer sequences, annealing temperatures the length of amplified fragments and the number of polymorphic and species-specific sites

Fragment	Primer name	Primer sequence (5′-3′)†	T_a	Fragment size (bp)‡	Polymorphic sites	*Aphrodes* species-specific sites§			
						A. bicincta	*A. makarovi*	*A. aestuarina*	*A. diminuta*
A	APH-COI-F1	TAGATTTATTATTCGTATTGAAC	58	244	34	4	4	6	11
	APH-COI-R1	ATAAACAGTTCAACCAGTACCA							
B	APH-COI-F1		54	100	9	1	0	2	4
	APH-COIint-R1	AATTYCCAAACCCRCCAATT							
C	APH-COI-F1		54	140	17	2	2	2	7
	APH-COIint-R2	AATGCYATATCAGGRGCACC							
D	APH-COIint-F1	AATTGGYGGGTTTGGRAATT	54	124	22	2	3	4	6
	APH-COI-R1								
E	APH-COIint-F2	GGTGCYCCTGATATRGCATT	54	84	15	2	2	3	4
	APH-COI-R1								

COI, cytochrome oxidase subunit I.
†Degenerate base codes; Y = C/T, R = A/G.
‡Fragment lengths include amplified target sequence only.
§For details, see Fig. S2 (Supporting Information).

Table 2 Details of archived *Aphrodes* specimens used in the study

Museum number	Collection date	Gender	Species as listed†	Extraction	Preservation technique	Lysis time (h)	Amplified fragment‡	Species revealed§	Species-specific sites¶
NMW.2.1984-054	17 August 1985	Female	*Aphrodes* sp.	Body + legs	Dried, pinned	12	A	*makarovi*	4/4
NMW.2.1984-054	17 August 1985	Female	*Aphrodes* sp.	Body	Dried, pinned	12	A	*makarovi*	4/4
				Legs		12	A	*makarovi*	4/4
NMW.2.1984-054	17 August 1985	Female	*Aphrodes* sp.	Body + legs	Dried, pinned	24	A	*makarovi*	4/4
NMW.2.1984-054	17 August 1985	Female	*Aphrodes* sp.	Body	Dried, pinned	24	A	*makarovi*	4/4
				Legs		24	A	*makarovi*	4/4
1929.20.325	31 August 1907	Male	*bicincta*	Body + legs	Dried, glued on card	12	C, D	*diminuta*	11/11
1929.20.326	31 August 1907	Male	*bicincta*	Body	Dried, glued on card	12	C, D	*makarovi*	4/4
				Legs		12	C, D	*makarovi*	4/4
1929.20.327	31 August 1907	Male	*bicincta*	Body + legs	Dried, glued on card	12	C, D	*makarovi*	4/4
1929.20.328††	28 August 1907	Female	*aestuarina*	Legs	Dried, glued on card	24	C, D	*makarovi*	4/4
1929.20.329††	28 August 1907	Female	*aestuarina*	Legs	Dried, glued on card	24	D	*makarovi*	3/3
1929.20.330††	28 August 1907	Female	*aestuarina*	Legs	Dried, glued on card	24	–	–	–
1929.20.331††	28 August 1907	Male	*aestuarina*	Legs	Dried, glued on card	24	C	*makarovi*	2/2
1929.20.332††	28 August 1907	Male	*aestuarina*	Legs	Dried, glued on card	24	B, C, D, E	*aestuarina*	6/6
1929.20.333††	28 August 1907	Male	*aestuarina*	Legs	Dried, glued on card	24	B, C, D, E	*aestuarina*	5/6‡‡
1929.20.334††	28 August 1907	Male	*aestuarina*	Legs	Dried, glued on card	24	B, C, D, E	*aestuarina*	6/6

†Species name under which specimens were archived.
‡See Table 1 for details.
§Species revealed from the mtDNA sequence.
¶Number of species-specific sites in recovered sequences.
††Specimens taken from the syntype series for *Aphrodes aestuarina*.
‡‡Only five of six diagnostic bases could be resolved for this specimen because of an ambiguous base found in the sequence for fragment B and C.

during the extraction process. Each fragment was amplified separately using the Qiagen multiplex kit in 10-μL reactions containing 1 or 2 μl of extracted DNA, 5 μl of 2× Multiplex PCR mix, 0.2 μM of each of the primer and 0.2 μg BSA. PCRs were carried out in BioRad DNAengine PTC 200 thermal cycler (Hemel Hempsted, UK) under the following conditions: 95 °C for 15 min, followed by 35 cycles of 94 °C for 30 s, appropriate annealing temperature (Table 1) for 90 s, 72 °C for 90 s and a final cycle of 72 °C for 10 min. PCR products were visualized under UV light on 2.5% agarose gels stained with ethidium bromide. Negative controls (extraction and PCR negatives) and a positive control (*Aphrodes* sp. DNA) were included in each PCR and subsequently sequenced. The PCR products were cleaned using 0.25 U each of exonuclease I and shrimp alkaline phosphatase incubated at 37 °C for 45 min and 80 °C for 15 min. Products were sequenced directly using the three sets of primers on a 3130xl Genetic Analyser (Applied Biosystems, Warrington, UK) using BigDye (version 3.1) sequencing chemistry in both forward and reverse direction.

DNA was successfully extracted and at least one fragment was recovered from 13 museum specimens (Table 2). Extractions from the whole body or legs both yielded DNA suitable for PCR amplification and sequencing. Younger specimens (25 years old) did not show significant external distortion and/or body damage even after 24 h lysis (except some loss of pigmentation.) However, the integrity of older museum samples (103 years old) was poor after DNA extraction (see Fig. S1, Supporting Information).

The complete 244-bp fragment was obtained only from specimens up to 25 years old (GenBank accession numbers FR727154–FR727157). Nevertheless, sequences obtained from shorter overlapping fragments enabled reconstruction of the full 244-bp fragment for 103-year-old material (see Methods S2 and Fig. S3, Supporting Information) (GenBank accession numbers FR727158–FR727166). Furthermore, any one of the shorter fragments was sufficient to discriminate between species by itself (see Figs S4 and S5, Supporting Information). Results showed that when *Aphrodes* specimens were labelled with species names in museum collections, the majority (78%) of archived specimens were assigned to the wrong species (Table 2).

This study showed that whole-body nondestructive extraction of DNA from up to 25-year-old dried archived insects is possible using a commercially available extraction kit. Furthermore, extraction from leafhopper legs yielded sufficient DNA for many rounds of PCR amplification. However, because of various factors, DNA degradation in museum specimens can differ substantially (Gilbert *et al*. 2007) and the amount of extracted DNA suitable for PCR amplification can differ greatly between specimens. As archived type specimens should be returned back into a collection, we recommend that only legs should be used for extraction. In leafhoppers, legs are not used as a systematic character and therefore removing them does not significantly reduce the value of the specimen.

Single-nucleotide misincorporations have been detected in mitochondrial DNA sequences obtained from archived specimens (Nyström *et al*. 2006; Stiller *et al*. 2006; Sefc *et al*. 2007; Zimmermann *et al*. 2008). While this could lead to overestimation of haplotypes present in archived specimens (Sefc *et al*. 2007), it could also potentially lead to incorrect species assignment when only short sequences (~ 100 bp) with a low number of informative sites are used for species resolution within a closely related cryptic species complex. In this study, however, we found no new haplotypes among the analysed museum specimens.

As archive specimens kept in museum collections become increasingly important in genetic analyses (e.g. Harper *et al*. 2006; Stuart & Fritz 2008; Crandall *et al*. 2009), errors in taxonomic identification can represent a major issue (Graham *et al*. 2004; Wandeler *et al*. 2007). For taxa like *Aphrodes*, for which accurate taxonomical identification based on morphological characters is difficult, incorrectly labelled museum specimens that are used to verify species status of fresh material can create long-lasting identification errors that are hard to resolve. This is underlined by the fact that at least three of the specimens that represent a syntype series for the purported UK endemic species *A. aestuarina* belong to another species (only seven of the 13 syntype specimens were tested in this study). These specimens have already been used in taxonomic studies (Tishechkin 1998). The high rate of incorrectly labelled specimens found in this study indicates that, whenever possible, archived specimens of such difficult taxa should be verified by molecular methods. Museum specimens should be validated by comparison with freshly collected individuals that have been unequivocally identified as members of a particular species, based on a multidisciplinary approach (morphological, behavioural and molecular). Only sequences obtained from such specimens could provide a reliable standard for species discrimination of archived material.

Resolving the status of taxa and synonyms within the *Aphrodes* complex is beyond the scope of the present study; however, the approach described here offers an opportunity to use archived type specimens to resolve the taxonomy of this group.

Acknowledgements

We are grateful to Tony Irwin, Norfolk Museum Service, for the kind loan of the 103-year-old archive specimens. We would like

to thank Mike Wilson, National Museum of Wales, for advice regarding *Aphrodes* taxonomy, the kind loan of the 25-year-old archive specimens and for the use of photographic equipment and imaging software. We thank Alan Stewart, University of Sussex, for help in the field and for advice on *Aphrodes*. Isa-Rita Russo, Cardiff University, is thanked for constructive comments on earlier drafts of the manuscript and for help with phylogenetic analyses.

References

Austin JJ, Melville J (2006) Incorporating historical museum specimens into molecular systematic and conservation genetics research. *Molecular Ecology Notes*, **6**, 1089–1092.

Crandall KA, Robinson HW, Buhay JE (2009) Avoidance of extinction through non-existence: the use of museum specimens and molecular genetics to determine the taxonomic status of an endangered freshwater crayfish. *Conservation Genetics*, **10**, 177–189.

Folmer O, Black M, Hoeh W *et al.* (1994) DNA primers for the amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Gilbert MTP, Moore W, Melchior L *et al.* (2007) DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS ONE*, **2**, e272. doi: 10.1371/journal.pone.0000272.

Graham CH, Ferrier S, Huettman F *et al.* (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution*, **19**, 497–503.

Hajibabaei M, Smith AM, Janzen DH *et al.* (2006) A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, **6**, 959–964.

Hamilton KGA (1975) A review of the Northern hemisphere *Aphrodina* (Rhynchota: Homoptera: Cicadellidae) with special reference to the Nearctic fauna. *The Canadian Entomologist*, **107**, 1009–1027.

Hamilton KGA (1983) Introduced and native leafhoppers common to the old and new worlds (Rhynchota: Homoptera: Cicadellidae). *The Canadian Entomologist*, **115**, 473–511.

Harper GL, Maclean N, Goulson D (2006) Analysis of museum specimens suggests extreme genetic drift in Adonis blue butterfly (*Polyommatus bellargus*). *Biological Journal of the Linnean Society*, **88**, 447–452.

Jarman SN (2004) Amplicon: software for designing PCR primers on aligned DNA sequences. *Bioinformatics*, **20**, 1644–1645.

King RA, Tibble AL, Symondson WOC (2008) Opening a can of worms: unprecedented sympatric cryptic diversity within British lumbricid earthworms. *Molecular Ecology*, **17**, 4684–4698.

Le Quesne WJ (1988) The *Aphrodes bicinctus* complex. Proceedings of the 6th Auchenorryncha meeting, Turin, Italy, pp. 413–415.

Nickel H, Achtziger R (2005) Do they ever come back? Responses of leafhopper communities to extensification of land use. *Journal of insect Conservation*, **9**, 319–333.

Nyström V, Angerbjörn A, Dalén L (2006) Genetic consequences of a demographic bottleneck in the Scandinavian artic fox. *Oikos*, **114**, 84–94.

Rowley DL, Coddington JA, Gates MW *et al.* (2007) Vouchering DNA-barcoded specimens: test of a nondestructive extraction protocol for terrestrial arthropods. *Molecular Ecology Notes*, **7**, 915–924.

Sefc KM, Payne RB, Sorenson MD (2007) Single base errors in PCR products from avian museum specimens and their effect on estimates of historical genetic diversity. *Conservation Genetics*, **8**, 879–884.

Stiller M, Green RE, Ronan M *et al.* (2006) patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proceedings of the National Academy of Sciences USA*, **103**, 13578–13584.

Stuart BL, Fritz U (2008) Historical DNA from museum type specimens clarifies diversity of Asian leaf turtles. *Biological Journal of the Linnean Society*, **94**, 131–141.

Thompson JD, Gibson TJ, Plewiak F *et al.* (1997) CLUSTAL X Windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research*, **25**, 4876–4882.

Tishechkin DY (1998) Acoustic signals and morphological characters of leafhoppers belonging to *Aphrodes bicinctus* group (Homoptera, Cicadellidae) from European Russia. *Entomological Review*, **78**, 370–377.

Wandeler P, Hoeck PEA, Keller LF (2007) Back to the future: museum specimens in population genetics. *Trends in Ecology and Evolution*, **22**, 634–642.

Weintraub PG, Beanland LA (2006) Insect vectors of phytoplasmas. *Annual Review of Entomology*, **51**, 91–111.

Zimmermann J, Hajibabaei M, Blackburn DC *et al.* (2008) DNA damage in preserved specimens and tissue samples: a molecular assessment. *Frontiers in Zoology*, **5**, 18. doi: 10.1186/1742-9994-5-18.

Supporting Information

Additional supporting information may be found in the online version of this article.

Methods S1 Species determination, DNA extraction from archived specimens and morphological analysis pre- and post extraction.

Methods S2 Phylogenetic analysis.

Fig. S1 *Aphrodes* museum specimens before (A, B) and after (a, b) 24 h lysis. (A, a) 25 year old specimen (A: magnification 2×, 219 mp/mm; a: magnification 1.6×; 177 mp/mm).

Fig. S2 Sequences of a 244 bp fragment (fragment A Table 1) of the cytochrome oxidase subunit *I* gene for four species in the genus *Aphrodes*.

Fig. S3 Neighbour joining phylogeny using Kimura 2-parameter distances for the 244 bp fragment of cytochrome oxidase I sequence depicting the relationships between 13 museum specimens and all currently known haplotypes for *Aphrodes makarovi*, *Aphrodes bicincta*, *Aphrodes aestuarina* and *Aphrodes diminuta*.

Fig. S4 Neighbour joining phylogeny using Kimura 2-parameter distances for fragment C (140 bp) of cytochrome oxidase I sequence depicting the relationship between 12 museum specimens and all currently known haplotypes for *Aphrodes makarovi*, *Aphrodes bicincta*, *Aphrodes aestuarina* and *Aphrodes diminuta*.

Fig. S5 Neighbour joining phylogeny using Kimura 2-parameter distances for fragment D (124 bp) of cytochrome oxidase I sequence obtained from 12 museum specimens and all currently known haplotypes for *Aphrodes makarovi*, *Aphrodes bicincta*, *Aphrodes aestuarina* and *Aphrodes diminuta*.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

2.1. Supporting Information

Species determination

The species identity of all individuals used in this study, except museum specimens, was determined by recording their vibrational signals using a laser vibrometer (PDV 100, Polytec GmbH, Waldbronn, Germany) (methods described in Virant-Doberlet & Žežlina 2007; Mazzoni *et al*. 2009). *Aphrodes bicincta*, *A. makarovi* and *A. diminuta* (= *A. centrorossica*) were determined by comparing recorded signals with the ones previously described by Tishechkin (1998). Vibrational signals of *A. aestuarina* had not been previously recorded. Identification of this species was based on specimens from the site where syntype specimens had been collected, the species description of Edwards (1908), specific ecology (daily inundation during high tide) and vibrational signals that differ from the ones previously described for other species.

DNA extraction from archived specimens

All extractions involving museum specimens were conducted in an isolated laboratory and precautions were taken to detect and minimise contamination (Gilbert *et al.* 2007; Wandeler *et al.* 2007; Bantock *et al.* 2008; Lee & Prys-Jones 2008). A new DNeasy Blood and Tissue Kit (Qiagen) was purchased and used only in this laboratory. The work surfaces and equipment were cleaned with 10% bleach and the room, equipment and consumables were decontaminated by 24h UV treatment prior to extraction. When legs were used for extraction, after each dissection forceps were thoroughly cleaned by flaming. Each series of extractions included only up to seven specimens and a negative extraction control. Either the whole specimen, body without legs or only the legs were incubated for 12 or 24 h and a 100 µL elution with AE buffer were carried out in the final step.

Morphological analysis pre- and post extraction

Museum specimens were imaged before and after DNA extraction (Fig. S1). They were photographed using AUTO-MONTAGE PRO version 5.0 (Synoptics) imaging software and a JVC KY F70 3CCD digital camera mounted on a Leica M28 stereo microscope (1.6x or 2.0x magnification depending on size of the specimen) with a Planapo chromatic 1x lens attached. Before extraction they were photographed dry still pinned or glued to the card. After extraction the specimens were fixed in a trough made of white-tack inside an excavated block filled with 100% ethanol.

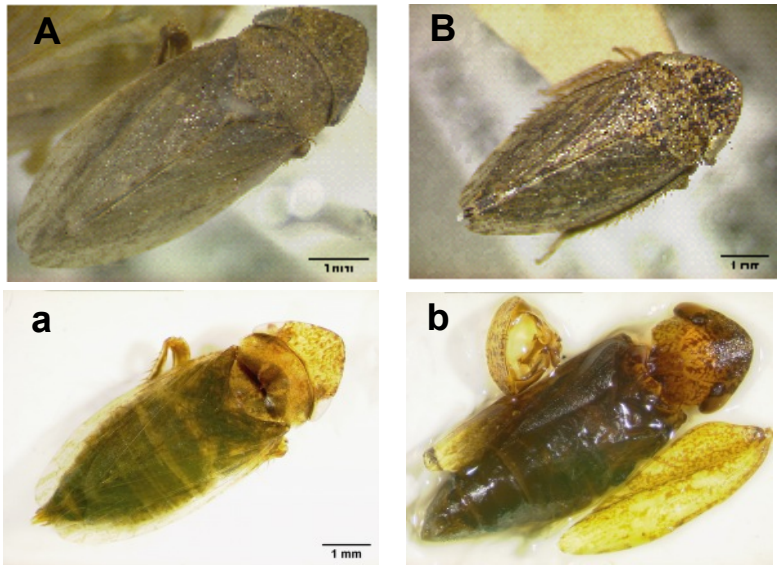


Fig. S1 *Aphrodes* museum specimens before (A, B) and after (a, b) 24 h lysis. (A, a) 25 year old specimen (A: magnification 2.0x, 219mp/mm; a: magnification 1.6x; 177mp/mm). (B, b) 100 year old specimen (magnification 1.6x; 177 mp/mm). Specimens A and B were photographed dry, a and b in 100% ethanol.

```
A. bicincta            ---------- ---C------ ---------- ---------- ---------- ---------- ----------
A. makarovi            ---------- ---------- ---------- ---------- ---------- ---------- ----------
A. aestuarina          -C-----G-- ---------- ---------- ---------- ---------- -----G---- ----------
A. diminuta            ---------- ---G------ C--------- ---------- ---------- --G--A---- ----------
Consensus sequence   1-TTTCACAACC AGGTTCATTT TTGGGGAATG ACCAAATTTA TAATGTAGTT GTTACTTCTC ATGCATTTGT

A. bicincta            ---------- ---------- -G-------- ---------- -----G---- -A-------- ----------
A. makarovi            ---------- ---------- -G-------- ------C--- ---------- -A-------- -C--------
A. aestuarina          ---------- ---------- ---------- ---------- ---------- ---------- ---------G
A. diminuta            ---------- --C------- ---------- ------A--- ---------- ---------- T-----G---
Consensus sequence     AATGATTTTT TTTATAGTTA TACCTATTAT AATTGGTGGG TTTGGAAATT GGCTTGTTCC ATTAATATTA

A. bicincta            -----C---- ---------- ---G------ ---------- -------G-- ---------- -----G----
A. makarovi            ---------- ---------- ---T------ ---------- -------G-- ---------G ----------
A. aestuarina          ---------- ----G----- ---------- ---------- ---------- ------C--- --G-------
A. diminuta            -----C---- ---------- ------C--- ---------- ---------- G--------- -----G----
Consensus sequence     GGTGCTCCTG ATATAGCATT TCCACGAATA AATAATATGA GATTTTGATT ATTGCCTCCA TCATTAATTT

A. bicincta             -G-------- ---------- ---------- ----
A. makarovi             ---------- ---------- ---------- ----
A. aestuarina           -------G-- -------G-- A--------- ----
A. diminuta             ---------- T--------- ---------- -C--
Consensus sequence      TATTATTAAT GAGATCAATT GTTGAAATAG GTTC-244
```

Fig. S2 Sequences of a 244 bp fragment (fragment A Table 1) of the cytochrome oxidase subunit *I* gene for four species in the genus *Aphrodes*. For each species, the most common haplotype is shown. Positions of species-specific and diagnostic nucleotides are shown. Dashes (-) represent identical bases to those of the consensus sequence (244 bp). For *A. diminuta* at position 167 bp, the haplotypes also include T instead of C, however, both are diagnostic at this position. For *A. makarovi* at position 200 bp, haplotypes include T instead of G but both are diagnostic at this position.

Phylogenetic analysis

The 244 bp sequences obtained from museum specimens (GenBank Accession numbers; FR727154-FR727166) were aligned with all haplotypes obtained from fresh material (GenBank Accession numbers; FR727167-FR727179), using SEQUENCHER version 4.9 (Gene Codes). Neighbour-joining (NJ) and maximum likelihood (ML) analyses were used to determine phylogenetic relationships between haplotypes for the four *Aphrodes* species including the museum specimen sequences using PAUP version 4.0 beta (Swofford 2001) and Bayesian Inference (BI) using MRBAYES version 3.1.2 (Ronquist & Huelsenbeck 2003). For two museum specimens where the whole 244 bp fragment was not obtained (1929.20.329 and 1929.20.331, Table 2) the unknown bases were coded as missing data. All trees were rooted using the closely related species *Anoscopus limicola* (Edwards 1908) (GenBank Accession number FR729924).

Pairwise Kimura-2-parameter (K2P) distances (Kimura 1980) between haplotypes were used to carry out NJ analyses with 1000 bootstrap replicates to estimate nodal support. A likelihood ratio test as implemented in JMODELTEST version 0.1.1 (Guindon & Gascuel 2003; Posada 2008) was used to statistically select the best-fit model of nucleotide substitution for the data. The best-fit model for the full 244 bp alignment (fragment A), chosen using the Akaike information criterion corrected for small sample size (AICc), was the TPM1uf+G model (base frequencies of A = 0.3065, C = 0.1280, G = 0.1591, T = 0.4065, gamma distribution shape parameter = 0.1030). For fragment C (140 bp) the HKY+I+G model was selected using AICc (base frequencies of A = 0.2706, C = 0.1180, G = 0.1958, T = 0.4157, transition / transversion ratio = 5.5886, proportion of invariable sites (I) = 0.4940, gamma distribution shape parameter = 0.0530). For fragment D (124 bp) the TPM3 model was selected using AICc (base frequencies = equal, rates = equal). Maximum likelihood analyses were conducted using the heuristic search option with 10 random addition replicates and the tree-bisection-reconnection (TBR) branch-swapping algorithm, with 1000 bootstrap replicates (Felsenstein 1985) to estimate nodal support. The Bayesian analyses were conducted using MRBAYES version 3.1.2 (Ronquist & Huelsenbeck 2003). Four chains were run for 5 x 10^6 generations using random starting trees and flat priors. Trees and parameters were recorded every 100^{th} generation. Two runs were performed simultaneously and split frequencies were compared every 100^{th} generation to ensure convergence of the

runs. Both runs used the default heating and swap parameters. The first 5000 generations were excluded as the burn-in.

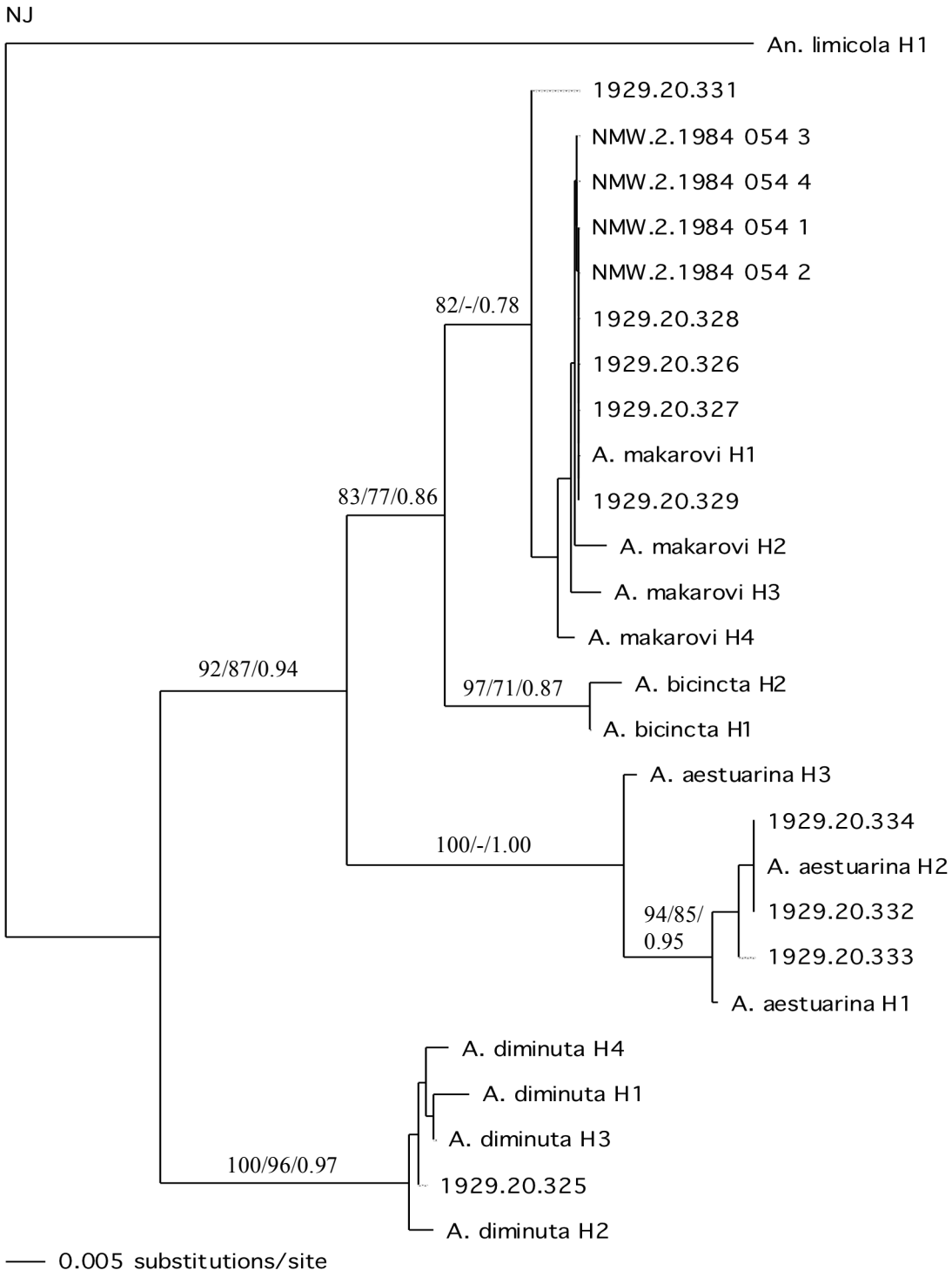


Fig. S3 Neighbour joining phylogeny using Kimura 2-parameter distances for the 244 bp fragment of cytochrome oxidase subunit I gene sequence depicting the relationships between 13 museum specimens and all currently known haplotypes for *Aphrodes*

makarovi, Aphrodes bicincta, Aphrodes aestuarina and *Aphrodes diminuta.* Bootstrap support values greater than 70% are shown above the branches for neighbour-joining, maximum likelihood analyses and posterior probability support values greater than 0.7 are shown for Bayesian analyses, respectively. A dash is presented when a node could not be recovered by one or more of the analyses described. Museum specimens 1929.20.329 and 1929.20.331 are only partial sequences. The phylogram is rooted using *Anoscopus limicola*. The scale bar represents 0.005 substitutions per site. H = haplotype.

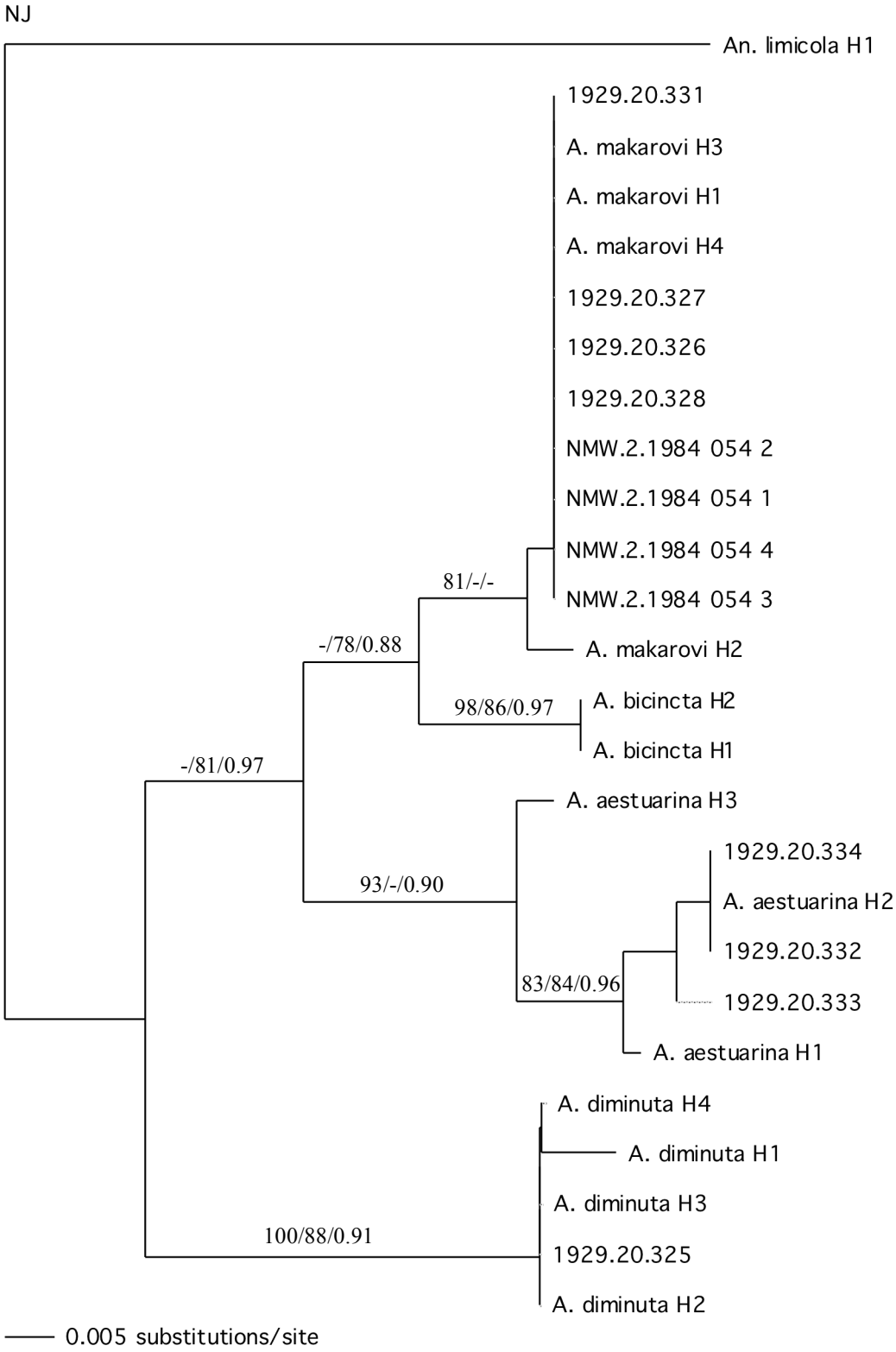


Fig. S4 Neighbour joining phylogeny using Kimura 2-parameter distances for fragment C (140 bp) of cytochrome oxidase subunit I gene sequence depicting the relationship between 12 museum specimens and all currently known haplotypes for *Aphrodes makarovi, Aphrodes bicincta, Aphrodes aestuarina* and *Aphrodes diminuta.* Bootstrap

support values greater than 70% are shown above the branches for neighbour-joining, maximum likelihood analyses and posterior probability support values greater than 0.7 are shown for Bayesian analyses, respectively. A dash is presented when a node could not be recovered by one or more of the analyses described. The phylogram is rooted using *Anoscopus limicola*. The scale bar represents 0.005 substitutions per site. Museum specimen 1929.20.329 did not amplify for this fragment and was omitted from the analysis. H = haplotype.

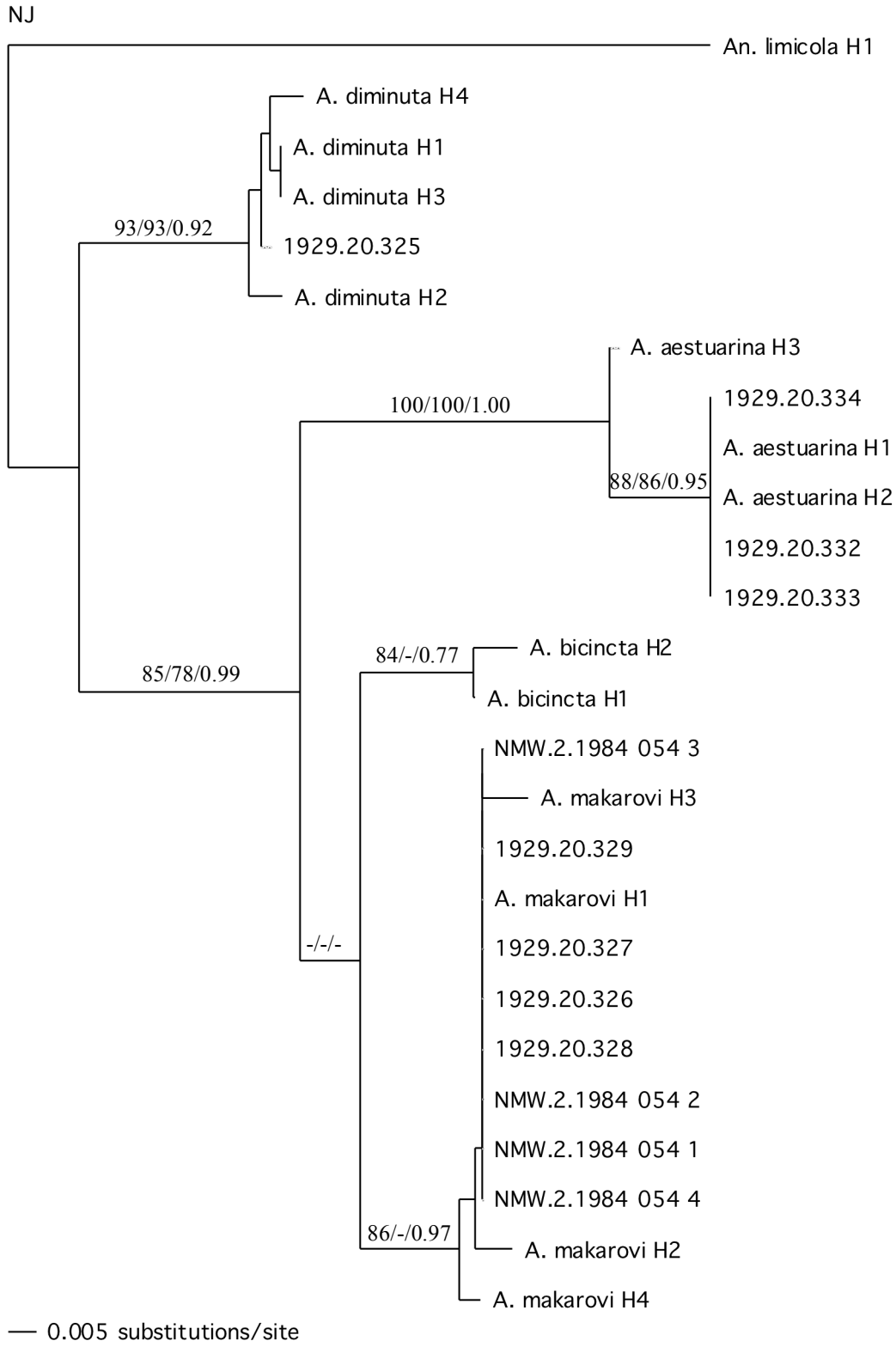


Fig. S5 Neighbour joining phylogeny using Kimura 2-parameter distances for fragment D (124 bp) of cytochrome oxidase subunit I gene sequence obtained from 12 museum specimens and all currently known haplotypes for *Aphrodes makarovi, Aphrodes*

bicincta, Aphrodes aestuarina and *Aphrodes diminuta.* Bootstrap support values greater than 70% are shown above the branches for neighbour-joining, maximum likelihood analyses and posterior probability support values greater than 0.7 are shown for Bayesian analyses, respectively. A dash is presented when a node could not be recovered by one or more of the analyses described. The phylogram is rooted using *Anoscopus limicola*. The scale bar represents 0.005 substitutions per site. Museum specimen 1929.20.331 did not amplify for this fragment and was omitted from the analysis. H = haplotype.

References

Bantock, T. M., Prys-Jones, R. P., & Lee, P. L. M. (2008). New and improved molecular sexing methods for museum bird specimens. *Molecular Ecology Resources*, *8,* 519-528.

Edwards, J. (1908). On some British Homoptera hitherto undescribed and unrecorded. *Entomologist's Monthly Magazine*, *44*, 55-59.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, *39*, 783-791.

Gilbert, M. T. P., Moore, W., Melchior, L., & Worobey, M. (2007). DNA extraction from dry museum beetles without conferring external morphological damage. *PLoS ONE 2(3)*: e272. doi: 10.1371/journal.pone.0000272

Guindon, S., & Gascuel, O. (2003). A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology, 52*, 696-704.

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *15*, 111–120.

Lee, P. L. M., & Prys-Jones, R. P. (2008). Extracting DNA from museum bird eggs, and whole genome amplification of archive DNA. *Molecular Ecology Resources*, *8*, 551-560.

Mazzoni, V., Prešern, J., Lucchi, A., & Virant-Doberlet, M. (2009). Reproductive strategy of the Nearctic leafhopper *Scaphoideus titanus* Ball (Hemiptera: Cicadellidae). *Bulletin of Entomological Research*, *99*, 401-413.

Posada, D. (2008). jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*, *25*, 1253-1256.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics, 19*, 1572-1574.

Swofford, D. L. (2001). PAUP: Phylogenetic Analysis Using Parsimony (*and other methods), version 4.0 b10. Sinauer Associates, Sunderland, Massachusetts.

Tishechkin, D. Y. (1998). Acoustic signals and morphological characters of leafhoppers belonging to *Aphrodes bicinctus* group (Homoptera, Cicadellidae) from European Russia. *Entomological Review*, *78*, 370-377.

Virant-Doberlet, M., & Žežlina, I. (2007). Vibrational communication of *Metcalfa pruinosa* (Hemiptera: Fulgoroidea: Flatidae). *Annals of the Entomological Society of America*, *100*, 73-82.

Wandeler, P., Hoeck, P. E. A., & Keller, L. F. (2007). Back to the future: museum specimens in population genetics. *Trends in Ecology and Evolution*, *22*, 634-642.

Chapter 3:

Differentiation among species within the *Aphrodes* leafhopper genus (Hemiptera, Cicadellidae), comparing morphological, bioacoustic and DNA-based taxonomy.

3.1. Abstract

A multi-disciplinary approach was taken to explore the variation among species within the *Aphrodes* leafhopper genus (Hemiptera, Cicadellidae). This chapter aimed to provide validated fresh material that had been unequivocally identified as belonging to a particular *Aphrodes* species, which was subsequently used for ancient DNA primer design and identification of museum specimens (including the *Aphrodes aestuarina* syntype series) by Bluemel *et al.* (2011 – Chapter 2). This was carried out so that reliable identification of *Aphrodes* could be achieved (particularly those found in estuarine habitats) and to provide a basis to explore further evolutionary questions discussed in the remainder of this thesis. A combination of male vibrational mating signal/female preference and phylogenetic analyses of mitochondrial cytochrome oxidase subunit I gene (COI) sequence data provided good support for the existence of four *Aphrodes* species, that remain distinct in sympatry, despite considerable morphological similarity. This result was congruent across all UK sampling localities, except the Medway estuary where individuals were found that emitted the mating signal of *A. aestuarina* (or females that responded to *A. aestuarina* playback signals) but their mitochondrial DNA sequences matched *A. makarovi*. The dis-concordant pattern identified in this population could be explained by either introgression or retention of an ancient ancestral polymorphism. Distinct habitat, host plant, and morphological differences were identified between *A. makarovi* populations inhabiting inland and estuarine environments. Furthermore, very few mitochondrial DNA haplotype sequences were shared between inland and estuarine *A. makarovi* populations. Host plant shifts or adaptation along environmental gradients may play a role in promoting and maintaining the differentiation identified among inland and estuarine *A. makarovi* populations. Alternatively, random processes (genetic drift) occurring at different geographic localities may be responsible for the pattern reported. Using an array of tools to reveal patterns of behavioural, morphological and molecular diversity among morphologically cryptic species suggests that the *Aphrodes* genus is an ideal model for exploring processes involved in driving and maintaining biological differentiation, above and below the species level.

3.2. Introduction

An array of character types have been used to identify different taxa, but distinct species often show variation in morphological characters with little or no overlap between species (discrete morphological clusters) (Coyne & Orr 2004). The occurrence of morphological distinction in areas of sympatry is often used to infer reproductive isolation among morphologically distinct taxa, but when no consistent morphological differences are identified, these groups are commonly referred to cryptic or sibling species (Mayr 1963; Quicke 1993). The biological or recognition species concept (Mayr 1963; Paterson 1985), however, underlines the importance of reproductive isolation and mate recognition, with no *a priori* assumptions of an association between speciation and morphology. These concepts have become increasingly more important due to the mounting evidence that overlapping morphological variability and cryptic species are a fairly common phenomenon among invertebrates (Henry *et al.* 1999a, 2002; Hebert *et al.* 2004; Pfenninger *et al.* 2006; Price *et al.* 2007). A multidisciplinary approach is therefore advised when identifying species (Sites & Marshall 2004) and has been employed in studies involving morphologically cryptic taxa (Pfenninger *et al.* 2006; Price *et al.* 2007; Towes & Irwin 2008).

Aphrodes leafhoppers belong to the order Hemiptera (Suborder: Auchenorrhyncha, Family: Cicadellidae, Subfamily: Aphrodinae) and have a widespread distribution across the northern hemisphere (Nickel & Remane 2002). The four currently recognised species, *A. makarovi* (Zakhvatkin 1948), *A. bicincta* (Schrank 1776), *A. diminuta* (Ribaut 1952) and *A. aestuarina* (Edwards 1908) are very similar morphologically, making identification problematic (Hamilton 1975; Le Quesne 1988; Tishechkin 1998). There are divergent opinions on the taxonomic classification of the *Aphrodes* genus (Le Quesne & Payne 1981; Hamilton 1983; Kirby 1992; Tishechkin 1998; Nickel 2003). Sexual communication (conspecific mate recognition) in leafhoppers (and all Auchenorrhyncha, except cicadas) is facilitated exclusively by species-specific vibrational mating signals (Claridge 1985; Čokl & Virant-Doberlet 2003). In the *Aphrodes* genus, substrate borne vibrational signals have been shown to discriminate among species (Tishechkin 1998; Virant-Doberlet *et al.* 2005, 2006), as found in other phytophagous insects taxa (Virant-Doberlet & Čokl 2004). Different mating signals are likely to evolve due to adaptations to the signal transmission properties of their

respective hosts (Cocroft & Rodrigues 2005; McNett & Cocroft 2008). Recognition of species based solely on mating signals, however, is problematic for taxonomists because living specimens need to be examined and training to identify song phenotypes and to carry out laboratory based playback experiments to identify females, is often required.

Recent evidence from ancient mitochondrial DNA (mtDNA) sequencing of four short, overlapping fragments of the cytochrome oxidase subunit I gene (COI) from *Aphrodes* museum specimens identified numerous misclassified specimens (Bluemel *et al.* 2011 – Chapter 2), including individuals from the syntype series of *A. aestuarina* (Edwards 1908). This earlier study identified the need to validate the identity of museum specimens by comparison with freshly collected specimens that have been unequivocally identified as members of a particular species. Validated representatives of each *Aphrodes* species were necessary for designing the *Aphrodes* specific mtDNA COI primers that were used by Bluemel *et al.* (2011 – Chapter 2). In this chapter, the analyses undertaken that provided the validated fresh material for this purpose are described. Analyses involved the combined use of morphological, behavioural (vibrational mating signals) and molecular (mtDNA COI sequencing) techniques. A brief history of the similarities and differences reported between *Aphrodes* species is also outlined, based on the current knowledge of characteristics used to distinguish species of this genus.

3.2.1. *Aphrodes* leafhoppers

Aphrodes leafhoppers (Hemiptera: Cicadellidae: Aphrodinae) are univoltine, with egg development occurring overwinter (Nickel & Remane 2002; Nickel 2003). Adult *Aphrodes* males produce species-specific vibrational mating signals (Tishechkin 1998) to which con-specific females respond to, which is thought to be associated with the prevention of inter-specific matings, as a form of reproductive isolation (Virant-Doberlet unpublished data). Due to the morphological similarities among *Aphrodes* species, the male species-specific mating signals have been used for identification, which currently represents the most robust method for distinguishing among *Aphrodes* (Tishechkin 1998; Virant-Doberlet *et al.* 2005, 2006). Females of all four *Aphrodes* species emit similar vibrational mating signals that differ in their duration and in the click repetition time (Virant-Doberlet, unpublished data). However, playback

experiments are technically demanding and time consuming and therefore not practical. In addition, mated females won't respond to male signals (Virant-Doberlet, personal communication).

Previous morphological identification of *Aphrodes* relied on the differences in the positions of the spines on the male aedeagus (Nast 1976; Tishechkin 1998, Chapter 1). The use of male genitalia for identifying species is common in many insect groups (Quicke 1993). Other characteristics used to determine species status of males include the proportions of fore wings and vertex, shape of the penis shaft, body size and the shape of the sternal apodemes of abdominal segment II (Tishechkin 1998). These morphological characteristics are unreliable for classification because variation is continuous and considerable overlap occurs between species (Tishechkin 1998). Additionally, descriptions were only based on small sample sizes. *Aphrodes* females are morphologically cryptic to the extent that no consistent differences have been found between them (Tishechkin 1998; Biedermann & Niedringhaus 2004). A possible reason for no morphological differentiation reported in females, may be that within mixed communities of *Aphrodes,* females may have been attributed incorrectly to a particular species, when identified only by the species identification of males present at the same locality. Thus, no consideration in previous studies was given to the possibility that several species may be present at a single locality.

Ecological differences between species have been observed, such as host plant preference, height found above sea level and habitat type (Kirby 1992; Tishechkin 1998; Nickel 2003; Biedermann & Niedringhaus 2004, Chapter 1), however, ranges of some species overlap and to what degree these preferences are fixed is debatable. *Aphrodes* males also signal on non-host plants (Virant-Doberlet, personal communication) and other studies have reported the apparent polyphagous nature of some species (e.g. *A. makarovi*, Nickel & Remane 2002; Biedermann & Niedringhaus 2004).

While the species status of *A. makarovi, A. bicincta* and *A. diminuta* is well acknowledged by taxonomists (following Tishechkin 1998), classification of the halophilous *A. aestuarina* has been questioned due to a lack of consistent morphological differences between it and *A. makarovi* (Biedermann & Niedringhaus 2004) and mating signal data has been lacking for estuarine populations of *Aphrodes.*

Aphrodes aestuarina was classified as a distinct species due to its specific niche and slight differences in size (longer and thinner) and colouration (Edwards 1908), but with much caution as it is also documented as a possible synonym of *A. makarovi* (Kirby 1992; Nickel 2003). *Aphrodes aestuarina* is thought to be endemic to the UK (Tishechkin 1998) but specimens thought to be this species have been reported in coastal regions of the North and Baltic Sea, including Germany and Poland (Kirby 1992; Biedermann & Niedringhaus 2004). *Aphrodes aestuarina* is described as being habitat specific, occurring only in saline meadows and coastal saltmarshes (Nickel 2003), adapted to survive daily inundation on Shrubby Seablite, *Suaeda vera* (Amaranthaceae, Subfamily Suaedoideae; previously known as *S. fruticosa*), and occasionally on annual seablite, *S. maritima* (Edwards 1908; Kirby 1992). Sampling data from Virant-Doberlet (personal communication) suggests that *A. aestuarina* was collected from locations where little or no *S. vera* was present but mainly Sea Purslane, *Atriplex portulacoides* (Amaranthaceae, Subfamily Chenopodioideae; previously known as *Halimione portulacoides*, see also Kirby 1992). So the degree of host plant specificity is questionable.

Recent evidence from ancient mtDNA sequencing of 244 base pair (bp) COI sequences from c. 100 year old museum specimens representing the syntype series for *A. aestuarina* (found in Wells, Norfolk; Edwards 1908) identified three (out of a total of seven analysed) misclassified specimens, that matched sequences of *A. makarovi* (Bluemel *et al.* 2011 – Chapter 2). This showed that *A. makarovi* was historically found on saltmarshes in Norfolk in sympatry with *A. aestuarina.* Based on the 244 bp fragment sequenced, considerable mtDNA sequence divergence between *A. aestuarina* and *A. makarovi* exists (forming well supported clades in phylogenetic analysis, Bluemel *et al.* 2011, 2.1 Supporting Information – Chapter 2), which continue to remain distinct when they are found in sympatry (at the syntype location), it is clear that *A. aestuarina* is likely to represent a distinct species. Whether there are any morphological differences between *A. makarovi* and *A. aestuarina* remains unclear, as previous studies could have included a mixed species sample (*A. makarovi* and *A. aestuarina*) when they thought they were sampling just *A. aestuarina*. Further analysis of estuarine populations of *Aphrodes* is needed to examine this.

Based on the taxonomic uncertainties outlined within the *Aphrodes* genus (and other related genera) and the numerous synonyms identified in the past (the extent of which

has only briefly been discussed), there is a clear need for a robust and multidisciplinary approach for species identification.

3.2.2. Aims and hypotheses

The aim of this study was to obtain the validated information about the species identity of freshly collected *Aphrodes* using a multi-disciplinary approach, incorporating morphology, behaviour (mating signals) and molecular (mtDNA COI sequencing) analyses. The hypothesis that the previously defined *Aphrodes* species do represent distinct species is tested. Specimens representing each species that are analysed using all of the methods described should provide a reliable standard upon which further investigations of *Aphrodes* species can be based.

Additionally, this study aimed to identify the mtDNA haplotype diversity, phylogenetic relationships, vibrational signals and morphological differences between *Aphrodes* species with particular attention to saltmarsh sites where both estuarine species (*A. makarovi* and *A. aestuarina*) may be found in sympatry (as reported in Chapter 2 – Bluemel *et al.* 2011). Aims to determine their current distributions around the coast of the UK and to establish whether they are frequently found in sympatry or whether populations of single species are more common, were undertaken. Based on the current knowledge of differentiation among *Aphrodes* species, it is hypothesised that estuarine *A. aestuarina* and *A. makarovi* will remain distinct in sympatry when analysed for mtDNA sequence variation along with male vibrational mating signals, which provide good evidence for their species status, but these putative species are unlikely to express significant differences in morphology.

3.3. Materials and Methods

3.3.1. Sample collection

Adult specimens of the four *Aphrodes* species, *A. makarovi*, *A. bicincta*, *A. diminuta* and *A. aestuarina* were collected from 12 locations across the British Isles (n= 403, Table 3.1 and Fig. 3.1). Specimens were collected from June-August during the years 2005-2009 using a converted leaf blower (D-vac suction sampler, Electrolux, BVM 250). Sampling sites around the coast were chosen in part based on the presence of suitable saltmarsh habitats with relevant host plants (Chapter 1). Specimens were stored in absolute ethanol at -80°C either directly from the field or after vibrational signals had been recorded. See 3.8, Appendix I for a list of sampling sites and geographic coordinates (GPS).

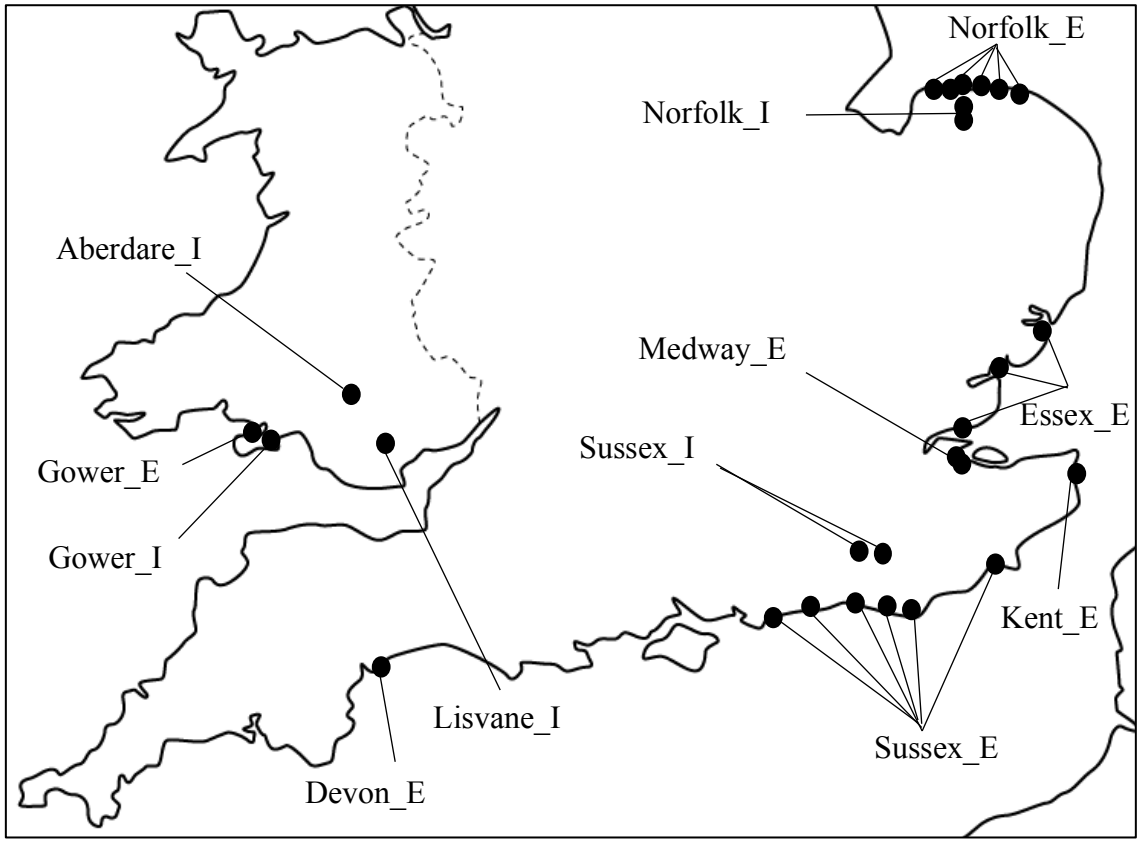


Figure 3.1. Map of the UK illustrating the sampling locations used in this study. E = estuarine/saltmarsh sites and I = inland sites. See 3.8, Appendix I for geographic coordinates (GPS) for each sampling locality.

Table 3.1. Information for each of the 12 sampled populations, including host plant and habitat type, location name, number of validated male and female (M/F) specimens from each location recorded for mating signal, sequenced for mitochondrial cytochrome oxidase subunit I (COI) gene and for morphological analysis (whole body morphology and/or male aedeagus morphology). Additional specimens were included from each locality and were included in at least one type of analysis. Any of these additional specimens included in morphological analyses had been included in at least one other analysis (molecular or bioacoustic). I=inland locations, E=estuarine locations.

Host plant / Habitat type	**Location**	**Validated (M/F)**	**Additional (M/F)**
Atriplex - saltmarsh	Kent_E	/	10 (8/2)
Atriplex - saltmarsh	Essex_E	/	26 (15/11)
Atriplex - saltmarsh	Gower_E*	5 (5/0)	15 (6/9)
Atriplex - saltmarsh	Norfolk_E	22 (22/0)	39 (29/10)
Atriplex - saltmarsh	Medway_E*	24 (17/7)	76 (43/33)
Atriplex - saltmarsh	Devon_E	/	10 (8/2)
Atriplex - saltmarsh	Sussex_E*	10 (7/3)	47 (9/38)
Urtica - grassland	Gower_I	1 (1/0)	19 (7/12)
Urtica - grassland	Norfolk_I	3 (3/0)	17 (8/9)
Urtica - grassland	Lisvane_I*	7 (7/0)	14 (9/5)
Urtica + Fabaceae – Grassland/heathland	Sussex_I*	12 (8/4)	16 (14/2)
Fabaceae - brown field site	Aberdare_I*	8 (5/3)	23 (17/6)
Total	**12**	**92 (75/17)**	**312 (173/139)**

* = locations where signal recordings and sample collections were mainly carried out by Virant-Doberlet *et al.* (2005-2007).

See 3.8, Appendix I for geographic coordinates (GPS) for each sampling locality.

A subset of 92 specimens (75 males and 17 females) was chosen for species validation (Table 3.1), which was included in all three analyses (morphological, bioacoustic and mitochondrial COI sequencing). Initially, individuals were determined by recording their vibrational mating signal. Legs were subsequently removed and DNA was extracted for molecular analyses. The rest of the body was stored in absolute ethanol at -80°C and used for morphological analyses.

Specimens collected in the year 2005 were dissected completely, so that they could not be used in morphological analyses of the whole body. Thus not all of the 92 validated specimens were included in both whole body and aedeagus analyses. Methods described for mtDNA COI primer design for ancient analysis of museum specimens (Bluemel *et*

al. 2011 – Chapter 2) were based upon this validated subset of 92 specimens, which were unequivocally identified as being a member of a particular species. An additional 311 specimens (173 males and 138 females) were analysed using at least one analysis. Any of these additional specimens that were subsequently included in morphological analyses had been analysed using either molecular or bioacoustic methods for species identification.

3.3.2. Vibrational signal recordings

Male vibrational mating signals were recorded using a laser vibrometer (PDV 100, Polytec GmbH, Waldbronn, Germany). Signals were digitised with 48 kHz sample rate and 16-bit depth and stored directly onto a hard drive of a notebook computer using a Sound Blaster Audigy 2 ZS sound card (Creative Labs Inc.) and the RAVEN version 1.2.1 software program (Charif et al., 2004) (methods described in Virant-Doberlet & Žežlina 2007; Mazzoni *et al.* 2009). At a number of sampling localities (Table 3.1) spontaneously emitted male vibrational signals were recorded, while females were identified by their responsiveness in playback experiments by Virant-Doberlet (unpublished data). In the following years not all individuals could be recorded due to equipment availability. A total of 150 vibrational mating signals were recorded (125 males and 25 females, including the 92 specimens used for species validation). Male calling songs were compared to those reported in Tishechkin (1998). The mating signal of *A. aestuarina* had never been recorded previously. Identification of *A. aestuarina* was based on specimens from the site where syntype series had been collected (Wells, Norfolk), the species description of Edwards (1908), specific ecology (daily inundation during high tide) and vibrational signals that differ from the ones previously described for the other three species (see also mtDNA COI sequencing results for the *A. aestuarina* syntype series in Bluemel *et al.* 2011 – Chapter 2).

3.3.3. Adult morphology

Whole specimens were photographed using AUTO-MONTAGE PRO version 5.0 (Synoptics) imaging software and a JVC KY F70 3CCD digital camera. This was mounted on a Leica M28 stereo microscope (1.6x or 2.0x magnification depending on size of specimen) with a Planapo chromatic 1x lens attached. Only specimens with high morphological integrity were measured (i.e. those decapitated during leg dissection were excluded). A total of 223 specimens were analysed (130 males and 93 females).

The genitalia of a subset of 57 male specimens were examined. The abdomen was dissected and macerated by heating in 10% potassium hydroxide solution until the soft tissue had dissolved. This was then dissected further to retrieve the aedeagus and stored in glycerol. Aedeagus images were obtained as described above, 5x magnification was used with a 1.5x Apo chromatic lens attached. Measurements were obtained using the measuring tools in IMAGEJ version 1.40e (Rasband 2007). Measurements taken for whole body analyses are shown in Fig. 3.2 (a), and measurements for the front and side view of the male aedeagus are outlined in Fig. 3.2 (b) and (c), respectively. The corresponding ratios taken and analysed are shown in Table 3.2. Ratios are more likely to provide more information than absolute values because they are independent of variability in the overall size of insects (Quicke 1993). Ratios were chosen to represent the general shape and size of the whole body, thorax and head, male aedeagus and also the distance between the spines on the male aedeagus with resect to other measurements of length.

Principle Components Analysis (PCA) was performed in MINITAB version 16 using a standardised correlation matrix. Whole body male and female data sets were analysed separately due to size differences between males and females of each species (sexual size dimorphism where females are typically larger than males). Male aedeagus and whole body data sets differed in the total number of individuals due to reasons described above (3.3.1).

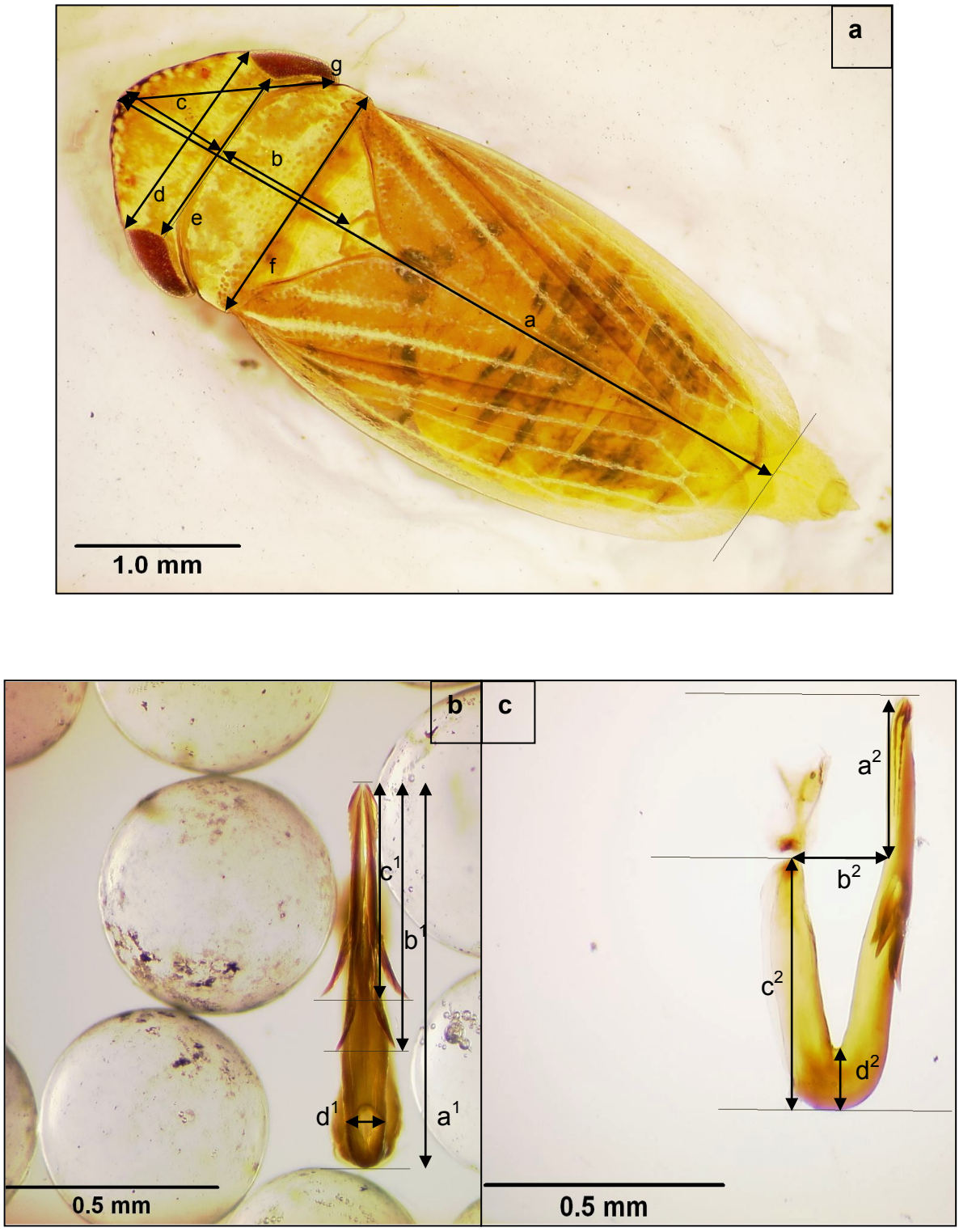


Figure 3.2. Measurements taken for a) whole male and female specimens, b) male aedeagus – front view, c) male aedeagus – side view. Whole body images were taken with specimens mounted in a trough of white-tack submersed in absolute ethanol in an excavated glass block. Aedeagus images were taken whilst mounted on a cavity slide filled with glycerol and small glass beads (see image b) that were used to aid orientation of the aedeagus. Scale bars are shown for each image in mm. The measurements taken (small case letters) correspond to those in Table 3.2 for each image type.

Table 3.2. The measurements and ratios used for analysis of *Aphrodes* species, showing whole body, male aedeagus – front view and male aedeagus – side view measurements and ratios. The letters in small case correspond to those in Fig. 3.2 (a-c) for each image type.

WHOLE BODY	**AEDEAGUS**	
Measurements:	**Measurements:**	
	Aedeagus front view[1]	**Aedeagus side view**[2]
a = length (head to wing) (L)	a^1 = length (AL)	a^2 = length – hook (L-H)
b = head to thorax (H2T)	b^1 = lower spine (LS)	b^2 = hook-tip to shaft (H-S)
c = head length (HL)	c^1 = upper spine (US)	c^2 = hook length (HKL)
d = head width (HW)	d^1 = base width (BW)	d^2 = base height (BH)
e = eye width (EW)	z = distance between spines ($b^1 – c^1$) (Z)	
f = thorax width (TW)		
g = head to eye (H2E)		
Ratios:	**Ratios:**	**Ratios:**
L / TW	AL / H-S	H-S / Z
TW / H2T	AL / HKL	HKL / H-S
TW / HL	AL / BW	HKL / Z
HW / HL	AL / BH	BW / BH
EW / HL	AL / Z	L-H / H-S
H2E / HL	LS / US	H-S / BH
	Z / BW	

3.3.4. DNA extraction, Polymerase Chain Reaction (PCR) amplification and sequencing

Legs of 363 *Aphrodes* specimens (216 males and 147 females, including 92 specimens used for species validation) were dissected and genomic DNA was extracted using a DNeasy Tissue Kit (Qiagen). Universal invertebrate primers LCO1490 (5`-GGTCAACAAATCATAAAGATATTGG-3`) and HCO2198 (5`-TAAACTTCAGGGTGACCAAA-AAATCA-3`) (Folmer *et al.* 1994) were used to amplify a 710 bp fragment (including primer sequence) of the cytochrome oxidase subunit I (COI) gene. Polymerase chain reactions were carried out in total volumes of 12.5µl consisting of 1x PCR buffer (Invitrogen), 4mM $MgCl_2$ (Invitrogen), 0.1mM dNTP (New England Biolabs, NEB), 0.1µg/µl bovine serum albumin (NEB), 0.1µM of each forward and reverse primer, 0.3125 U Taq polymerase (Invitrogen) and 1µl DNA extract, made up to the final volume with ddH_20. Amplification initiated with a denaturing step at 94°C for 2 min 30 s, followed by 35 cycles at 94°C for 30 s, 52°C for 30 s, 72°C for 40 s and a final extension at 72°C for 10 min.

The PCR products were purified and reactions were carried out in a total volume of 10µl consisting of 4U EXO (Exonuclease I, NEB), 1U AP (Antarctic Phosphatase, NEB), 0.2x AP Buffer (Antarctic Phosphatase Buffer, NEB), 6µl of PCR product and made up to the final volume with ddH_20 and incubated at 37°C for 45 min followed by 80°C for a further 15 min. Dye-terminator cycle sequencing was performed in a reaction volume of 10µl, for both the light and heavy strands using the same primers as for the PCR reaction (0.16µM) the ABI PRISM Big DyeTM Terminator, version 3.1 Cycle Sequencing Ready Reaction Kit (Applied Biosystems) following the manufactures protocol. Cycle sequencing was carried out with 25 cycles of 94°C for 10 s, followed by 50°C for 5 s, and 60°C for 2 min. Cycle sequencing products were subsequently precipitated using isopropanol (2-propanol) (Applied Biosystems). Nucleotide sequences were determined using an ABI 3130xl automated sequencer (Applied Biosystems), carried out by the Cardiff University Molecular Biology Support Unit.

3.3.5. Sequence alignment, diversity and phylogenetic analyses

The 710 bp mtDNA raw sequences obtained from *Aphrodes* specimens were checked for quality and aligned using SEQUENCHER version 4.9 (Gene Codes). A consensus sequence for each individual from forward and reverse sequences was determined. Unique haplotype sequences were identified and deposited in GenBank under the Accession numbers; FR727167–FR727179, HE587025–HE587045. Descriptive statistics and diversity indices such as haplotype diversity (Nei 1987) and nucleotide diversity, π (Tajima 1983; Nei 1987) were calculated for the whole sample and each species separately excluding missing data using DNASP, version 4.10.9 (Rozas et al., 2003) and ARLEQUIN version 3.5.1.2 (Excoffier & Lischer 2010). Neighbour-joining (NJ) and Maximum Likelihood (ML) analyses were used to determine phylogenetic relationships between haplotypes for the four *Aphrodes* species using PAUP version 4.0 beta (Swofford 2002) and Bayesian Inference (BI) using MRBAYES version 3.1.2 (Ronquist & Huelsenbeck 2003). All trees were rooted using the closely related estuarine species *Anoscopus limicola* (Edwards 1908) (Hemiptera: Cicadellidae: Aphrodinae) (GeneBank Accession number FR729924 and HE587046).

Pairwise Kimura-2-parameter (K2P) distances (Kimura 1980) between haplotypes were used to carry out NJ analyses with 1000 bootstrap replicates to estimate nodal support. A likelihood ratio test as implemented in JMODELTEST version 0.1.1 (Guindon & Gascuel 2003; Posada 2008) was used to statistically select the best-fit model of nucleotide substitution for the data. Maximum Likelihood analyses were conducted using the heuristic search option with 10 random addition replicates and the tree-bisection-reconnection (TBR) branch-swapping algorithm, with 1000 bootstrap replicates (Felsenstein 1985) to estimate nodal support. The Bayesian analyses were conducted using MRBAYES. Four chains were run for 5×10^6 generations using random starting trees and flat priors. Trees and parameters were recorded every 100th generation. Two runs were performed simultaneously and split frequencies were compared every 100th generation to ensure convergence of the runs. Both runs used the default heating and swap parameters. The first 5000 generations were excluded as the burn-in. Percentage sequence divergence estimates within and among species was also calculated based on pairwise Kimura-2-parameter (K2P) distances calculated in PAUP.

3.4. Results

3.4.1. Vibrational signals

All vibrational signals recorded for *A. makarovi, A. bicincta* and *A. diminuta* matched those previously recorded by Tishechkin (1998). For a map of signal types recorded at each sampling location see Fig. 3.3. The vibrational mating song patterns recorded from each *Aphrodes* species are shown in Fig. 3.4. Signal recordings were not obtained from all locations in Fig. 3.1. The signal of *A. diminuta* was recorded from two inland locations in Sussex_I, which is the first record of this species in the UK.

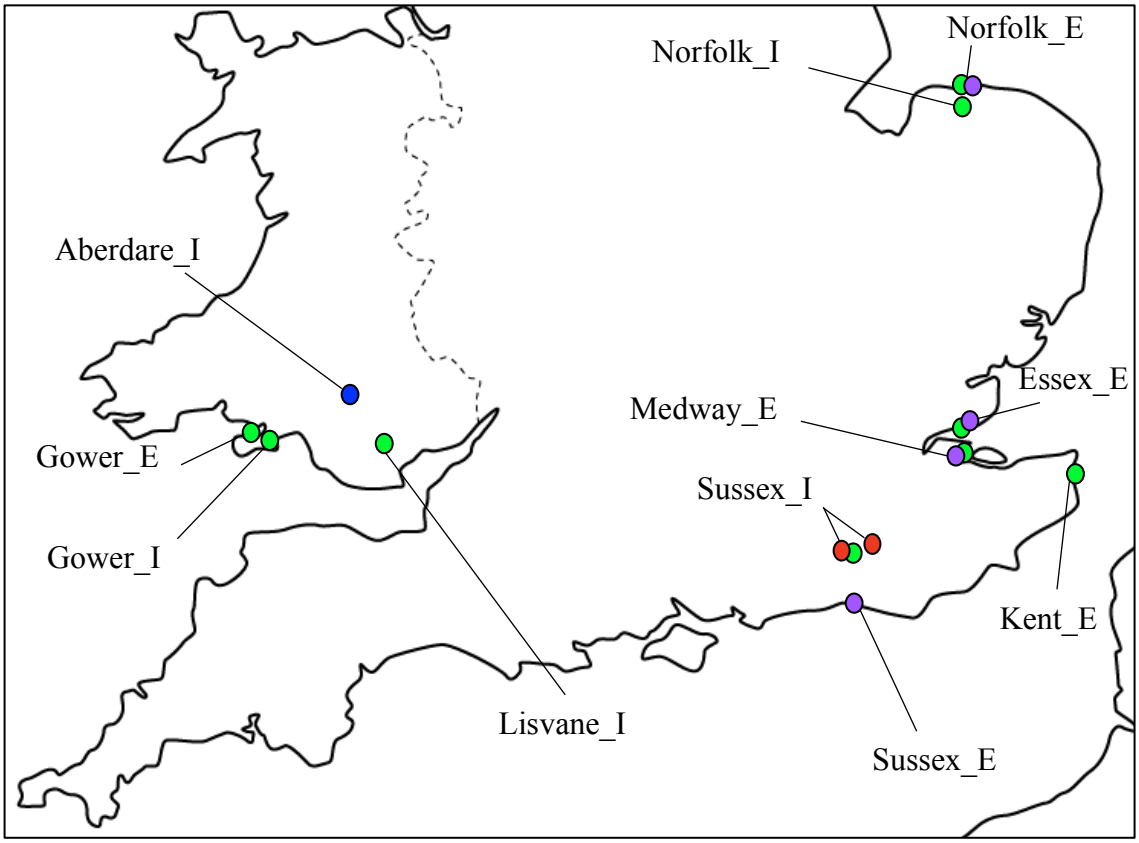


Figure 3.3. Vibrational mating signal results for *Aphrodes* populations sampled in the UK. Green circles indicate where *Aphrodes makarovi* signal was recorded and purple circles represent *A. aestuarina* signals, blue circle = *A. bicincta* signals and red circles = *A. diminuta* signals. Marginally overlapping circles indicate where more than one species mating signal was recorded.

Identification of *A. aestuarina* male mating signal was carried out by analysing individuals collected from the original syntype location (Norfolk, Wells) and by identifying the signal that differed from other *Aphrodes* species. This was confirmed by

comparing mtDNA sequences to those found for ancient museum syntype specimens of *A. aestuarina* (Bluemel *et al.* 2011 – Chapter 2). At this location, both *A. makarovi* and *A. aestuarina* signals (Fig. 3.3) were recorded from freshly collected specimens correlating with results found for the ancient syntype specimens that showed a mixture of both *A. makarovi* and *A. aestuarina* species based on a small fragment (244 bp) of the mtDNA COI gene.

Two other estuarine populations were identified where both *A. makarovi* and *A. aestuarina* vibrational signals were recorded in sympatry, including Essex_E and Medway_E locations. Only *A. aestuarina* mating signals were recorded from the Sussex_E population. *Aphrodes makarovi* mating signals were recorded at most estuarine locations and also from inland sites where one of their primary host plants occurs (*Urtica* sp.).

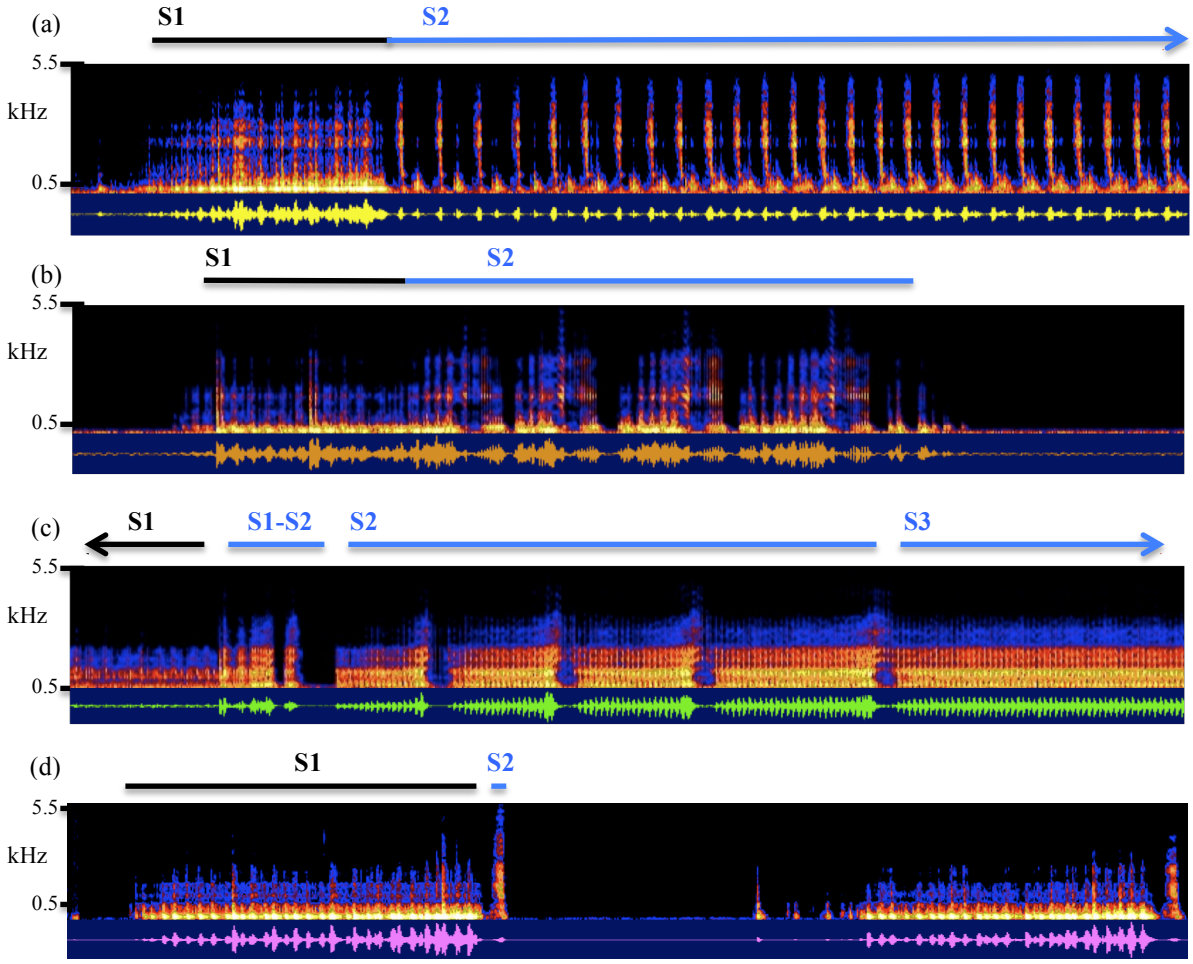


Figure 3.4. Vibrational mating signal acoustic patterns and sonograms illustrating the composition of typical male calling songs of *Aphrodes* species. (a) = *A. bicincta*, (b) = *A. diminuta*, (c) = *A. makarovi*, (d) = *A. aestuarina*. Modified from Virant-Doberlet *et al.* (2005).

Aphrodes mating signals are composed of several sections that differ in structure (Fig. 3.4). Section 1 (S1) is present in all three previously recorded species (Fig. 3.4a-c) and contains no species information, while other sections are highly specific and are composed of species-specific song elements (Virant-Doberlet *et al.* 2005). The song pattern of *A. aestuarina* is composed of elements also present in typical *A. bicincta* males found in Aberdare (Virant-Doberlet *et al.* 2005) (Fig. 3.4d).

3.4.2. Adult morphology

An example of male and female specimens of each *Aphrodes* species can be seen in Fig. 3.5 (except female *A. diminuta*). Results of PCA analysis using ratios obtained from male aedeagus and male and female whole body images indicate a considerable overlap in morphological characters between *Aphrodes* species (Fig. 3.6 - 3.8). *Aphrodes diminuta* is the most distinct in terms of male aedeagus (Fig. 3.6), although sample sizes are low for this species and no female representatives could be included in whole body morphology. A number of individuals collected from the Medway estuary locations are illustrated separately (Fig. 3.6 - 3.8) due to a dis-concordance found between mating signal and mtDNA analyses for a number of these specimens (section 3.4.3). Individuals collected from the Medway_E population that were found to match *A. makarovi* for mating signal and/or mtDNA analyses and are grouped with *A. makarovi* (Fig. 3.6 - 3.8).

There is some degree of clustering of the aedeagus form of the four *Aphrodes* species but with substantial overlap (Fig. 3.6). Principle component 1 (PC1) was correlated with ratios relating to the aedeagus length in comparison to the distance between the spines and explains 56.7% of the overall variation. *Aphrodes makarovi* inland and estuarine populations, *A. aestuarina* and the mismatched Medway estuary specimens show typically larger aedeagus length compared to the distance between the spines, whereas *A. diminuta* shows the opposite, with larger distance between the spines compared with overall aedeagus length. *Aphrodes bicincta* is intermediate between these groups. PC2 is related with ratios involving the angle measurement b^2 (Fig. 3.2b) and explains 30.4% of the variation in the data. Estuarine *A. makarovi* and some Medway estuary and *A. diminuta* specimens show the largest angle between the base of the aedeagus and the

shaft in comparison to aedeagus length measurements (negative end of PC2, Fig. 3.6), although there is much overlap along this PC axis.

There is considerable overlap for all species for whole body male and female morphology (Fig. 3.7 and Fig. 3.8, respectively). The largest sample sizes analysed were for *A. makarovi* and *A. aestuarina* whole body morphology. Results suggest some clustering of the two species but with considerable overlap in all morphological analyses. For males, *A. bicincta, A. diminuta* and estuarine *A. makarovi* overlap considerably (Fig. 3.7). Inland *A. makarovi* males show some clustering towards the positive end of PC1 but with much overlap with other species. PC1 explains 61.8% of the variation and is associated with body width in comparison to aspects of head length. These inland *A. makarovi* specimens have the largest head/thorax width in comparison to head length, suggesting specimens are wider with more rounded heads. No clustering of any species was seen along PC2 (20.9% variation), however, PC3 explained 12.6% of the variation and was associated with the ratio of overall body length to width (Fig. 3.7). *Aphrodes aestuarina* and some Medway estuary specimens cluster towards the negative end of PC3, correlating with longer and thinner body shape in comparison to other species, although considerable overlap is seen.

For females, PC1 and PC2 (explaining 58.5% and 17.5% of variation, respectively) did not separate any *Aphrodes* species. The third PC3 explained 15.9% of the variation in the data and was associated with overall length to width ratio. Inland *A. makarovi* cluster toward the negative end of PC3, correlating with shorter and wider body shape compared with other species although some overlap with *A. bicincta* and estuarine *A. makarovi* specimens was seen (Fig. 3.8). Results suggest that estuarine *A. makarovi* is more similar to *A. aestuarina* (longer and thinner) than inland *A. makarovi* (shorter and wider), with respect to whole body morphology, especially in females. It was also observed during analyses that the body pigmentation of estuarine *A. makarovi* is more similar to that of *A. aestuarina* (Fig. 3.8) showing a lesser degree of banding and dark pigmentation compared with *A. makarovi* from inland sites (personal observation), a characteristic previously used to identify *A. aestuarina* prior to this study (section 3.2.1). Inland *A. makarovi* populations show a much more varied array of morphs, typically darker with higher numbers of banded or spotted morphs compared with the more uniform lighter sandy coloured morph identified in both species occurring in estuarine habitats (Fig. 3.5, see also Chapter 4).

Morphology alone is confirmed to be unsuitable for *Aphrodes* identification purposes although sample sizes are low for aedeagus morphology and for *A. diminuta* and *A. bicincta* whole body morphology. The majority of male whole body and aedeagus data sets could not be analysed together as most aedeagus measurements were taken from individuals that were dissected and therefore could not be included in whole body analyses. Sample sizes for this combined analysis were roughly five individuals per species (with no estuarine *A. makarovi* representatives), and tests using PCA on this data set provided no additional separation of *Aphrodes* species to that presented here and therefore these results are not reported.

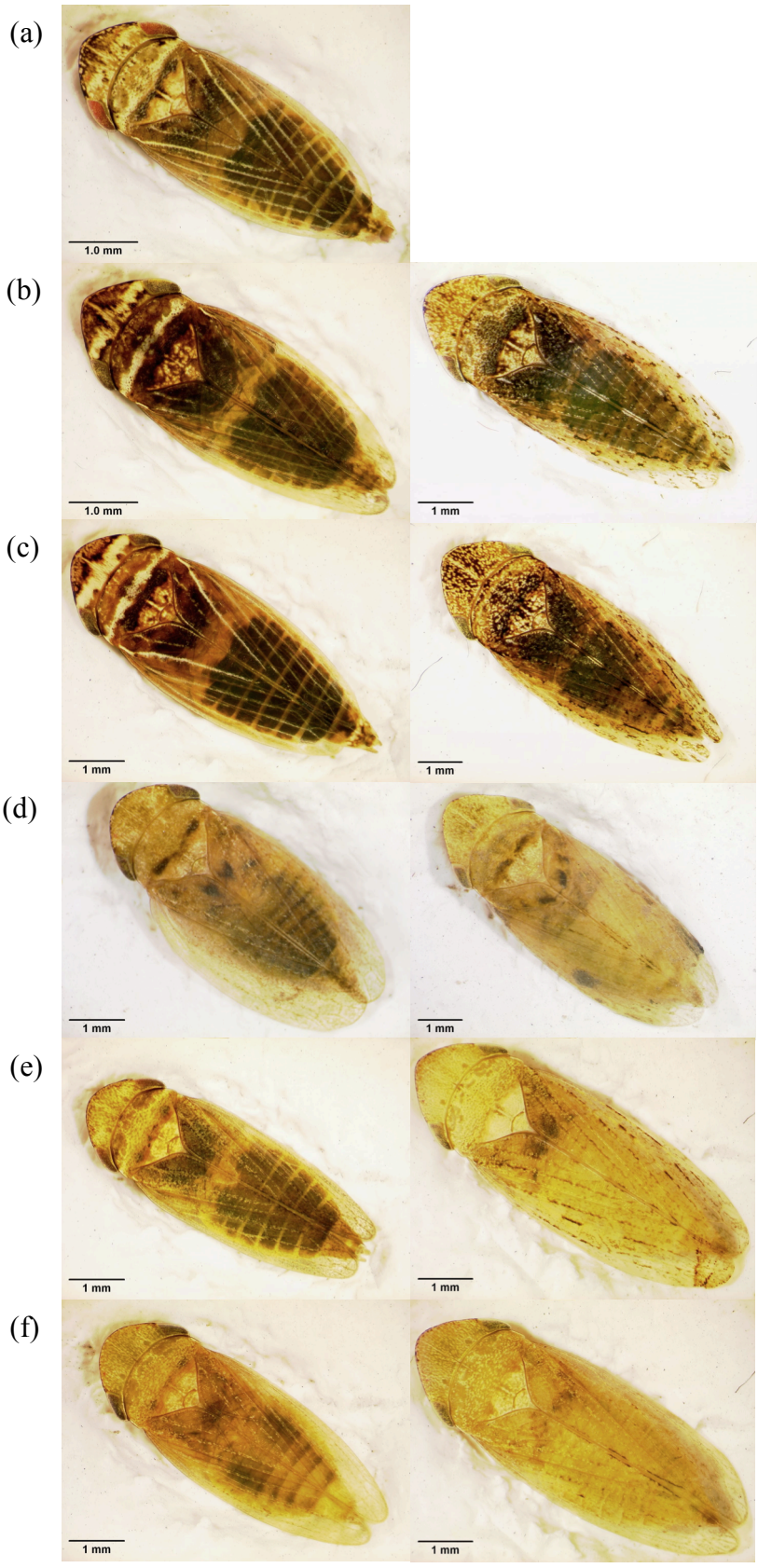


Figure 3.5. Examples of images obtained for each *Aphrodes* species from which whole body morphology measurements were taken. a) *A. diminuta* male, b) *A. bicincta* male and female, respectively, c) inland *A. makarovi* male and female, d) estuarine *A. makarovi* male and female, e) *A. aestuarina* male and female f) Medway estuary male and female specimens, showing a mismatch between vibrational mating signal and mtDNA sequence. Scale bar represents 1mm, images were taken at 1.6x or 2x magnification depending on the size of the specimen.

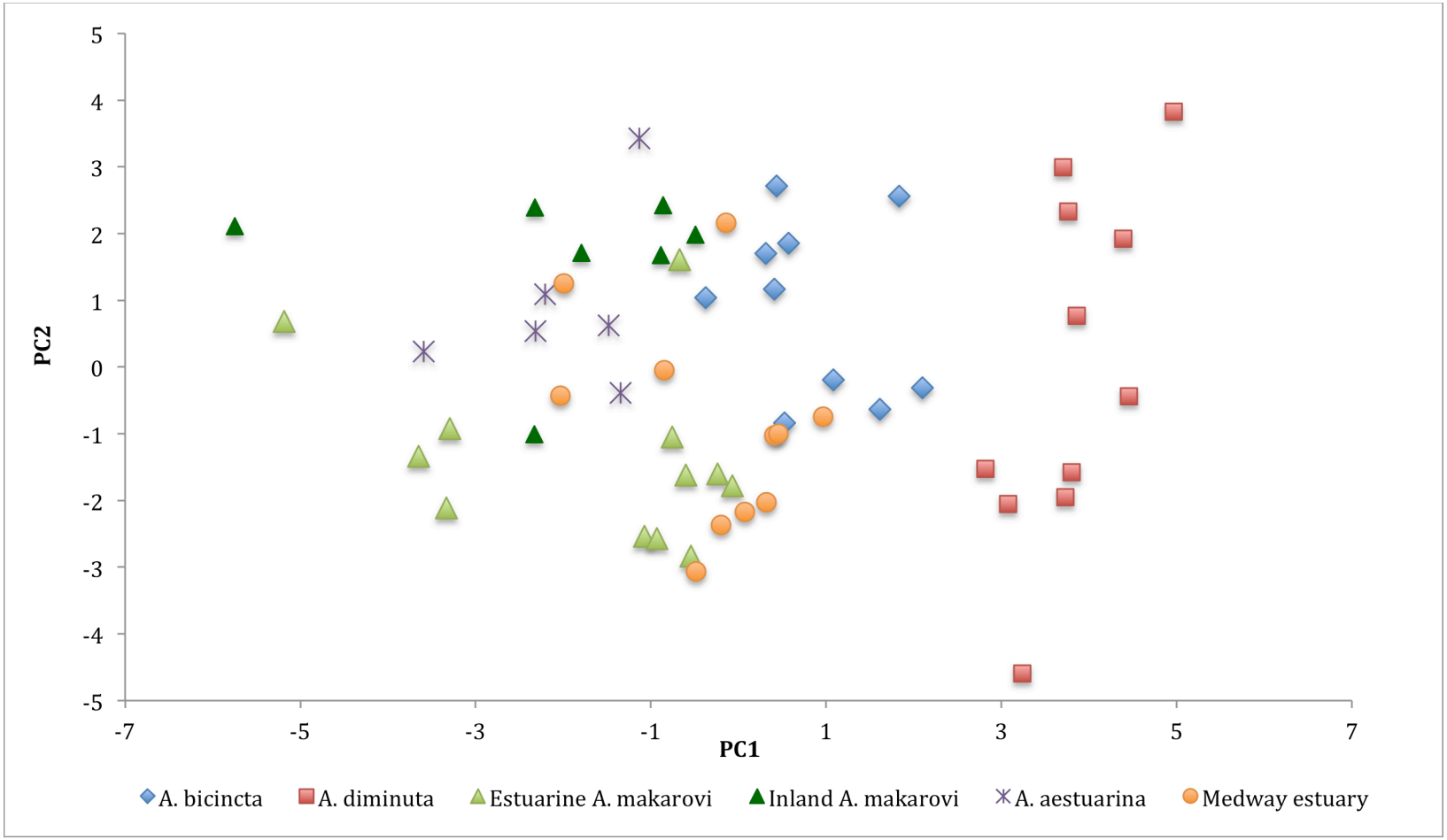


Figure 3.6. Principle component analysis using a standardised correlation matrix of ratios taken for aedeagus morphology for 57 male *Aphrodes* specimens. Principle component (PC) 1 versus PC2 is shown explaining 56.7% and 30.4% variation, respectively. Blue = *Aphrodes bicincta,* red = *A. diminuta,* yellow and green = inland and estuarine *A. makarovi,* respectively and purple = *A. aestuarina.* A number of specimens collected from the Medway estuary (orange) are shown separately due to a dis-concordance found between mating signal and mitochondrial DNA sequences in these specimens (section 3.5.3 and Chapter 5).

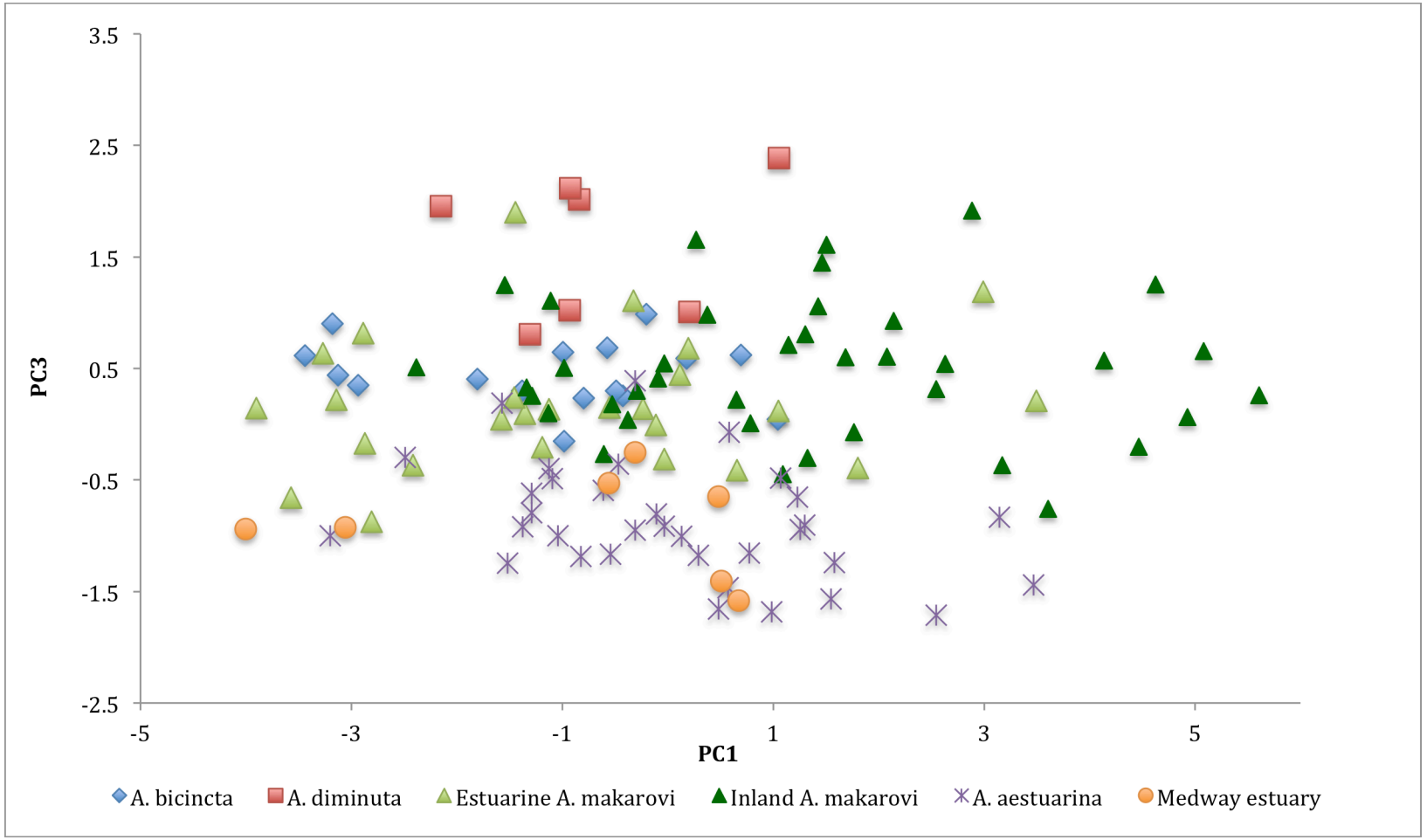


Figure 3.7. Principle component analysis using a standardised correlation matrix of ratios taken for whole body morphology for 129 male *Aphrodes* specimens. Principle component (PC) 1 versus PC3 is shown explaining 61.8% and 12.6% of the variation, respectively. Blue = *Aphrodes bicincta,* red = *A. diminuta,* yellow and green = inland and estuarine *A. makarovi,* respectively and purple = *A. aestuarina.* A number of specimens collected from the Medway estuary (orange) are shown separately due to a dis-concordance found between mating signal and mitochondrial DNA sequences in these specimens (section 3.5.3 and Chapter 5).

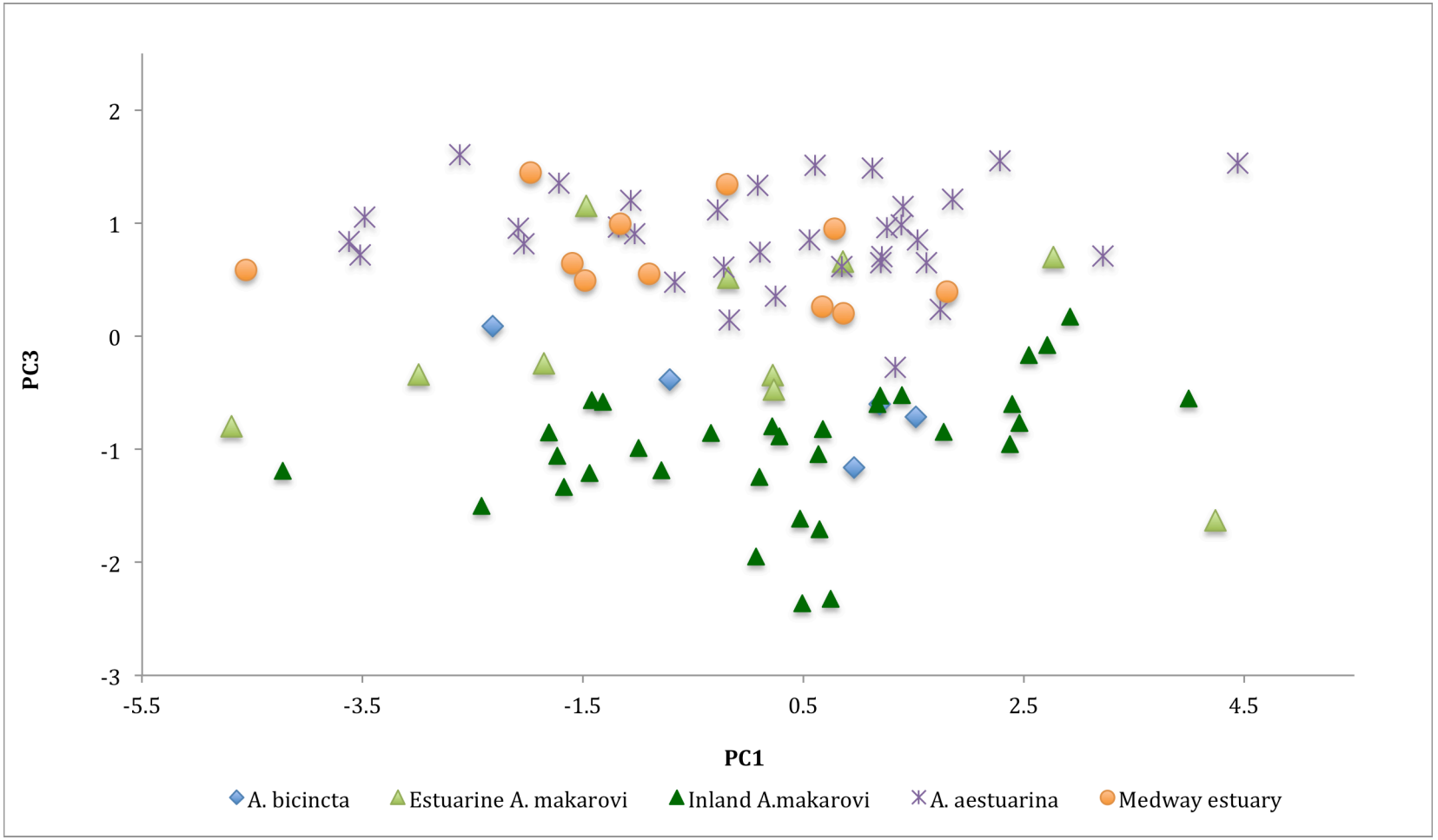


Figure 3.8. Principle component analysis using a standardised correlation matrix of ratios taken for whole body morphology for 93 female *Aphrodes* specimens. Principle component (PC) 1 versus PC3 is shown explaining 58.5% and 15.9% of the variation, respectively. Blue = *Aphrodes bicincta,* yellow and green = inland and estuarine *A. makarovi,* respectively and purple = *A. aestuarina*. A number of specimens collected from the Medway estuary (orange) are shown separately due to a dis-concordance found between mating signal and mitochondrial DNA sequences in these specimens (section 3.5.3 and Chapter 5).

3.4.3. Mitochondrial DNA sequence diversity and phylogenetic analyses

Thirty-four unique mtDNA haplotypes for 658 bp (inter primer length) of the COI gene (GenBank Accession numbers; FR727167–FR727179, HE587025–HE587045) were identified for a total sample of 355 *Aphrodes* specimens. Eight individuals sequences could not be determined due to ambiguous bases. The *A. bicincta* haplotype 5 was only 618 bp long with 40 bp of missing data due to poor amplification of one strand sequenced. In total across all 33 haplotypes (excluding *A. bicincta* haplotype 5) there are 83 polymorphic sites and 575 invariable monomorphic sites. Of the 83 polymorphic sites, ten were singleton variable sites, 73 were parsimony informative sites (two variants = 62 and three variants = 11). There were a total of 95 substitutions, and the ratio of transitions to transversions was 2.5:1. The percentage of the 658 bp consisting of the nucleic acid bases Adenine, Thymine, Cytosine and Guanine were 25.70%, 43.37%, 14.30% and 16.62%, respectively. The mean number of pairwise differences overall was 20.9. The 658 bp COI region was protein coding (3.8, Appendix II for consensus amino acid protein translation). A total of 83 synonymous and 12 non-synonymous substitutions were identified.

For phylogenetic analyses rooted using the closely related species *Anoscopus limicola* (GenBank Accession numbers FR729924 and HE587046) the best-fit substitution model chosen for the data set using the Akaike information criterion (AIC), was the TPM3uf+G model (base frequencies of A = 0.2715, C = 0.1449, G = 0.1595, T = 0.4241, gamma distribution shape parameter = 0.2790). Phylogenetic trees rooted with *A. limicola* (Fig. 3.9) suggest that this genus is less closely related than previously thought (until recently classified in the same genus as *Aphrodes*), due to the deep branch connecting the *Aphrodes* ingroup with the *A. limicola*. A BLAST search on GenBank to identify a more suitable outgroup was unsuccessful as no sequences were identified with a higher than 81% similarity (*Cicadella viridis* mtDNA COI sequence, Virant-Doberlet *et al.* 2011). *Anoscopus limicola* shows c. 85% match to *Aphrodes* species. Using *A. limicola* as an outgroup, phylogenetic analyses resulted in an unresolved node in all three analysis methods depicting the relationship between *A. aestuarina* and the two sister species *A. makarovi/A. bicincta* (Fig. 3.9).

All *Aphrodes* species formed a monophyletic group and clustered into distinct clades with high nodal support for NJ, ML and BI phylogenies (Fig. 3.9) and the phylogenetic topologies were largely congruent with similarly high support values across analysis types. Each *Aphrodes* clade is characterised by short internal branches with a lack of support for within clade structure seen. In particular *A. makarovi* shows a large number of haplotypes many differing by only a single base pair, with shallow structure at the terminal branches. Table 3.3 shows the mean percentage of sequence divergence within and between species based on Kimura-2-parameter distance measures. Within species percentage sequence divergence was between 0.3 – 0.4% and between species ranges from 4.2 – 7.4%. The lowest sequence divergence seen was between *A. makarovi* and *A. bicincta* (4.2%) and the highest *A. aestuarina* and *A. bicincta* (7.4%).

Table 3.3. Average percentage sequence divergence within *Aphrodes* species (across the diagonal) and between *Aphrodes* species (below the diagonal) based on Kimura-2-parameter distance measure (Kimura 1980).

Sequence divergence (%) – Kimura-2-parameter				
	A. bicincta	***A. diminuta***	***A. makarovi***	***A. aestuarina***
A. bicincta	0.32			
A. diminuta	7.03	0.30		
A. makarovi	4.19	7.17	0.40	
A. aestuarina	7.42	6.70	6.93	0.36

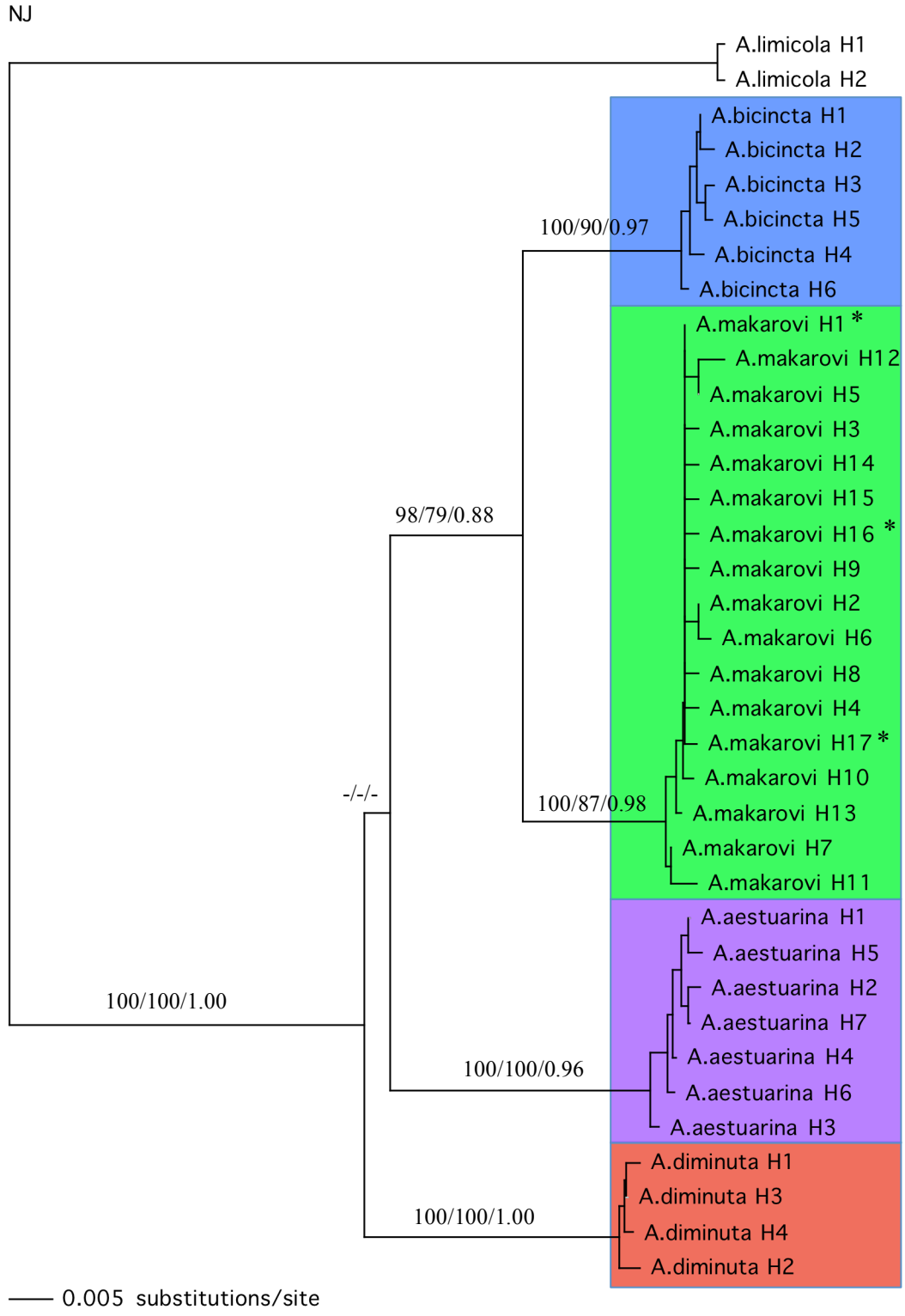


Figure 3.9. Neighbour-joining phylogeny using Kimura 2-parameter distances (Kimura 1980) for 658 bp of cytochrome oxidase subunit I gene for 34 *Aphrodes* haplotypes from four species, *Aphrodes makarovi* (green), *A. bicincta* (blue), *A. aestuarina* (purple) and *A. diminuta* (red). * Specimens from the Medway estuary were found emitting or responding to *A. aestuarina* vibrational mating signals but possessed mtDNA sequences of *Aphrodes makarovi.* Bootstrap support values greater than 75% are shown above the branches for neighbour-joining, maximum likelihood analysis and posterior probability support values greater than 0.75 are shown for Bayesian analysis, respectively. The phylogram is rooted using *Anoscopus limicola*. The scale bar represents 0.005 substitutions per site. H = haplotype. A dash (-) is presented when a node could not be recovered by one or more of the analyses described.

Table 3.4 gives the descriptive statistics for each species separately, including the number of polymorphic sites and substitution types, also showing molecular diversity indices. All species show low nucleotide diversity (between 0.1 – 0.2%). Both *A. bicincta* and *A. diminuta* show a higher haplotype diversity (0.7 and 0.9 respectively) compared with the moderate to low haplotype diversity in *A. aestuarina* (0.56) and *A. makarovi* (0.47).

Table 3.4. Descriptive statistics for each *Aphrodes* species (*Aphrodes bicincta, A. diminuta, A. makarovi* and *A. aestuarina*) as identified in phylogenetic analysis of 658 bp of the mitochondrial cytochrome oxidase subunit I (COI) gene.

Descriptive statistics				
	A. bicincta	***A. diminuta***	***A. makarovi***	***A. aestuarina***
Useable loci (bp)	618	658	658	658
Missing data (bp)	40*	/	/	/
Invariable sites	613	654	640	651
Polymorphic sites	5	4	18	7
- Singleton (2 variants)	3	4	14	5
- Parsimony informative (2 variants)	2	/	4	2
Substitutions	5	4	18	7
- Transitions/transversion ratio	4:1	3:1	1.6:1	6:1
- Synonymous	5	3	15	7
- Non-synonymous	1	1	3	0
Mean pairwise differences	1.08	1.53	0.68	0.74
Nucleotide diversity % (+/- SD)	0.17 (0.10)	0.23 (0.20)	0.10 (0.09)	0.11 (0.09)
Haplotype diversity (+/- SD)	0.68 (0.13)	0.87 (0.13)	0.47 (0.04)	0.56 (0.04)

*40 nucleotide sites with missing data for *Aphrodes bicincta* for haplotype H5

SD – standard deviation

A total of 18 specimens from the Medway estuary that were included in the validated species data set (i.e. included in all three analysis methods), either emitted or responded to *A. aestuarina* mating signals (11 males and seven females possessed *A. makarovi* mtDNA haplotypes (either haplotype H1, H16 or H17) (Table 3.5 and Fig. 3.10 (d)). Pure *A. makarovi* (matching mating signal and mtDNA COI sequences) were also identified in this population (seven males and one female). A number of specimens (n = 70) from this region were not recorded for their mating signals and so their species

status cannot be reliably determined based solely on mtDNA data (but see Chapter 5). It is not known if this mismatch between mtDNA and mating signal occurs in other locations; however, results show the mismatch is only present in the Medway_E population. All individuals from this region possessed *A. makarovi* mtDNA (n = 96) and none was found to match *A. aestuarina* mtDNA even though *A. aestuarina* mating signals were recorded from many individuals.

At all other localities, except the Medway estuary, individuals included in the validated data set (Table 3.5) gave concordant results for vibrational mating signals and mtDNA sequences (as well as specimens included in the additional data set that were included in both of these analyses).

Table 3.5. Information for each of the 13 sampled populations, including the host plant, population name, *Aphrodes* species present based on both mating signal and mitochondrial cytochrome oxidase subunit I gene (COI) sequence data. I=inland, E=estuarine.

Host plant	**Population**	**Species signal**	**Species COI**
Atriplex	Kent_E	*A. makarovi*	*A. makarovi*
Atriplex	Essex_E	*A. makarovi*	*A. makarovi*
		A. aestuarina	*A. aestuarina*
Atriplex	Gower_E	*A. makarovi*	*A. makarovi*
Atriplex	Norfolk_E	*A. makarovi*	*A. makarovi*
		A. aestuarina	*A. aestuarina*
Atriplex	Medway_E	*A. makarovi*	*A. makarovi*
		A. aestuarina*	***A. makarovi****
Atriplex	Devon_E	*A. makarovi*	*A. makarovi*
Atriplex	Sussex_E	*A. aestuarina*	*A. aestuarina*
Urtica	Gower_I	*A. makarovi*	*A. makarovi*
Urtica	Norfolk_I	*A. makarovi*	*A. makarovi*
Urtica	Lisvane_I	*A. makarovi*	*A. makarovi*
Urtica + Fabaceae	Sussex_I	*A. makarovi*	*A. makarovi*
		A. diminuta	*A. diminuta*
Fabaceae	Aberdare_I	*A. bicincta*	*A. bicincta*

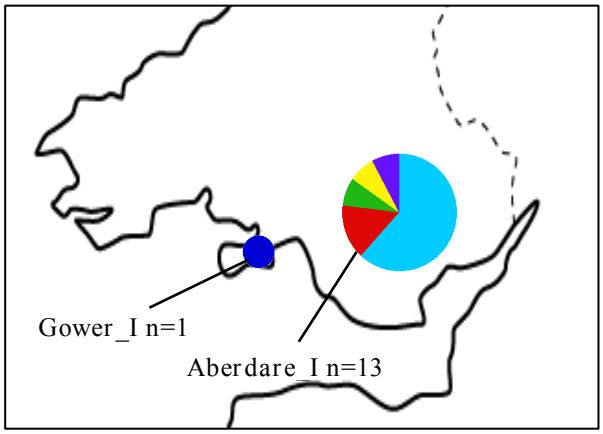
* Specimens from the Medway estuary were found emitting or responding to *Aphrodes aestuarina* vibrational mating signals but possessed mtDNA sequences of *Aphrodes makarovi.*

Figure 3.10 (a-d) illustrates the mtDNA haplotype distribution for all *Aphrodes* species at the locations sampled. Distributions for *A. diminuta* and *A. bicincta* are from only a single location with low sample sizes and are therefore not representative of their overall species distributions. One individual collected from the Gower_I population (typically *A. makarovi* host, *Urtica* species) gave *A. bicincta* haplotype H6 suggesting some degree of overlap for these two species ranges depending on host plant distributions. It is common to find Fabaceae in and around grassland habitats amongst other *Aphrodes* species (not *A. aestuarina*) host plant types.

Aphrodes aestuarina (Fig. 3.10 (c)) mtDNA haplotypes were only identified at three estuarine localities (Norfolk_E, Essex_E and Sussex_E). The *A. aestuarina* haplotype H1 is common along the coast of Essex_E and Sussex_E sites and haplotype H2 common in Norfolk_E. No haplotypes present in the Norfolk_E population were found in the other locations and vice versa. No *A. aestuarina* populations were identified in the West of the UK; however, this may be due to insufficient sampling of the area. No saltmarsh habitats were identified on the North coast of Devon or further south of Budleigh-Salterton where the shoreline tends to be rocky rather than saltmarsh habitat.

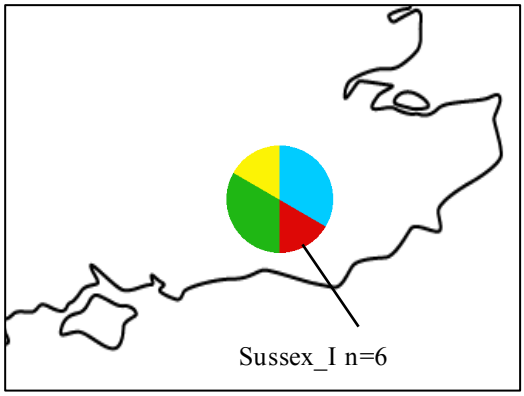
For *A. makarovi* a considerable number of haplotypes were identified (17) compared to other *Aphrodes* species (between four and seven) although the sample size is also the highest for this species as *A. makarovi* is common in both grassland and estuarine habitats, on two host plants/ecological niche types. *Aphrodes makarovi* haplotype H1 is present across all sampled locations and other haplotypes are present in lower frequency and generally site specific (Fig. 3.10 (d)).

(a) *A. bicincta* – South Wales



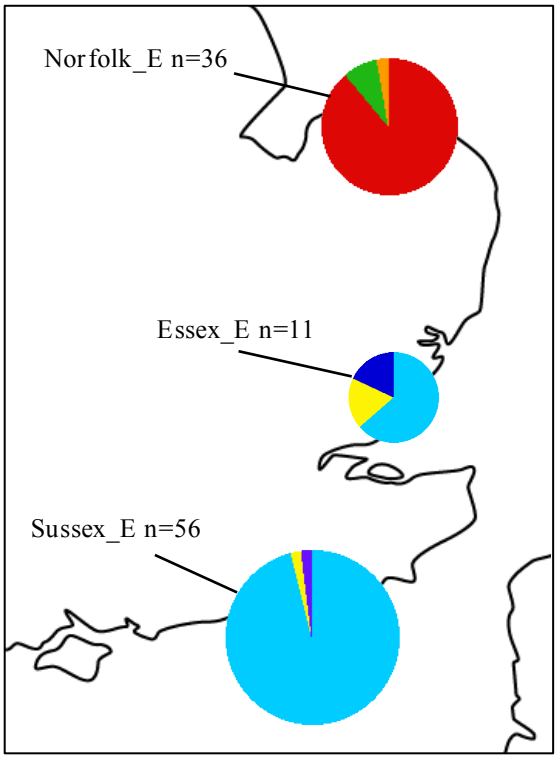
Haplotype 1
Haplotype 2
Haplotype 3
Haplotype 4
Haplotype 5
Haplotype 6

(b) *A. diminuta* – South East England



Haplotype 1
Haplotype 2
Haplotype 3
Haplotype 4

(c) *A. aestuarina* –East & South East England



Haplotype 1
Haplotype 2
Haplotype 3
Haplotype 4
Haplotype 5
Haplotype 6
Haplotype 7

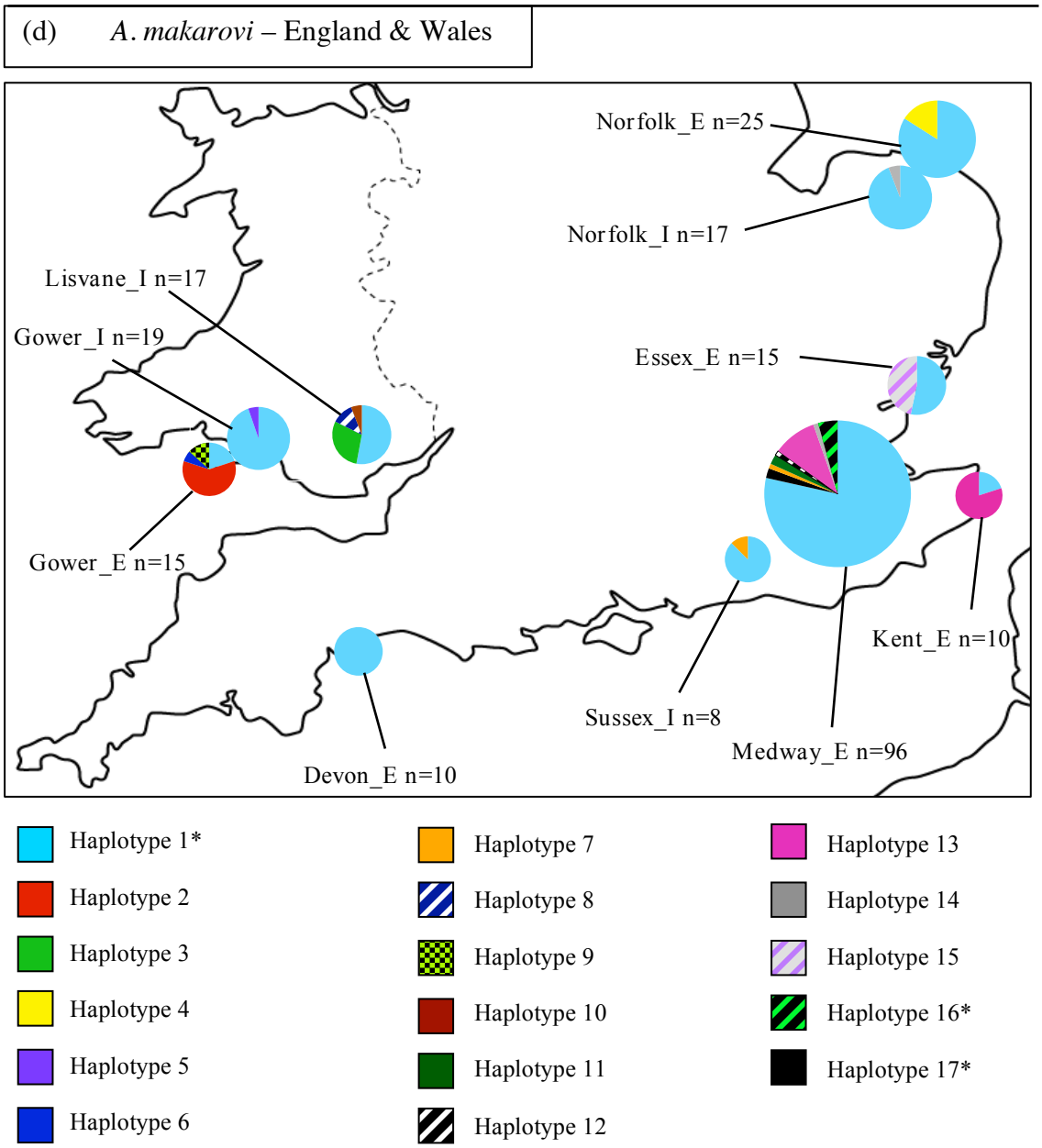


Figure 3.10. Mitochondrial cytochrome oxidase subunit I gene haplotype distribution maps for four *Aphrodes* species identified across UK sampling locations. (a) = *Aphrodes bicincta*, (b) = *A. diminuta*, (c) = *A. aestuarina,* (d) = *A. makarovi.* * = *A. makarovi* haplotypes identified from Medway_E specimens that emitted or responded to *A. aestuarina* mating signals. I = Inland locations and E = Estuarine habitats.

3.5. Discussion

Using a multidisciplinary approach to examine morphological, behavioural and genetic differentiation among *Aphrodes* species provided validated specimens that have been unequivocally identified as representatives of each *Aphrodes* species. Evidence from recorded male vibrational calling signals (and female response to playback signals) and phylogenetic analyses of mtDNA COI sequences provided strong support for four distinct species. This result was congruent across all sampling localities (except the Medway estuary). Morphological analyses, however, proved unreliable for species identification, as suspected. Each species was discovered co-occurring with another *Aphrodes* species in at least one sampling locality giving evidence that they can be found and do remain distinct in sympatry. The evolution of species-specific mating signals suggests reproductive isolation is near complete, regardless of apparently overlapping morphology. In light of the behavioural and genetic differentiation identified between *Aphrodes* species in sympatry, most species concepts would recognise that the four *Aphrodes* species described clearly represent four distinct taxa (Coyne & Orr 2004; Mallet 2008).

3.5.1. Vibrational signals

The mating signals that were recorded correspond to those previously described by Tishechkin (1998) for *A. makarovi, A. bicincta* and *A. diminuta*. *Aphrodes diminuta* has also been identified in the UK at two locations in Sussex_I, which is the first documented evidence for this species occurring in the UK. The previously unrecorded mating signal of *A. aestuarina* has been identified and confirmed using a combination of analyses to correlate results from museum specimens (Bluemel *et al.* 2011 – Chapter 2) and freshly collected specimens from saltmarsh habitats at the syntype location (Norfolk). The combined use of historical, ecological, behavioural and molecular data to validate freshly collected specimens that were compared to museum type specimens using DNA analysis, clearly represents a valuable approach when analysing morphologically similar taxa, with a history of taxonomic uncertainty (Austin & Arnold 2001; Austin & Melville 2006). In the syntype museum collection for *A. aestuarina* (Edwards 1908) both *A. makarovi* and *A. aestuarina* were identified using molecular

tools in Chapter 2 (Bluemel *et al.* 2011). This correlates with results for specimens also found at a local population in Norfolk c. 100 years later. These concordant results for this location suggest that the identification of museum specimens using an array of tools can provide reliable results.

Calling signals emitted by males of each species are distinct and contain highly species-specific song elements (Virant-Doberlet *et al.* 2005). Within the boundary of this study, no intermediate signals have been found and it would be impossible to mistake the song of one species for that of another (Fig. 3.4). Identification of the *A. aestuarina* mating signal that is distinct from other *Aphrodes* species gives good evidence for the species status of *A. aestuarina* (as well as distinct mtDNA with c. 6 – 7% divergence from other *Aphrodes* species, see 3.5.3) even when both *A. makarovi* and *A. aestuarina* occur in sympatry (Norfolk_E and Essex_E). However, the song pattern of *A. aestuarina* is composed of elements also present in *A. bicincta* males found in Aberdare (Fig. 3.4, Virant-Doberlet *et al.* 2005), and female preference experiments indicated no behavioural barrier based on signals between *A. bicincta* and *A. aestuarina* from Medway_E and Sussex_E populations. *Aphrodes bicincta* females made no discrimination and readily responded to males producing the *A. aestuarina* mating signal and also, females from Sussex_E and Medway_E populations made no discrimination between the signal types of *A. bicincta* or *A. aestuarina* and readily responded to both (Virant-Doberlet, personal communication).

The occasional identification of similar/overlapping song features in morphologically cryptic song species has also been documented (Henry *et al.* 1999b) although only in allopatric species pairs. This pattern may be expected as prezygotic isolation is likely to evolve more rapidly in sympatric compared to allopatric species pairs (Coyne & Orr 2004). To date, *A. bicincta* has not been identified in sympatry with *A. aestuarina* and they are both found in very different habitats (inland and estuarine, respectively) suggesting that song similarity due to similar environmental constraints or selection pressures is unlikely. Furthermore, other *Aphrodes* species are found in similar habitats even on the same host plants (e.g. estuarine *A. makarovi* and *A. aestuarina*), but differ considerably in signal type. Although, the possibility of evolutionary convergence due to environmental adaptation cannot be ruled out without further investigations of habitat preference, host fidelity and signal transmission properties in different host plant substrates. Alternatively, *A. bicincta* and *A. aestuarina* could exhibit similarities in song

features due to the retention of an ancestral (plesiomorphic) state from a more distant common ancestor (Henry *et al.* 1999b). They are not members of the same species and do not share the most recent common ancestor based on the mtDNA COI phylogeny presented here. Random mutations may also explain the convergence seen in vibrational signal song elements in these species if signal traits can be altered considerably by single mutations (Henry *et al.* 1999b).

3.5.2. Morphology

A large overlap in morphological form between *Aphrodes* species confirms that morphological measures used in this study are unsuitable for distinguishing *Aphrodes* species with any confidence. Analyses of male aedeagus form clearly identified *A. diminuta* as the most distinct of the four species (although sample sizes are low for aedeagus morphology), with this species showing a larger distance between the spines compared with overall aedeagus length than other *Aphrodes* species, a morphological feature previously proposed for species delimitation (Tishechkin 1998). Given the considerable time and effort required to analyse aedeagal morphology and the degree of overlap in characters identified in this study, morphological methods for *Aphrodes* identification is not recommended. If larger sample sizes were included it is likely that a greater overlap would be seen based on the degree of overlap and degree of within-species variation reported. Further sampling would confirm this possibility.

Both estuarine *A. makarovi* and *A. aestuarina* share the same host/habitat type and are often in sympatry occurring on the same host plant (*Atriplex portulacoides*), as found at two locations (Norfolk_E, Essex_E). *Aphrodes makarovi* is also common in inland habitats on the previously identified host plant types, *Urtica, Taraxacum* and *Cirsium* species (Biedermann & Niedringhaus 2004). Very little evidence for morphological differences between the two saltmarsh-adapted species was observed and considerable overlap was identified. Morphological differences, however, between *A. makarovi* occurring in different habitats (inland and estuarine) were observed, particularly in females, where estuarine forms of both species were most similar to each other (longer thinner) compared to inland *A. makarovi* (shorter and wider), but with considerable overlap. Males on the other hand, formed clusters, with inland *A. makarovi* and *A. aestuarina* being more distinct compared to other species, but again with much overlap.

Distinct differences in the banding patterns on the head of inland and estuarine *A. makarovi* populations were also observed. Inland populations were typically made up of specimens with varied banding patterns but typically with a higher degree of banding with darker pigmentation than their estuarine counterparts. The estuarine adapted *A. makarovi* specimens showed less varied pigmentation (lighter and less banded) very similar to *A. aestuarina* occurring in the same habitat. The importance of processes in contributing to the maintenance of colour polymorphism and the evolution of reproductive isolation is well acknowledged (Gray & McKinnon 2007). Among other factors, random genetic drift can maintain colour polymorphism, although may be difficult to detect due to the likely involvement of a number of processes (Gray & McKinnon 2007). The similarities between the estuarine adapted *Aphrodes* species may represent a physiological response to varying predation levels and the need for morphological background crypsis in different habitats resulting in evolutionary convergence of body pigmentation when in the same habitat (Nosil & Crespi 2006; Rosenblum 2006). No information is currently known about the effect that *Aphrodes* predators may have on *Aphrodes* colour morphology in different habitats. Morphological analysis of banding patterns and pigmentation intensity was explored to initially identify the significance of this observation (Chapter 4) and to formulate further hypotheses relating to this topic.

3.5.3. Mitochondrial DNA

Phylogenetic analyses of mtDNA COI sequences resulted in well-supported clades implying the existence of four genetically distinct species. With exception of the Medway estuary population there are no shared haplotypes among specimens that had been behaviourally classified as belonging to a particular *Aphrodes* species, even in areas of sympatry. The overall mtDNA COI gene sequence divergence between *Aphrodes* species was similar to that reported for other congeneric taxa (Hebert *et al.* 2003). Hebert *et al.,* (2003) reported that sequence divergences of greater than 3% were common between lepidopteran species pairs (98% in this study). Lower divergence values of less than 2% were suggestive of recent origin. Considering a 3% sequence threshold as a guide for species identification, the lowest divergence seen between the two sister species *A. makarovi* and *A. bicincta* showing c. 4% sequence divergence and divergence in other pairwise species comparisons was similar (average 7%) giving good

evidence for four distinct taxa. Although there are limitations that need to be considered when interpreting sequence divergence as a standard for species identification (Meyer & Paulay 2005) and the effectiveness of mtDNA has been questioned for use in phylogeograpahical studies (Ballard & Whitlock 2004), this marker still remains useful for phylogenetic and taxonomic inference, as long as the limitations are comprehended (Bensasson *et al.* 2001; Meyer & Paulay 2005; Löytynoja & Goldman 2008). Thus, a holistic approach for species identification was employed, examining behavioural, morphological and molecular differentiation among *Aphrodes* species.

Within the *A. makarovi* lineage a considerable number of haplotypes were identified (17) compared to other *Aphrodes* species (between four and seven) although the sample size was also the highest for this species as *A. makarovi* was common in both grassland and estuarine habitats, on two host plants/ecological niche types. One mtDNA haplotype was widespread and likely to be ancestral (*A. makarovi* haplotype H1), while other haplotypes were geographically local, lower in lower frequency compared with *A. makarovi* haplotype H1, and are likely to be recent mutations that have not spread throughout populations (Avise *et al.* 1987). Species with this mtDNA haplotype distribution indicate phylogeographic continuity and life histories associated with intermediate gene flow with weak long-term barriers to gene flow (Avise *et al.* 1987). Based on our sampling effort, the distribution of *A. aestuarina* was restricted to salt-marsh habitats around the East and Southeast coasts of England. Two more common haplotypes were found in this species, one in Norfolk and the other in Essex and Sussex regions. An additional five haplotypes were identified in lower frequency and typically location specific.

If the patterns of mtDNA diversity reflect the true demographic history of *Aphrodes* species then the low nucleotide diversity (0.10 and 0.11), moderate haplotype diversity (0.47 and 0.56) and shallow phylogenetic structure identified for both *A. makarovi* and *A. aestuarina,* respectively, do not reflect patterns suggested for a stable population that has a large effective population size (Avise *et al* 1987; Grant & Bowen 1998; Avise 2000). The patterns identified are more suggestive of abundant species that have undergone a population bottleneck and recent population expansion from a small effective population size (Avise *et al.* 1987; Grant & Bowen 1998; Avise 2000).

To determine the phylogeographic history of *Aphrodes* species was not the main aim of this study and due to the low sample sizes (collected from a single region) for both *A. bicincta* and *A. diminuta,* it would be unwise to do so for these species. Furthermore, inference of phylogeographic history based solely on mtDNA can lead to incorrect conclusions due to other processes (introgression, selective sweeps and cytoplasmic infections) that can influence patterns of mtDNA variation (Ballard & Whitlock 2004). Mitochondrial DNA should also not be assumed to be a neutral marker without first undertaking appropriate tests to satisfy assumptions of neutral model of evolution (Ballard & Whitlock 2004). Sampling on a larger geographical scale (including possible glacial refugia locations across Europe), employing the use of a number of molecular markers (including mitochondrial and nuclear DNA) and appropriate statistical tests of selective neutrality are needed to further explore the phylogeographic history of the *Aphrodes* genus.

Identification of a mismatch between mating signal and mtDNA COI sequence data for 18 individuals from the Medway estuary highlight the importance of the limitations emphasised above, when inferences concerning the history of species are based solely on patterns of mtDNA variation. These mismatched individuals were those that emitted the mating signal of *A. aestuarina* (or females that responded to *A. aestuarina* playback signals) but their mtDNA sequences matched *A. makarovi*. Non-mismatched *A. makarovi* specimens were also found in the Medway estuary, but non-mismatched *A. aestuarina* was absent (closest location of non-mismatched *A. aestuarina* to the Medway estuary was the Essex_E population). Results presented here highlight another important avenue of research to be explored within the *Aphrodes* genus concerning possible reasons for the mismatch identified (Chapter 5).

The dis-concordance between signal pattern and mtDNA could be the result of introgression of *A. makarovi* mtDNA into the genetic background of *A. aestuarina* and thus, could present possible evidence that *A. makarovi* and *A. aestuarina* can remain distinct despite some level of hybridisation and gene flow in the past. This phenomenon is more common than previously though (Ballard & Whitlock 2004; Mallet 2005; Mallet 2008). Although, such patterns of dis-concordance are also generated through retention of an ancient ancestral polymorphism due to incomplete lineage sorting, but this is more likely in very recently diverged species (Ballard & Whitlock 2004). Because mtDNA alone cannot be used to decipher among such hypotheses, the nuclear

genetic differentiation between *A. aestuarina* and *A. makarovi* was explored to evaluate the likelihood of introgression between saltmarsh adapted *Aphrodes* and evidence for hybrid origin of mismatched *Aphrodes* present in the Medway estuary (Chapter 5).
It is also possible that because vibrational mating signals were recorded for a subset of individuals in this study, and larger numbers of specimens were analysed using molecular methods, the other unrecorded males or females from other sampled populations of estuarine *A. makarovi* and *A. aestuarina* may also in fact be misclassified when based solely on mitochondrial data. With this in mind it is also possible that, without vibrational signal data, the museum specimens identified as *A. makarovi* in the syntype series of *A. aestuarina* based on mtDNA (Bluemel *et al.* 2011 – Chapter 2) may also be incorrectly identified in this analysis. However, based on the current data the mismatched specimens are limited to sites around the Medway estuary. Further analysis using nuclear markers will hopefully elucidate these findings.

In addition to environment/host plant differences, slight morphological differentiation in body shape between inland and estuarine *A. makarovi* (and the observation of colour and head pattern differences), no mtDNA haplotypes were shared between inland and estuarine populations, apart from the common *A. makarovi* haplotype H1 and the less frequent haplotype H14 (shared between Norfolk_I and Medway_E). If the distribution of mtDNA variation can be attributed to host/ecological type rather than random processes (genetic drift) occurring at different geographic localities, this may give insights into whether ecological selection is an important factor in determining the pattern of genetic variation between inland and estuarine *A. makarovi* populations (Chapter 4).

Adaptation to different host plants is thought to be a main reason for the high diversity of phytophagous insects observed in nature (Berlocher & Feder 2002; Funk *et al.* 2002), whereby ecologically driven selection leads to reproductive isolation (Via 2001; Drès & Mallet 2002). It is known that leafhoppers oviposit directly into host plant tissues and emergence and development times of nymphs are likely to be linked with host phenology (Dietrich *et al*. 1999). Alternatively, adaptation along an environmental abiotic gradient such as salinity or altitude, could also explain the diversification of many species (Mallet 2008). It is likely that a number of factors contribute to ecological race formation. Further morphological and genetic analysis of inland and estuarine adapted *A. makarovi* using multiple unlinked nuclear loci was performed to test the

hypothesis that these populations are differentiated and represent the initial stages of host/ecological race formation (Chapter 4). Genetic analysis of nuclear variation is required to identify if there is any genetic differentiation between these ecological forms when found in sympatry (at Gower_E and Gower_I or Norfolk_E and Norfolk_I sampling locations) (Chapter 4).

3.6. Conclusion

A combination of male mating signal/female preference and mtDNA COI sequence data provided a robust method for identification of *Aphrodes* and presented good support for the existence of four distinct species, despite considerable morphological similarity. More importantly this study revealed new research avenues worthy of exploration concerning the evolutionary processes acting within the *Aphrodes* genus, including the possibility of ecological adaptation, convergent evolution and introgression, within and between *Aphrodes* species. Such hypotheses cannot be distinguished using only mtDNA as a molecular marker and further examination of the patterns of nuclear variation, employing the use of multiple unlinked nuclear loci, is required (Chapter 4, Chapter 5).

This study highlights that using a holistic approach to define the framework of biological diversity, whether concealed or obvious, is of great importance to unravelling the processes involved in driving and maintaining the evolution of species, above and below the species level.

3.7. References

Austin, J. J., & Arnold, E. N. (2001). Ancient mitochondrial DNA and morphology elucidate an extinct island radiation of Indian Ocean giant tortoises (*Cylindraspis*). *Proceedings of the Royal Society B: Biological Sciences*, *268*, 2515-2523.

Austin, J. J., & Melville, J. (2006). Incorporating historical museum specimens into molecular systematic and conservation genetics research. *Molecular Ecology Notes, 6*, 1089–1092.

Avise, J. C., Arnold, R. M., Ball, E., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., & Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review Ecology and Systematics*, *18*, 489-522.

Avise, J. C. (2000). *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*, 729-744.

Bensasson, D., Zhang, D.-X., Hartl, D. L., & Hewitt, G. M. (2001). Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology & Evolution*, *16*, 314-321.

Berlocher, S. H., & Feder, J. L. (2002). Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Review of Entomology*, *47,* 773-815.

Biedermann, R., & Niedringhaus, R. (2004). *Die Zikaden Deutschlands*. WABV, Scheeßel, (pp. 214-215).

Bluemel, J. K., King, R. A., Virant-Doberlet, M., & Symondson, W. O. C. (2011). Primers for identification of type and other archived specimens of *Aphrodes* leafhoppers (Hemiptera, Cicadellidae). *Molecular Ecology Resources, 11*, 770-774.

Charif, R. A., Clark, C. W., & Fristup, K. M. (2004). Raven 1.2. User's manual, Cornell Laboratory of Ornithology, Ithaca, NY.

Claridge, M. F. (1985). Acoustic signals in the Homoptera: behaviour, taxonomy, and evolution. *Annual Review of Entomology, 30*, 297-317.

Cocroft, R. B., & Rodríguez, R. L. (2005). The behavioral ecology of insect vibrational communication. *BioScience*, *55*, 323-334.

Čokl, A., & Virant-Doberlet, M. (2003). Communication with substrate-borne signals in small plant-dwelling insects. *Annual Review of Entomology, 48*, 29-50.

Coyne, J. A., & Orr, H. A. (2004). *Speciation.* Sinauer Associates, Sunderland, USA.

Dietrich, C. H. (1999). The role of grasslands in the diversification of leafhoppers (Homoptera: Cicadellidae): a phylogenetic perspective. *Proceeding of the Fifteenth North American Prairie Conference,* (pp. 44-49).

Drès, M., & Mallet, J. (2002). Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *357*, 471-492.

Edwards, J. (1908). On some British Homoptera hitherto undescribed or unrecorded. *Entomology Monthly Magazine, 44*, 56-59.

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

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution, 39*, 783-791.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & 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.

Funk, D. J., Filchak, K. E., & Feder, J. L. (2002). Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica*, 116, 251-267.

Grant, S. W., & Bowen, B. W. (1998). Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *Journal of Heredity*, *89*, 415-426.

Gray, S. M., & McKinnon, J. S. (2007). Linking color polymorphism maintenance and speciation. *Trends in ecology & evolution*, *22*, 71-9.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, *52*, 696-704.

Hamilton, K. G. A. (1975). A review of the Northern hemisphere *Aphrodina* (Rhynchota: Homoptera: Cicadellidae) with special reference to the Nearctic fauna. *The Canadian Entomologist*, *107*, 1009–1027.

Hamilton, K. G. A. (1983). Introduced and native leafhoppers common to the old and new worlds (Rhynchota: Homoptera: Cicadellidae). *Canadian Entomologist, 115*, 473-511.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of Royal Society Biological Sciences*, *270*, 313-321.

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences, 101*, 14812-14817.

Henry, C. S., Brooks, S. J., Johnson, J. B., & Duelli, P. (1999a). Revised concept of *Chrysoperla mediterranea* (Hölzel), a green lacewing associated with conifers: courtship songs across 2800 kilometres of Europe (Neuroptera: Chrysopidae). *Systematic Entomology*, *24*, 335-350.

Henry, C. S., Wells, M. L. M., & Simon, C. M. (1999b). Convergent evolution of courtship songs among cryptic species of the *carnea* group of green lacewings (Neuroptera: Chrysopidae: *Chrysoperla*). *Evolution, 53*, 1165-1179.

Henry, C. S., Brooks, S. J., Duelli, P., & Johnson, J. B. (2002). Discovering the true *Chrysoperla carnea* (Insecta: Neuroptera: Chrysopidae) using song analysis,

morphology, and ecology. *Annals of the Entomological Society of America*, *95*, 172-191.

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *15*, 111–120.

Kirby, P. (1992). *A Review of the Scarce and Threatened Hemiptera of Great Britain.* No. 2. The Joint Nature Conservation Committee. (pp. 154).

Le Quesne, W. J., & Payne, K. R. (1981). *Cicadellidae (Typhlocybinae) with a Checklist of the British Auchenorrhyncha (Hemiptera, Homoptera)*. Vol. II, Part 2(c). Fitton, M.G. (ed.). Royal Entomological Society of London. (pp. 45).

Le Quesne, W. J. (1988). The *Aphrodes bicinctus* complex. *Proceedings of the 6th Auchenorryncha meeting, Turin, Italy,* (pp. 413–415).

Löytynoja, A., & Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science*, *320*, 1632-1635.

Mallet, J. (2005). Hybridisation as an invasion of the genome. *Trends in Ecology and Evolution, 20*, 229-237.

Mallet, J. (2008). Hybridisation, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363,* 2971-2986.

Mayr, E. (1963). *Animal species and evolution.* Harvard University Press, Cambridge.

Mazzoni, V., Lucchi, A., Čokl, A, Prešern, J., & Virant-Doberlet, M. (2009). Disruption of reproductive behaviour of *Scaphoideus titanus* by playback of vibrational signals. *Entomologia experimentalis et applicata*, *133*, 174-185.

McNett, G. D., & Cocroft, R. B. (2008). Host shifts favor vibrational signal divergence in *Enchenopa binotata* treehoppers. *Behavioral Ecology*, *19*, 650-656.

Meyer, C. P., & Paulay, G. (2005). DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, *3*, 1-10.

Nast, J. (1976). Peiwiki (Homoptera, Auchenorrhyncha) Peinin. *Fragmenta Faunistica*. XXI, 157-160.

Nei, M. (1987). *Molecular Evolutionary Genetics*. Columbia University Press, New York, USA. (pp. 180, 257).

Nickel, H., & Remane, R. (2002). *Check list of the planthoppers and leafhoppers of Germany, with notes on food plants, diet width, life cycles, geographic range and conservation status (Hemiptera, Fulgoromorpha and Cicadomorpha)*. Artenliste der Zikaden Deutschlands.

Nickel, H. (2003). *The leafhoppers and planthoppers of Germany (Hemiptera, Auchenorrhyncha): patterns and strategies in a highly diverse group of phytophagous insects.* Pensoft Publishers, Sofia-Moscow. (pp. 103-105).

Nosil, P., & Crespi, B. J. (2006). Experimental evidence that predation promotes divergence in adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 9090-9095.

Paterson, H. E. H. (1985). The recognition concept of species. In *Species and Speciation* (E.S. Vrba, Ed.), Transvaal Museum Monograph No. 4. Transvaal Museum, Pretoria. (pp. 21-29).

Pfenninger, M., Cordellier, M., & Streit, B. (2006). Comparing the efficacy of morphological and DNA-based taxonomy in the freshwater gastropod genus *Radix* (Basommatophora, Pulmonata). *BMC Evolutionary Biology, 6*, 1-14.

Posada, D. (2008). jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*, *25*, 1253-1256.

Price, B. W., Barker, N. P., & Villet, M. H. (2007). Patterns and processes underlying evolutionary significant units in the *Platypleura stridula* L. species complex (Hemiptera: Cicadidae) in the Cape Floristic Region, South Africa. *Molecular Ecology, 16,* 2574-2588.

Quicke, D. L. J. (1993). *Principles and techniques of contemporary taxonomy*. Blackie Academic & Professional, London.

Rasband, W. S. (1997-2007). ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/.

Ronquist, F., & Huelsenbeck, J. P. (2003). MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*, 1572–1574.

Rosenblum, E. B. (2006). Convergent evolution and divergent selection: lizards at the white sands ecotone. *The American Naturalist*, *167*, 1-15.

Rozas J., Sánchez-DelBarrio, J. C., Messeguer, X., & Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, *19*, 2496–2497.

Sites, J. W., & Marshall, J. C. (2004). Operational criteria for delimiting species. *Annual Review of Ecology, Evolution, and Systematics*, *35*, 199-227.

Swofford, D. L. (2002). PAUP: Phylogenetic Analysis Using Parsimony (*and other methods), version 4.0 b10. Sinauer Associates, Sunderland, Massachusetts.

Tajima, F. (1983). Evolutionary relationship of DNA sequences in finite populations. *Genetics*, *105*, 437-460.

Tishechkin, D. Y. (1998). Acoustic signals and morphological characters of leafhoppers belonging to *Aphrodes bicinctus* group (Homoptera, Cicadellidae) from European Russia. *Entomological Review, 78*, 370-377.

Towes, D. P., & Irwin, D. E. (2008). Cryptic speciation in a Holarctic passerine revealed by genetic and bioacoustics analyses. *Molecular Ecology, 17*, 2691-2705.

Via, S. (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology and Evolution, 16*, 381-390.

Virant-Doberlet, M., & Čokl, A. (2004). Vibrational communication in insects. *Neotropical Entomology, 33*, 121–134.

Virant-Doberlet, M., Prešern, J., Stewart, A. J. A., & Wilson, M. R. (2005). The genus *Aphrodes*: bioacoustic approach to solve taxonomic problems. *Congress of International BioAcoustic Council (IBAC)*. (pp. 35).

Virant-Doberlet, M., Prešern, J., Bluemel, J., King, A. R., Wilson, M. R., Symondson, W. O. C., & Stewart, A. J. A. (2006). Vibrational signals and reproductive isolation in the genus *Aphrodes* (Hemiptera: Cicadellidae). *11th International Meeting on Invertebrate Sound and Vibration,* University of Toronto at Scarborough.

Virant-Doberlet, M., & Žežlina, I. (2007). Vibrational Communication of *Metcalfa pruinosa* (Hemiptera: Fulgoroidea: Flatidae). *Annals of the Entomological Society of America*, *100*, 73-82.

Virant-Doberlet, M., King, R. A., Polajnar, J., & Symondson, W. O. C. (2011). Molecular diagnostics reveal spiders that exploit prey vibrational signals used in sexual communication. *Molecular Ecology*, *20*, 2204-2216.

3.8. Appendix

Appendix I – Sampling locations and geographic coordinates

Location	Host plant/ habitat type	Country	Geographic co-ordinate
Kent_E			
Pegwell Bay	*Atriplex* - saltmarsh	England	N 51° 18.954' E 001° 21.590'
Essex_E			
Canvey Bridge	*Atriplex* - saltmarsh	England	N 51° 32.551' E 000° 33.834'
Canvey Island	*Atriplex* - saltmarsh	England	N 51° 31.343' E 000° 37.025'
Mersea Island	*Atriplex* - saltmarsh	England	N 51° 47.728’ E 000° 55.322’
Horsey Island	*Atriplex* - saltmarsh	England	N 51° 51.552’ E 001° 14.649’
Gower_E			
Penclaudd	*Atriplex* - saltmarsh	Wales	N 51° 38.646’ W 004° 06.659’
Norfolk_E			
Morston Quay	*Atriplex* - saltmarsh	England	N 52° 57.543' E 000° 59.060'
Blakney Quay	*Atriplex* - saltmarsh	England	N 52° 57.343' E 001° 00.827'
Stiffkey Marsh	*Atriplex* - saltmarsh	England	N 52° 57.621' E 000° 55.389'
Warham Marsh	*Atriplex* - saltmarsh	England	N 52° 57.367' E 000° 54.505'
Wells East Quay	*Atriplex* - saltmarsh	England	N 52° 57.417' E 000° 51.851'
Overy Marsh	*Atriplex* - saltmarsh	England	N 52° 57.792' E 000° 44.239'
Thornham	*Atriplex* - saltmarsh	England	N 52° 57.979' E 000° 35.135'
Holne-next-the-sea	*Atriplex* - saltmarsh	England	N 52° 58.064' E 000° 31.696'
Medway_E			
R (site destroyed)	*Atriplex* - saltmarsh	England	N 51° 23.460' E 000° 30.560
R1	*Atriplex* - saltmarsh	England	N 51° 23.597' E 000° 29.637'
R2	*Atriplex* - saltmarsh	England	N 51° 23.500' E 000° 29.670'
R3-Wouldham Marsh	*Atriplex* - saltmarsh	England	N 51° 22.367' E 000° 27.994'
R4-Rochester Castle	*Atriplex* - saltmarsh	England	N 51° 23.280' E 000° 29.952'
R5-Baty’s Marsh	*Atriplex* - saltmarsh	England	N 51° 22.588' E 000° 28.986'
R6-Riverside Walk	*Atriplex* - saltmarsh	England	N 51° 23.348’ E 000° 30.617’
R7-Gillingham Pier	*Atriplex* - saltmarsh	England	N 51° 23.849’ E 000° 33.327’
R8-Chatham Reach	*Atriplex* - saltmarsh	England	N 51° 24.015’ E 000° 30.945’
R9-Hoo St Werbergs	*Atriplex* - saltmarsh	England	N 51° 24.677’ E 000° 33.727’
R10-Stoke/Grain	*Atriplex* - saltmarsh	England	N 51° 26.959’ E 000° 39.356’
Lower Twydall	*Atriplex* - saltmarsh	England	N 51° 23.198' E 000° 35.610'
Funton Creek	*Atriplex* - saltmarsh	England	N 51° 22.807' E 000° 41.825'

Devon_E			
Budleigh-Salterton	*Atriplex* - saltmarsh	England	N 50° 37.579' W 003° 18.726'
Sussex_E			
Shoreham	*Atriplex* - saltmarsh	England	N 50° 50.456' W 000° 17.387'
Rye Harbour	*Atriplex* - saltmarsh	England	N 50° 56.223' E 000° 45.822'
Cuckmere Haven	*Atriplex* - saltmarsh	England	N 50° 46.384' E 000° 08.771'
Newhaven	*Atriplex* - saltmarsh	England	N 50° 47.889' E 000° 02.787'
Littlehampton	*Atriplex* - saltmarsh	England	N 50° 48.826’ W 000° 33.595’
Pagham Harbour	*Atriplex* - saltmarsh	England	N 50° 46.197' W 000° 45.354'
Gower_I			
Penclaudd	*Urtica* - Grassland	Wales	N 51° 38.609’ W 004° 05.742’
Church Lane	*Urtica* - Grassland	Wales	N 51° 38.020’ W 004° 05.992’
Graveyard	*Urtica* - Grassland	Wales	N 51° 38.262’ W 004° 05.925’
Near Ilston	*Urtica* - Grassland	Wales	N 51° 35.482’ W 004° 05.429’
Norfolk_I			
Warham Marsh coastal footpath	*Urtica* - Grassland	England	N 52° 57.367' E 000° 54.505'
Stiffkey Marsh coastal footpath	*Urtica* - Grassland	England	N 52° 57.409’ E 000° 55.384’
Stiffkey Marsh Road	*Urtica* - Grassland	England	N 52° 57.272’ E 000° 55.431’
Stiffkey-Wells Road	*Urtica* - Grassland	England	N 52° 56.946’ E 000° 54.404’
Lisvane_I			
Lisvane	*Urtica* - Grassland	Wales	N 51° 32.160’ W 003° 10.173’
Aberdare_I			
Mountain Ash	Fabaceae - Brown field site	Wales	N 51° 41.616’ W 003° 24.406’
Sussex_I			
Castle Hill	Fabaceae - Grassland	England	N 50° 50.868’ W 000° 03.313’
Castle Hill	*Urtica* - Grassland	England	N 50° 50.473’ W 000° 04.400’
Lullington Heath	Fabaceae - Heathland	England	N 50° 47.690' E 000° 11.680'

Appendix II – Mitochondrial COI sequence and protein translation

Consensus sequence for 34 *Aphrodes* 658 bp mitochondrial cytochrome oxidase subunit I gene (COI) haplotype sequences and the corresponding protein translations below. Translated protein abbreviations correspond to the protein codes for the invertebrate mitochondrial code (identified using SEQUENCHER version 4.9 (Gene Codes Corporation)).

```
#1      AACTTTGTAT TTCATTTTTG GGTTATGGTC TGGTATATTG
#1        T   L  Y    F   I   F   G    L  W   S   G   M   L

#41     GGTATGATRC TTAGATTTAT TATTCGTATT GAACTTTCAC
#41     G   M   M   L   S   F   I   I   R   I   E   L  S   Q

 #81    AACCAGGTTC ATTTTTGGGG AATGACCAAA TTTATAATGT
 #81      P   G   S   F   L   G   N   D   Q   I   Y   N   V

 #121   AGTTGTTACT TCTCATGCAT TTGTAATGAT TTTTTTTATA
 #121    V   V   T   S   H   A   F   V   M   I   F   F   M

 #161   GTTATGCCTA TTATAATTGG CGGGTTTGGA AATTGACTTG
 #161   V   M   P   I   M   I   G   G   F   G   N   W   L   V

 #201   TTCCAYTAAT ATTAGGTGCT CCTGATATAG CATTTCCTCG
#201      P   L   M   L   G   A   P   D   M   A   F   P   R

#241    AATAAATAAT ATGAGATTTT GGTTATTGCC TCCATCATTA
 #241    M   N   N   M   S   F   W   L   L   P   P   S   L

 #281   ATTTTATTAT TAATGAGATC AATTGTTGAA ATAGGTTCTG
 #281   I   L   L   L   M   S   S   I   V   E   M   G   S   G

 #321   GTACTGGTTG AACTGTTTAT CCACCCCTAT CTTCTAATAT
 #321     T   G   W   T   V   Y   P   P   L   S   S   N   I

 #361   TTCTCACTCT GGTCCTAGAG TAGATTTAAC TATTTTTTCT
 #361    S   H   S   G   P   S   V   D   L   T   I   F   S

 #401    TTACATTTRG CTGGTATTTC ATCTATTCTT GGGGCTATTA
 #401   L   H   L   A   G   I   S   S   I   L   G   A   I   N

 #441   ATTTTATTTC AACTATTATT AATATGCGAA TTCAGGGCAT
 #441     F   I   S   T   I   I   N   M   R   I   Q   G   M

#481    AAAGATAGAT AAAATACCTT TATTTGTTTG ATCAGTTTTT
 #481    K   M   D   K   M   P   L   F   V   W   S   V   F

 #521    GTTACGGCTA TTCTTTTAAT GCTTTCATTA CCTGTTTTAG
#521    V   T   A   I   L   L   M   L   S   L   P   V   L   A

 #561   CAGGAGCTAT TACTATATTA TTAACAGATC GTAATTTAAA
 #561      G   A   I   T   M   L   L   T   D   R   N   L   N

 #601   TACAACYTTT TTTGACCCTT CAGGTGGAGG GGATCCTATT
 #601     T   T   F   F   D   P   S   G   G   G   D   P   I

 #641   TTGTATCAAC ATTTATTT
 #641   L   Y   Q   H   L   F
```

Chapter 4:

Ecological adaptation and early-stage sympatric divergence of *Aphrodes makarovi* (Hemiptera, Cicadellidae) into inland and estuarine lineages? Evidence from morphology, mtDNA and AFLP markers.

4.1. Abstract

Adaptation to different host plants is thought to be a main reason for the high diversity of phytophagous insects observed in nature. Alternatively, selection pressure differences along environmental gradients may drive diversification and speciation. The leafhopper *Aphrodes makarovi* is found inhabiting inland and estuarine habitats on two different primary host plants (*Urtica* sp. and *Atriplex portulacoides* respectively), likely to be associated (in the case of estuarine populations) with strong selection pressures (salt tolerance and tidal inundation). A genome scan of 495 amplified fragment length polymorphism (AFLP) markers was carried out to examine the genetic structuring of nine UK populations of adult *A. makarovi,* collected from inland and estuarine habitats (including, at two sites, adjacent inland/estuarine populations). Morphological analyses suggest differentiation in the degree of banding pattern and pigmentation intensity on the head and thorax of inland and estuarine specimens. Significant nuclear genetic structure associated with habitat type was identified, compared to that explained by geographic locality. Phylogenetic analyses resulted in near monophyletic habitat associated lineages, suggesting the importance of habitat type in structuring the genetic diversity of this species. Mitochondrial DNA sequence data however, revealed no structure relating to habitat or geographic locality. The lack of fixed divergent AFLP loci or significant mitochondrial DNA structure suggests that *A. makarovi* populations have diverged very recently and are in the earliest stages of ecotype formation. Lack of significant isolation-by-distance pattern, and estimates of appreciable levels of gene flow, argues that the effects of selection between inland and estuarine *A. makarovi* may be diluted due to appreciable gene flow among populations of this mobile species. Mitochondrial DNA data revealed a 'star shaped' network indicative of a recent population expansion or a selective sweep. Initial divergence of inland and estuarine *A. makarovi* populations in allopatry cannot be ruled out, however, the pattern observed here does not suggest that populations have experienced a long period of vicariance. Additional work exploring the historical genetic structuring of *A. makarovi* (including populations from continental Europe), the degree of host/habitat fidelity, fitness costs associated with each habitat/host plant type, intrinsic genetic incompatibilities, hybrid fitness are needed to further elucidate the significance of these findings and likelihood of ecotype formation into inland and estuarine *A. makarovi* lineages.

4.2. Introduction

Sympatric speciation requires the evolution of reproductive isolation in sympatry (natural selection on different traits facilitating the direct or indirect evolution of reproductive isolation), despite limited gene flow between populations (Rice 1987; Dieckmann & Doebeli 1999; Kondrashov & Kondrashov 1999; Tregenza & Butlin 1999; Via 2001). Divergence can initiate when different populations evolve adaptations to different niches with varying ecological conditions (divergent natural selection), that provides a barrier to gene flow and allows long-term coexistence (Coyne & Orr 2004). Such models of speciation involving disruptive natural selection for specialisation (selection against intermediates) require genetically determined niche preference (such as feeding, oviposition or mating), niche adaptation (trade-offs in fitness, alleles that increase ability to survive in one environment but reduce survival in others), and assortative mating (preferential mating with individuals inhabiting the same niche) (Rice 1987; Coyne & Orr 2004). Reproductive isolation can evolve in sympatry either due to direct disruptive/divergent selection affecting loci influencing habitat choice or indirectly due to pleiotropic effects (via disruptive selection on other traits) (reviewed in Via 2001).

Ecological speciation can occur in sympatry or allopatry (Schulter 2001; Rundle & Nosil 2005) and thus the historical genetic patterns of species are of great importance to rule out divergence in allopatry before inferring sympatric origins. Under the allopatric scenario (Mayr 1963), genetic differentiation initiates in allopatry (due to ecologically driven natural selection), and reproductive isolation may then follow as a consequence of the accumulation of genetic differentiation between populations showing differential fitness in their respective habitats (Schluter 2001; Rundle & Nosil 2005). Speciation is then completed once populations come into secondary contact and reproductive isolation can be finalised through reinforcement, although other mechanisms do exist (Schluter 2001). Divergence and speciation in allopatry can also occur due to non-ecological factors such as genetic drift, demographic events, hybridisation and polyploidisation (reviewed in Coyne & Orr 2004).

4.2.1. Ecological adaptation, ecotypes and host races

Adaptation to different host plants is thought to be a main reason for the high diversity of phytophagous insects observed in nature (Berlocher & Feder 2002; Funk *et al.* 2002), whereby ecologically driven selection leads to reproductive isolation (Via 2001; Drès & Mallet 2002). This would permit closely related species to co-exist in the same habitat on different host plants (host races). A genetically determined host plant preference and assortative mating, with respect to host plant, are required for host-mediated speciation to occur, although host race formation is often viewed as a transitionary step towards sympatric speciation (Berlocher & Feder 2002). Host switching increases the potential for reproductive isolation (Mitter *et al.* 1988) and insect herbivores are the most widely studied models for this form of speciation.

Many studies suggest that species with restricted host ranges and specialist feeding habits are more likely to promote host race formation (Emilianov *et al.* 2004), due to the patchy distribution of host plants, facilitating isolation and reduced gene flow (reviewed in Futuyma & Moreno 1988). Alternatively, for generalist feeders and widespread insects, local adaptation through host-switching may be more likely to arise due to the high diversity of potential host types (Janz & Nylin 2008; Nylin & Walhberg 2008; Nylin & Janz 2009). Adaptation along an environmental abiotic gradient, such as salinity or altitude, could also explain the diversification of a widespread species (Bonin *et al.* 2006; Mallet 2008). Reproductive isolation could evolve due to genetic drift, if populations are geographically isolated, or due to divergent natural selection on host plant/ecological type, or a combination of both (Schulter 2001). Many examples in nature suggest that speciation can be initiated through ecologically driven natural selection but that this does not necessarily guarantee the evolution into full species (Nosil *et al.* 2009a).

4.2.2. *Aphrodes makarovi*

The phenomenon of ecotype formation is examined here, primarily by exploring the genetic diversity and population structure in the phytophagous leafhopper *A. makarovi* occurring in inland and estuarine habitats. Morphological differences associated with these habitats were also explored. An understanding of the genetic variation and

population structuring within *A. makarovi* was needed in order to gain insights into the potential of ecological (and/or other) processes that may be involved in shaping the genetic patterns in this species.

Leafhoppers oviposit eggs directly into host plant tissues and emergence and development times of nymphs are likely to be linked with host phenology (Dietrich 1999). *Aphrodes* leafhoppers (Hemiptera; Auchenorrhyncha; Cicadellidae) are univoltine and overwinter as eggs (Nickel 2003) and adults can be found from June-September. All four *Aphrodes* species are morphologically similar and rely on male vibrational mating signals for species recognition (Hamilton 1975; Le Quesne 1988; Tishechkin 1998; Virant-Doberlet *et al.* 2005; Chapter 3).

Aphrodes makarovi is an abundant species and occurs in most terrestrial habitats due to the ubiquitous nature of its host plants (*Urtica* (Urticaceae)*, Taraxacum* and *Cirsium* sp. (Asteraceae), Tishechkin 1998; Nickel 2003; Biedermann & Niedringhaus 2004). Recently, *Aphrodes makarovi* has been found inhabiting estuarine, saltmarsh environments (Chapter 3) that vary from inland habitats in abiotic factors such as salinity and support different communities of host plants. *Aphrodes makarovi* has been identified on the host *Atriplex portulacoides* (Amaranthaceae, Subfamily Chenopodioideae), often coexisting in sympatry with the related saltmarsh species *A. aestuarina* (Chapter 3) (previously thought to be a possible subspecies or ecological form of *A. makarovi,* Kirby 1992; Nickel 2003).

Originally, all salt-marsh inhabiting *Aphrodes* were classified as *A. aestuarina* (before knowledge that *A. aestuarina* and *A. makarovi* are found in sympatry), and were characterised by a halophilous host plant/habitat association and morphological characteristics such as lighter coloration, lack of banding pattern on the head/thorax and a narrower body shape (Edwards 1908). As the specimens studied at the *A. aestuarina* syntype location (and museum specimens from the syntype series, Chapter 2 – Bluemel *et al.* 2011) were in fact two species (Chapter 3) it is not surprising, that no major morphological differences were found between *A. aestuarina* and *A. makarovi* (Biedermann & Niedringhaus, 2004). However, recent evidence (Chapter 3) confirmed an overlap in morphology between these species (species identification verified with mating signal and molecular data). In females, estuarine adapted *A. makarovi* does show a longer and thinner body shape that is more similar to that of *A. aestuarina* than inland

A. makarovi (Chapter 3), although with a considerable degree of overlap. Notable differences in the banding/colouration pattern on the head and thorax between populations of *A. makarovi* utilising different hosts has been observed. While much variation exists within terrestrial populations of *A. makarovi,* a fairly consistent lighter and non-banded morphotype has been observed in estuarine *A. makarovi* populations, which is possibly more similar to that of *A. aestuarina* than its inland counterpart.

Aphrodes makarovi vibrational mating signals do not differ between inland and estuarine habitat types and much of the variation associated with song elements and signal composition lies within populations (Virant-Doberlet, unpublished data). *Aphrodes* species also seem to signal on any green plant (Virant-Doberlet, unpublished data). This suggests that different habitat associated populations are not reproductively isolated through mating signal recognition and this could indicate that host plant type does not significantly restrict signal-receiver transmission, although this has not been directly tested. Very little is known about the ecology of *Aphrodes* in relation to host/habitat specificity, dispersal and mobility, however they often show a patchy distribution at saltmarsh and grassland sites (personal observation). Dispersal is likely to be linked to female movement because females are the egg laying sex. Females have been identified in the field on the same host plant (*Urtica* sp.) for up to a week and due to this limited movement, it is likely that feeding occurs on the host that females are found on (Virant-Doberlet, personal observation).

Adaptation to an extreme environment may be as important in promoting divergence as adaptation to novel hosts for this species. Estuarine habitats supporting both *A. makarovi* and *A. aestuarina* often become fully submerged during tidal progressions (personal observation) suggesting that specific adaptations may be required to survive in such harsh environments. Additional differences in abiotic factors (salinity), characteristic of intertidal habitats, could present a major driving force for local adaptation. Temporal isolation has been known to promote divergence due to differences in emergence times of host plants (Coyne & Orr 2004). The role of predation may also be an important driving force of adaptation as suggested for other study systems (Wilding *et al.* 2001; Rosenblum 2006; Nosil & Crespi 2006). It is likely that a number of factors could contribute to ecological divergence, and multiple factors such as diet and habitat may increase the likelihood of speciation compared to just one factor alone (Nosil *et al.* 2009a). To date, no studies have been carried out to examine

genetic structuring and morphological differentiation within this species with respect to habitat type (and associated host plants) of *A. makarovi*.

4.2.3. Population genetics and identifying signatures of selection

A fundamental element of population genetics is that genome-wide effects (e.g. genetic drift and gene flow) affect all loci across the genome and allow reliable inference of phylogenetic history and population demography. On the other hand, locus-specific effects (selection, mutation and recombination) act on key genetic regions that are important for fitness and adaptation (Black *et al.* 2001; Luikart *et al.* 2003). To identify key genomic regions or specific genes associated with adaptive divergence and speciation is challenging and a key focus for many evolutionary biology studies (Luikart *et al.* 2003; Stinchcombe & Hoekstra 2008). Genome scans and associated analyses carried out to address this problem attempt to isolate specific loci that show high among population differentiation (F_{ST} values), known as outlier loci, when compared to F_{ST} estimates for neutral loci present across the genome. These outlier loci, predicted to be directly under selection or closely linked to a selected gene (Schlötterer 2003; Nosil *et al.* 2009b), are generated due to the increased frequency of locally advantageous alleles expected to result from natural selection (Black *et al,* 2001; Storz 2005).

The allele frequency differences seen between outlier and neutral loci suggests that gene exchange may be more restricted to some areas of the genome than others and is a pattern that could be seen as a signature of sympatric speciation, in contrast to allopatric or physical isolation which may induce genome-wide barriers to gene flow (Wilding *et al.* 2001). Allopatric divergence showing a uniform differentiation across a genome can be sustained following secondary contact and facilitated by the accumulation of genetic incompatibility at many loci (Barton & Hewitt 1985); however, it may be reduced by introgression. It can however be difficult to differentiate between the primary and secondary origins of clines (gradients in morphology or gene frequency) (Barton & Hewitt 1985; Bierne *et al.* 2011). The well supported (direct or indirect) inference of ecological divergent selection through identification of outlier loci has been challenged (Bierne *et al.* 2011). Pre-existing intrinsic (environment independent) genetic incompatibilities (known as tension zones, Barton & Hewitt 1985) that can become trapped by natural barriers (due to ecological selection) may be responsible for the

occurrence of high F_{ST} outlier loci instead of local adaptation. Genome scans identifying outlier loci may only identify the position of intrinsic barriers to gene flow rather than explaining their persistence (Barton & Hewitt 1985; Bierne *et al.* 2011).

Genome scans involve genotyping a large number of loci from across the genome, which are required to decipher genome-wide effects from locus-specific effects. This has become relatively easy and applicable to non-model species, using low cost techniques such as amplified fragment length polymorphism (AFLP) (Vos *et al.* 1995; Bensch & Åkesson 2005). However, without background knowledge of candidate genes or genetic linkage maps based on quantitative trait loci (QTL) the specific functionality of outlier loci will remain unknown in non-model organisms. Alternatively, studies have attempted to correlate highly differentiated genomic regions with ecological factors, such as host races in insects (Emelianov *et al.* 2004; Egan *et al.* 2008), different ecotypes in fish (Campbell & Bernatchez 2004), environmental gradients, such as altitude in the common frog (Bonin *et al.* 2006) and different morphological ecotypes in snails (Wilding *et al.* 2001). These are examples of the formation of genetic-environmental associations (Hedrick *et al.* 1976), but although these are relatively common, causal reasoning for such associations are rare in experimental literature (Bierne *et al.* 2011). Although the proposed direct inference of ecological divergent selection through identification of outlier loci may be less parsimonious than alternative explanations (to which future consideration should be given, Bierne *et al.* 2011), genome scans remain useful for revealing patterns of genetic diversity, differentiation and the presence of significant genetic breaks, whether the genetic markers are mapped or unmapped.

4.2.4. Aims

Because host plant or environmental associations may represent the outcome of local adaptation to distinct habitat types, the genome of *A. makarovi* was explored to identify signatures of genetic differentiation among populations utilising different host plants in different ecological habitats. Morphological analysis of banding pattern and pigmentation was also carried out to test the hypothesis that *A. makarovi* populations are morphologically and genetically divergent. Phenotypic and genetic diversity was investigated to address the following objectives: i) to identify if colour intensity and banding pattern of particular morphs can be associated with habitat type; ii) to assess the level of genetic differentiation among populations using mitochondrial DNA (mtDNA) and AFLP markers; iii) to determine the extent of gene flow between populations; iv) identify whether any genetic differentiation can be attributed to habitat type; v) reveal outlier loci (if any) that show high levels of genetic differentiation between habitat associated populations based on neutrality. The aim of this study was to identify potential candidate regions and assess if *A. makarovi* presents a suitable model system for further exploration into possible ecological factors driving local adaptation and ecotype formation.

Based on observations of phenotypic differences between inland *A. makarovi* and estuarine *A. makarovi* and similarities between both estuarine *Aphrodes* species (Chapter 3), the hypothesis that estuarine morphs of both species (*A. makarovi* and *A. aestuarina*) will be more similar to each other with regards to banding pattern and pigmentation than to inland *A. makarovi* populations, was tested. Mitochondrial DNA and AFLP marker analyses are used to test the hypothesis that genetic differentiation does exist between populations inhabiting different habitats in multiple locations and that habitat associated genetic differentiation is greater than that seen between same habitat type populations in different geographic locations. It is also predicted that different populations inhabiting alternative habitat types will possess nuclear AFLP markers showing high levels of genetic differentiation (high F_{ST} values) based on neutrality, which may indicate that positive selection is acting on populations of *A. makarovi* utilising different hosts in very different habitats.

4.3. Materials and Methods

4.3.1. Sample collection, species identification and DNA extraction

Specimens were collected across England and Wales from June to August during the years 2005 to 2009, using a converted leaf blower (D-vac suction sampler, Electrolux, BVM 250). Sampling sites were chosen based on habitat type and host plant. At inland sites *A. makarovi* were collected from patches of *Urtica* and/or *Cirsium* sp., and on estuaries they were sampled from *Atriplex portulacoides* (Sea Purslane). Estuarine habitats are often fringed by species of *Urtica* or *Cirsium* sp. and specimens were collected from contiguous habitats where possible. Specimens were stored in absolute ethanol at -80 °C either directly from the field or after bio-acoustic signals had been recorded (Chapter 3). Legs of *A. makarovi* specimens were dissected and genomic DNA was extracted using DNeasy Tissue Kits (Qiagen).

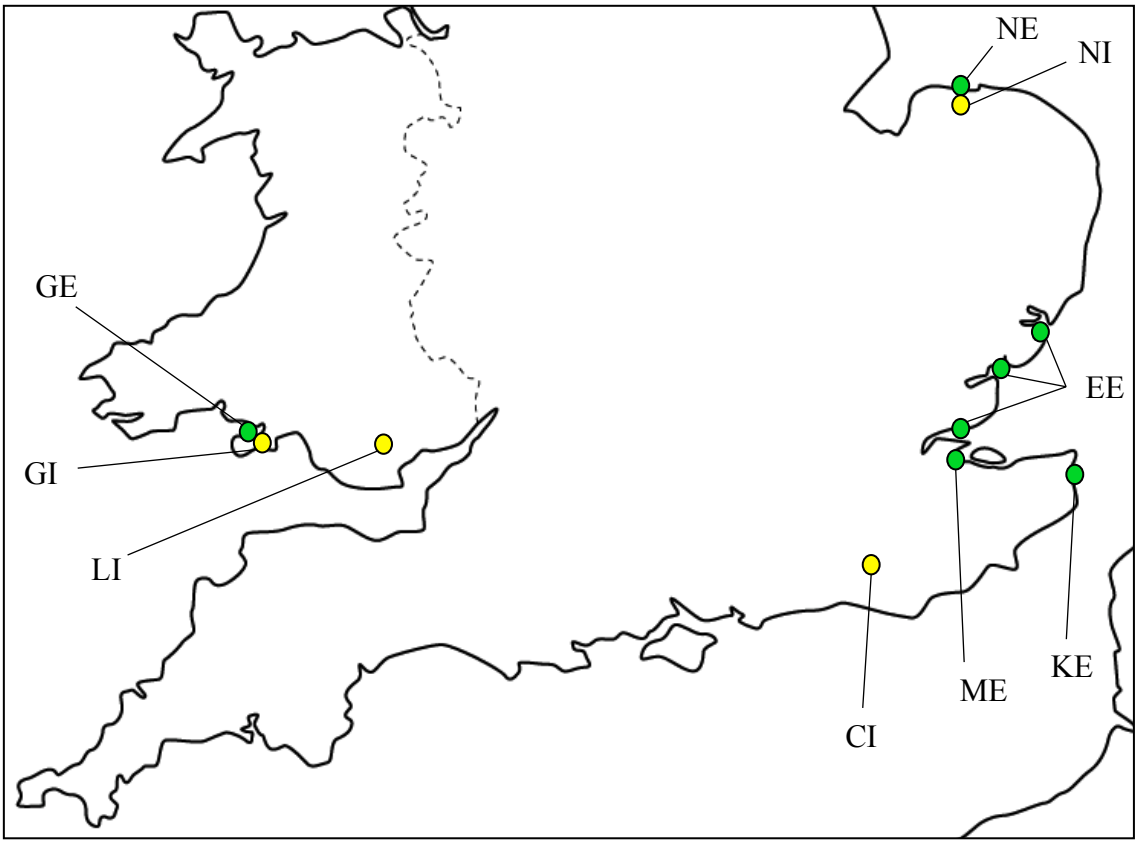


Figure 4.1. Sampling locations for *Aphrodes makarovi* in Wales and England. Green circles indicate estuarine populations utilising the host plant *Atriplex portulacoides* and yellow circles represent inland sites utilising *Urtica/Cirsium* sp. host plants. See Table 4.1 for sampling site abbreviations. Estuarine habitats (~E) and inland habitats (~I). See 4.8, Appendix I for GPS coordinates.

A sample of 171 *A. makarovi* specimens were used for AFLP analyses (with an additional five positive control repeats), collected from four inland sites (Castle Hill, Lisvane, Gower and Norfolk) and five estuarine sites (Kent, Medway, Essex, Gower and Norfolk) (Fig. 4.1). Gower and Norfolk are coupled sites where inland and estuarine habitats are naturally found adjacent to each other. Specimens were identified using either mitochondrial cytochrome oxidase subunit I gene (COI) or vibrational mating signals, or both (Chapter 3). Introgressed hybrid specimens have been identified in Medway (Chapter 3) and individuals included from this location in this study were confirmed to be pure *A. makarovi* in AFLP analyses carried out in Chapter 5. Introgressed hybrids were identified using vibrational signals, mtDNA and AFLP marker analyses (Chapter 5). The samples size for four Essex sites was low and specimens were pooled for the whole region. Samples sizes were from 8-48 individuals per population for AFLP analyses (Table 4.1). Geographical coordinates (GPS) can be found in section 4.8, Appendix I.

Table 4.1. The nine sampled populations, including the genus of host plant present and habitat type, number of male and female (M/F) specimens used for amplified fragment length polymorphism (AFLP) analysis. (~E) = estuarine habitats, (~I) = inland habitats.

Host / habitat	**Population**	**Abbreviation**	***n* AFLP (M/F)**
Atriplex / estuarine	Kent	KE	8 (6/2)
Atriplex / estuarine	Essex	EE	15 (8/7)
Atriplex / estuarine	Gower	GE	20 (10/10)
Atriplex / estuarine	Norfolk	NE	20 (13/7)
Atriplex / estuarine	Medway	ME	49 (30/19)
Urtica /inland	Lisvane	LI	10 (8/2)
Urtica /inland	Gower	GI	19 (7/12)
Urtica /inland	Norfolk	NI	20 (11/9)
Urtica /inland	Castle Hill	CI	10 (5/5)

4.3.2. Morphological analysis of banding pattern

To measure the intensity and degree of pigmentation and banding pattern across the head and thorax of *A. makarovi* specimens, greyscale intensity image analysis was used. Using the mean greyscale value and standard deviation (SD) across all pixels (for a standardised area measured in all individuals), differentiation between lighter colour morphs (higher greyscale mean) from darker morphs (lower grey scale mean) with or without banding (high SD or low SD, respectively) was achieved. The measure of SD cannot distinguish the type of pattern/contrast (striped or spotted) but it can give a reliable measure of the degree of contrast or variation around the mean intensity. Similarly the overall mean grey scale value cannot differentiate between pigmentation consisting of only midrange intensity values or more variable values, but used in conjunction with the SD (Mean:SD ratio) it gives an overall measure of pigmentation intensity and degree of banding.

A total of 47 females (15 estuarine and 32 inland) and 64 males (31 estuarine and 33 inland) *A. makarovi* were analysed. *Aphrodes aestuarina* were also included in morphological banding pattern analyses to test whether estuarine *A. makarovi* shows greater morphological similarities to *A. aestuarina* than to inland *A. makarovi*. To achieve this, 15 *A. aestuarina* females (collected from Sussex) and 26 *A. aestuarina* males (collected from Sussex and Norfolk) were also analysed (identified using COI and vibrational signals (Chapter 3) and AFLP analyses (Chapter 5)). Specimens were photographed using a JVC KY F70 3CCD digital camera mounted on a Leica M28 stereo microscope (magnification 1.6x, aperture setting 2) with a Planapo chromatic 1x lens attached. AUTO-MONTAGE PRO version 5.0 (Synoptics) imaging software was used to capture images. Adult specimens were imaged whole (legs removed) mounted in a trough formed with white-tack (to ensure horizontal positioning), in an excavated glass block and submerged in ethanol. The cold light source intensity and position (Schott KL 1500, intensity setting 3) were consistent throughout. All images were saved as bitmap format and exported into IMAGEJ version 1.40e (Rasband 2007). A standardised elliptical area of the head and thorax (Fig 4.2) was selected and the histogram option was used to obtain the mean pixel greyscale value (0-255) and SD. Imaging and grey scale intensity data collections were carried out by R. McFadyen (unpublished data).

Principle Components Analysis (PCA) was carried out in MINITAB version 16 using correlation standardised method. PCA was carried out for male and female specimens separately due to size differences between males and females of each species (females are typically larger than males).

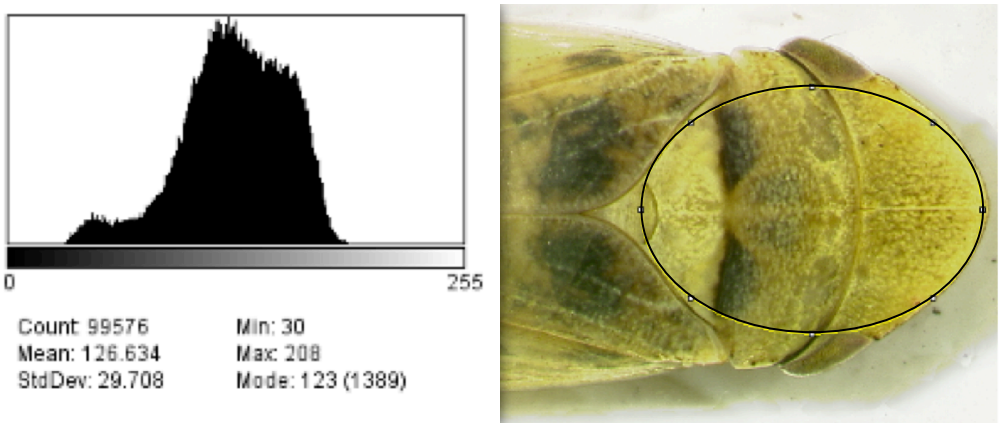


Figure 4.2. Example of grey scale intensity histogram (left) of an *Aphrodes makarovi* specimen illustrating the elliptical area used to measure the mean and standard deviation of grey scale intensity (right).

4.3.3. Mitochondrial sequencing and data analysis

The mitochondrial COI sequence dataset used in this chapter was obtained using methods described in Chapter 3 (GenBank Accession numbers; FR727167 – FR727170, HE587029 – HE587041). A total of 186 *A. makarovi* specimens from ten populations were sequenced (Table 4.1, including a population from Devon (DE) that was not included in AFLP analyses, see Chapter 3 for details). The 710 base pair (bp) COI sequences were amplified using polymerase chain reaction (PCR) and universal invertebrate primers LCO1490 and HCO2198 (Folmer *et al.*1994). Products of PCR reactions were sequenced directly using the original PCR primers (Chapter 3) using BigDye (version 3.1) sequencing chemistry in both forward and reverse directions. Reactions were run by the Cardiff University Molecular Biology Support Unit and analysed on an Applied Biosystems 3130x1 Genetic Analyser. Sequences were aligned and unique haplotypes were identified using SEQUENCHER version 4.9 (Gene Codes).

To evaluate how mtDNA variation was distributed with respect to geographic region and habitat association, a hierarchical analysis of molecular variance (AMOVA) in ARLEQUIN version 3.5.1.2 (Excoffier & Lischer 2010) was performed. Populations were assembled either specifying no population structure, regional structure (five or six groups) or structure relating to habitat (two groups). For regional structure populations were grouped into six regions (Wales (GE, GI and LI), Devon (DE), Medway/Essex (ME and EE), Norfolk (NI and NE), Sussex (CI) and Kent (KE)) (Table 4.1 for location abbreviations).

A median-joining network was constructed using NETWORK version 4.6 (Bandelt *et al.* 1999) to depict relationships between mtDNA haplotypes associated with different habitats. The parameter epsilon value was set to zero and character states were assigned equal weights.

4.3.4. AFLP protocol

AFLP analysis was performed according to Vos *et al.* (1995) with modifications described below. Initial AFLP tests, using the more typical four base/six base cutter enzyme combinations (*Taq*I and *Eco*RI, blunt and sticky end, respectively) produced AFLP profiles that were difficult to score due to the excessive numbers of bands obtained. The enzymes used in the protocol were *Pst*I and *Eco*RI (which are two six base recognition enzymes, both with sticky ends), with the aim of reducing the overall number of fragments obtained for each primer combination (Hawthorne 2001). Approximately 50ng of genomic DNA from each specimen was digested with *Eco*RI (2.5U, New England Biolabs (NEB)) and *Pst*I (2.5U, NEB), Buffer 3 (1x, NEB) adjusted with water to a final volume of 20µl and incubated for 2 h at 37°C. Adapters were ligated using 5pm/µl of double-stranded *Eco*RI adapter (5'-CTCGTAGACTGCGTACC – 3' and 5'- AATTGGTACGCAGTC – 3', Sigma), 50pm/µl of double-stranded *Pst*I adapter (5'- CTCGTAGACTGCGTACATGCA – 3' and 5'- TGTACGCAGTCTAC – 3', Sigma), ATP (1mM, Roche), T4 DNA ligase buffer (1x, NEB), and T4 DNA ligase (0.7U, NEB), adjusted to a final volume of 5µl with water, then added to the double digested genomic DNA (total volume 25µl) and incubated at 37°C for 4 h.

The digestion-ligation template was diluted (1:10) with low TE buffer (2M Tris-HCl (pH7.5), 0.5M EDTA (pH8)). Pre-selective PCRs contained 2.5µl diluted template DNA, GoTaq mater mix (3.75µl, Promega), pre-selective *Eco*RI primer (5`-GACTGCGTACCAATTC**A** – 3') and *Pst*I primer (5'- GACTGCGTACATGCAG**A** – 3') (each 2.5ng/µl, Sigma), adjusted to a total reaction volume of 10µl with water. Amplification initiated with a denaturing step at 95°C for 2 min, followed by 30 cycles at 95°C for 30 s, 56°C for 30 s, 72°C for 60 s. PCR products (2.5µl) were run on a 1.5% agarose electrophoresis gel. The pre-selective template was diluted (1:10) with low TE buffer for use in the selective amplifications.

Selective PCR reactions contained 2.5µl diluted pre-amplified template DNA, GoTaq mater mix (5µl, Promega) and *Eco*RI fluorescent-labelled primer (2.5ng/µl) and *Pst*I (15ng/µl) primer (Sigma), adjusted to a total reaction volume of 10µl with water. Three primer combinations were used each with three overhanging nucleotides at the 3' end:
PAAA/E44 (5'- GACTGCGTACATGCAG**AAA** – 3' / 5' -6Fam-GACTGCGTACCAATTC**ATC** – 3');
PAAG/E42 (5'- GACTGCGTACATGCAG**AAG** – 3' / 5' -6Fam-GACTGCGTACCAATTC**AGT** – 3');

PAAT/E35 (5'- GACTGCGTACATGCAG**AAT** – 3' / 5' -6Fam-GACTGCGTACCAATTC**ACA** – 3'). The touchdown thermal cycling programme initiated with a denaturing step at 95°C for 2 min, followed by 13 cycles of 95°C for 30 s, 65°C* for 30 s (*-0.7°C each cycle), 72°C for 60 s followed by 23 cycles of 94°C for 30 s, 56°C for 30 s, 72°C for 60 s. Amplified products (1µl) were each added to formamide (10µl, Applied Biosystems) and GeneScan ROX-500 size standard (0.25µl, Applied Biosystems). Reactions were run by the Cardiff University Molecular Biology Support Unit and analysed on an Applied Biosystems 3130x1 fragment analyser.

4.3.5. AFLP scoring and error rates

Electropherogram trace files were imported into GENEMARKER version 1.95 (SoftGenetics). GeneScan ROX 500 size standards were applied to the project and manually checked for quality and edited where required. Poor quality profiles (failed amplification) were removed from subsequent analyses. All peaks above 150 rfu (peak height identified as a suitable background noise threshold) and between 50-500 bp were scored using GENEMARKER. A panel was created automatically using all samples. Bin positions were manually checked to identify incorrect bin positioning and low quality or noise peaks (irregular shape or pull-ups). Overlapping bin positions were deleted from the data set to avoid ambiguous scoring due to possible size homoplasmy of co-migrating fragments (Vekemans *et al.* 2002). PCR negatives were checked for possible contaminants and any peaks above the background noise threshold were deleted from the respective primer combination.

AFLPSCORE version 1.4b (Whitlock *et al.* 2008) was used to identify thresholds (relating to average locus peak height and relative peak height across all loci) that resulted in acceptable mismatch error rate (< 5%) but maximised the number of AFLP markers retained for further analyses. Mismatch error rates (Bonin *et al.* 2004), based on five repeated genotype profiles were calculated using the data filtering option, a locus selection threshold of 400 rfu (18% of the total mean normalised peak height across all loci) and a relative phenotype calling threshold of 150 rfu (7% of the total mean normalised peak height across all loci). A binary matrix of retained AFLP markers was created in AFLPSCORE for the three primer combinations and a subset were compared to the original electropherograms to check for computational copying errors.

4.3.6. AFLP data analysis

4.3.6.1. Genetic diversity

Descriptive statistics were calculated assuming Hardy-Weinberg equilibrium (HWE) using AFLP-SURV version 1.0 (Vekemans *et al.* 2002) as the percentage of polymorphic loci (*P*) and the average expected heterozygosity (Nei's gene diversity) for each population. Allele frequency estimates were obtained assuming HWE, using a Bayesian method with non-uniform prior distribution of allele frequencies (Zhivotovsky 1999) and genetic diversity statistics were computed following the approach of Lynch & Milligan (1994) using 10,000 bootstraps.

Problems arise with traditional methods to evaluate genetic differentiation among populations using dominant markers compared with codominant markers. Therefore the hierarchical Bayesian approach implemented in HICKORY version 1.1 (Holsinger *et al.* 2002) was used, which does not assume any prior knowledge of the degree of inbreeding within populations. The $\theta^{(\Pi)}$ parameter, analogous to F_{ST} (based on Weir & Cockerham's approach, 1984), was estimated for each pairwise population comparison. The method most applicable to dominant data is the f-free model that does not attempt to estimate F_{IS} due to its unreliability when estimated using other models (Holsinger & Wallace 2004). Values derived for the deviance information criterion (DIC) calculated from the full model (which attempts to estimate F_{IS}), $\theta^{B} = 0$ (model that assumes no differentiation among populations) and the $F_{IS} = 0$ model (assuming no inbreeding among populations) were used to identify how well each model fitted the data (a better fit results in smaller DIC values).

Attempts to compute both F_{IS} and F_{ST} estimates based on the Approximate Bayesian Computation method (ABC) as implemented in ABC4F (Foll *et al.* 2008) were also undertaken. This method aims to avoid two biases identified in the original method proposed by Holsinger *et al.* (2002). These include the use of non-informative flat priors on the ancestral allele frequencies for estimating F_{IS} and the bias surrounding the selection of AFLP markers for analysis (Foll *et al.* 2008). Sample sizes of 5000 and 50,000 were tested and the ascertainment bias for hidden loci was set to 2 (loci only appearing in just one individual were previously removed from the data set) and for

fixed loci, 0, as no fixed loci were identified. All other parameter settings were set to default and the acceptance rates were set to 0.001.

4.3.6.2. Population structure

Population structure was inferred using a Bayesian model-based clustering method as implemented in STRUCTURE version 2.2 (Pritchard *et al.* 2000), which has been adapted to accommodate dominant data (Falush *et al.* 2007). STRUCTURE assigns individuals into genetic clusters (*K*) using multi-locus data, without using any information regarding population origin. The admixture model was used to estimate the proportion of each individual's genome that has descended from each source population (the proportion of ancestry). Ten independent runs for each value of *K* ranging from one to ten were performed with 1,000,000 Markov Chain Monte Carlo (MCMC) repetitions, following a burn-in of 250,000 and run on CONDOR (Litzkow *et al.* 1998) computational facility (ARCCA, Cardiff University). Runs were performed using both correlated and uncorrelated allele frequencies (Pritchard *et al.* 2000; Falush *et al.* 2003). The uncorrelated model assumes that different population allele frequencies are independent (Pritchard *et al.* 2000) but does not account for the fact that the allele frequencies may be similar in scenarios of subtle population structure. The correlated allele frequency model is more flexible, allowing for populations to have experienced different amounts of drift away from ancestral allele frequencies. Due to the prior assumption that allele frequencies in different populations are correlated, this model is better for detecting subtle admixture (Falush *et al.* 2003).

Analyses were carried out using all *A. makarovi* specimens from both inland and estuarine habitats to examine overall structure of all sampled populations. All inland and estuarine populations were also analysed in two separate runs to identify population structure within each habitat type. Site-specific comparisons of inland and estuarine populations were carried out for Gower and Norfolk separately to identify structure of different habitat associated populations in the absence of geographic isolation. The true number of clusters was deciphered using both the maximal log probability of the data, $\Pr(X|K)$ reported by STRUCTURE and the ΔK method of Evanno *et al.* (2005) that calculates the rate of change in the log probability of data between successive *K*-values.

Results were visualised using individual assignment values averaged over the 10 replicate runs.

Principal coordinate analysis (PCA) was carried out in GENALEX 6.4 (Peakall & Smouse 2006) using the standardised distance method. Analysis was carried out using Euclidian distance matrix calculated in ARLEQUIN. PCA was carried out comparing all sampled inland and estuarine populations and comparing only Gower and Norfolk coupled inland and estuarine sites separately.

AFLPOP version 1.1 (Duchesne & Bernatchez 2002) was used to give insights relating to the degree of gene flow between inland and estuarine *A. makarovi* populations. Re-allocation of reference genotypes (individuals from either habitat type) was performed, where each genotype is removed from the computation of frequencies within its known population and assigned as an unknown, to identify the assignment success of individuals back to their populations of origin (habitat type). Ten replicated simulations were carried out and a log likelihood difference of zero was used so that all individuals were assigned to a population. The log likelihood differences were then assessed to identify individuals with similar assignment probabilities for more than one category and thus questionable assignment. The success rate and log likelihood values can be used as a measure of how well the data set can distinguish between the different habitat type populations, and thus gives an indication of the degree of recent gene flow between parental populations (Manel *et al.* 2005). Assignment tests were carried out for the whole data set, divided by habitat type and also for each locally adjacent population at the Gower and Norfolk comparing inland populations with estuarine. Assignment tests are more suited for smaller geographic scales (Bonin *et al.* 2007) due to the influence of population history and the assumption of no mutation is likely to be violated over larger geographic scales. AFLPOP was also used to simulate 1000 pure parental genotypes from reference samples (identified as individuals collected from either habitat type). The success rate of assignment of the simulated populations to the parent classes can be used as a measure of how well the data set can distinguish parental populations and thus gives an indication of the degree of gene flow and genetic differentiation between parental populations. Both simulation and re-allocation of source populations should be carried out together as the simulation results can be used as the upper bound in allocation success and the re-allocation results can be used as the lower bound (Duchesne & Bernatchez 2002). The success rate is usually higher for the simulation

compared to the re-allocation procedure due to sampling errors on allele frequencies being masked, because simulations provide a larger sample of possible genotypes (Duchesne & Bernatchez 2002).

To evaluate how genetic variation was distributed with respect to geographic region and habitat association a hierarchical analysis of molecular variance (AMOVA) in ARLEQUIN version 3.5.1.2 (Excoffier & Lischer 2010) was performed. Analysis was carried out using Euclidean distance with 1,000 permutations. AMOVA's were calculated either specifying no population structure, regional structure (four groups) or structure relating to habitat type (two groups). For regional structure populations were grouped into four geographic regions: Wales (GI, GE, LI), Norfolk (NI, NE), Essex and Kent (EE, RE, KE) and Sussex (SE) (Table 4.1 for location abbreviations).

The correlation between genetic and geographical distance was examined with a MANTEL test in ARLEQUIN version 3.5.1.2 (Excoffier & Lischer 2010), with 10,000 permutations. Slatkin's linearised F_{ST} (Slatkin 1995) was calculated and compared to the natural logarithm (ln) of average pairwise population geographic distances (according to Rousset 1997). For geographic distances between coastal populations the natural log of coastal distance was used. For all distances between inland – inland or inland – estuarine populations, straight line geographic distances (calculated as shortest distance between populations) were used.

4.3.6.3. Phylogenetic structure

Using AFLP-SURV version 1.0 (Vekemans *et al.* 2002), 10,000 bootstrap matrices of Nei's genetic distance were generated and used to create UPGMA and neighbour-joining (NJ) 50% majority rule consensus trees using the NEIGHBOUR and CONDENSE programs in PHYLIP version 3.69 (Felsenstein 2005). The related species, *A. aestuarina*, was used to root the *A. makarovi* phylogenetic trees.

4.3.6.4. Outlier detection

Identification of potential outlier loci was performed using two current population genetic methods, BAYESCAN (Foll & Gaggiotti 2008) and MCHEZA (Antao & Beaumont 2011). MCHEZA is an online workbench that accommodates dominant data, composed of the program DFDIST (Beaumont & Nichols 1996) and a graphical user interface, similar to that of the codominant version LOTISAN (Antao *et al.* 2008). Both approaches are based on the idea that loci under balancing selection exhibit lower F_{ST} values compared to the null distribution (negative outliers), whereas loci under divergent selection are expected to show higher F_{ST} values (positive outliers) (Beaumont & Nichols 1996; Beaumont & Balding 2004). Both methods allow for control of the false discovery rate (FDR), which is used for multiple hypotheses testing in statistics, to correct for multiple comparisons, which is defined as the expected proportion of false positives out of the total number of significant results (Benjamini & Hochberg 1995). This reduces the possibility of detecting false positives (committing type-I errors) (Luikart *et al.* 2003; Bonin *et al.* 2007), which can also be minimised by using a conservative significance level (Beaumont & Balding 2004; Caballero *et al.* 2008). Both BAYESCAN and DFDIST have been tested using known levels of selection in simulated data sets (Caballero *et al.* 2008; Pérez-Figueroa *et al.* 2010). BAYESCAN was shown to be more efficient than DFDIST in certain situations (Pérez-Figueroa *et al.* 2010). MCHEZA, however, includes features that aim to improve the performance of DFDIST (Antao & Beaumont 2011), such as controlling the FDR.

DFDIST (implemented in MCHEZA) is a frequency based (frequentist), F_{ST} outlier method based on distribution of summary statistics and uses a Bayesian method to estimate allele frequencies from the proportion of recessive phenotypes in the sample (Zhivotovsky 1999). Coalescent simulations are carried out to calculate the null sampling distribution under neutral expectations, to which locus-specific F_{ST} values (Wier & Cockerham 1984) for populations under comparison are compared (Beaumont & Balding 2004). MCHEZA allows for user-friendly parameterisation for estimates of the average neutral F_{ST} of the real data by removing potential loci under selection to increase reliability and allows for correction when the average simulated F_{ST} is too dissimilar to that of the real data set. This can occur with the use of a stepwise mutation model or when the number of demes is low (Antao & Beaumont 2011). The combined use of both the forced mean F_{ST} and neutral mean F_{ST} algorithms was carried out as

advised (Antao & Beaumont 2011). The critical frequency for the most common allele was set to ≥ 0.99 and to estimate allele frequencies the scale for the Zhivotovosky (1999) parameters were set to 0.25. To obtain significance values the number of simulations for generating a null distribution of F_{ST} values was set to 50,000 and four independent runs were done. Theta was set to 0.04 (a range between 0.004 – 0.1 was tested, but no differences were seen due to the lack of sensitivity of theta, also shown by Beaumont & Nichols 1996). The sample size was set to zero to allow the estimation of sample size for each locus in MCHEZA, which when tested against default options gave the most consistent results across multiple independent runs. The FDR was set to 10% as advised by the author. By plotting heterozygosity (assuming HWE) against the observed F_{ST} values for each locus, outlier loci were identified as those positioned outside the neutral expectation given by the null distribution at two significance levels (95% and 99%). All loci with > 99% significance level were identified as outliers, as using a conservative threshold should reduce the number of false positives (Beaumont & Balding 2004; Caballero *et al.* 2008).

The demographic model used in DFDIST is one that has an infinite number of islands and assumes drift-migration equilibrium and violation of the model can lead to a high false-positive rate (Foll & Gaggiotti 2008; Excoffier *et al.* 2009). BAYESCAN implements a likelihood method and is an extension of that proposed by Beaumont & Balding (2004). BAYESCAN uses a Bayesian method to directly estimate the posterior probability that each locus is subject to selection (Foll & Gaggiotti 2008). A reversible jump Markov chain Monte Carlo (RJMCMC) approach is used to estimate posterior probabilities for two models, one that included the effects of selection and the other that excludes it, to identify potential outlier loci. Population specific F_{ST} coefficients are estimated which gives BAYESCAN the advantage by allowing consideration of different demographic histories and different levels of genetic drift between subpopulations. The model parameters were automatically adjusted with a series of pilot runs (10 pilot runs, length 2000) to check convergence of the MCMC chains, sample size was then set to 10,000 and the thinning interval to 50 (Foll & Gaggiotti 2008). A total of four independent repeats were performed with 500,000 iterations each to ensure that the detection of outlier loci was consistent. To incorporate the uncertainty of F_{IS} when using dominant markers, the Gaussian prior for the locus effects, α_i, was set to 0 mean and a standard deviation of 1 (Foll & Gaggiotti 2008; see also Holsinger *et al.* 2002). For the Gaussian prior for the population effects, β_j, the mean was set to -2 with a standard

deviation of 1.8 (Beaumont & Balding 2004). Estimated posterior probabilities >0.76 were retained as possible outliers, which corresponds to Bayes factor >3 (log_{10}PO (posterior odds) value >0.5 in BAYESCAN), which according to Jefferys model choice definition (Jefferys 1961) gives substantial support for acceptance for the model. FDR correction was used to control for multiple testing to obtain PO values that result in a FDR of no more than 5%.

To identify outlier loci associated with habitat type seven group-based structures were tested with both of the chosen approaches for identifying loci under selection. Regional structure, population origin and habitat type structures were tested. Analyses were repeated with a series of runs using only the Norfolk and Gower sampling sites, including pairwise site comparisons within these regions, as both habitat types are present across a small geographic scale. Outlier loci were identified when populations were grouped by:

1. Region – Geographic region of origin (n =4, based on results from AMOVA where the highest percentage of variability was explained by regional groupings – section 4.4.3.2), Wales (GE/GI/LI), Norfolk (NE/NI), Essex/Kent (EE/RE/KE) and Sussex (CI).
2. Geographic origin and habitat type (n = 9, all nine populations).
3. Norfolk and Gower region – Norfolk and Gower regions were compared omitting habitat type (n = 2).
4. Norfolk and Gower region and habitat type (n =4).
5. Norfolk and habitat type – Norfolk estuarine versus inland (n = 2)
6. Gower and habitat type – Gower estuarine versus inland (n = 2)
7. Habitat type only – All populations (n=2)

After the first round of outlier tests using both BAYESCAN and MCHEZA, a second round of analyses were performed after removing the outliers identified in the first round, using the same methods described above. All outliers identified before and after FDR correction were removed in the second round. This was done for all types of structure tested that yielded outliers in the first round of testing (i.e. group based structures were not tested a second time if no outliers were found in the first round).

Finally, all outliers detected by both methods were plotted to show the frequency distributions of the loci detected in each geographical population. After identification of possible habitat associated outliers, AMOVA and MANTEL analyses were repeated after the removal of the outliers from the data set using the settings described in section 4.3.6.2. This was done to identify the strength of these outliers in relation to their effects on habitat associated genetic differentiation. To determine the extent to which loci identified as being possible outliers affected overall phylogenetic structure, NJ and UPGMA trees were constructed omitting outlier loci and compared to those including all loci using methods described in section 4.3.6.3.

Loci identified under balancing selection should be taken with caution as this method cannot reliably identify such loci due to the lower 95% confidence limit often falling close to or lower than zero (Beaumont & Nichols 1996; Beaumont & Balding 2004). For this reason loci falling below the lower 95% confidence limit were conservatively not interpreted and were not removed from the data set during further simulations.

4.4. Results

4.4.1. Morphological analysis of banding pattern

The primary aim of the following analyses was to discover whether there were differences in the grey-scale pigmentation intensity and the degree of banding between inland and estuarine populations of *A. makarovi*, especially at sites where inland and estuarine population were adjoining. A supplementary aim was to identify whether there was evidence of convergent evolution of pigmentation between estuarine *A. makarovi* and the saltmarsh-specific species *A. aestuarina.* Images obtained for inland and estuarine *A. makarovi* and *A. aestuarina* from different populations illustrate some of the variation that exists within and between populations (Fig. 4.3).

For both males (Fig. 4.4) and females (Fig. 4.5), the PCAs show that estuarine *A. makarovi* cluster with *A. aestuarina* and are clearly different from inland *A. makarovi*, although with some overlap. Inland *A. makarovi* males and females group towards the negative end of PC1 (which explains 66.6% and 82.0% of the variation respectively) showing a high grey scale SD (higher degree of variation in grey scale values), low grey scale mean (darker) and low Mean:SD ratio (either darker or more varied grey scale intensities or a combination of both). Conversely, estuarine *A. makarovi* and *A. aestuarina* males and females cluster towards the positive end of PC1, showing a lower grey scale SD (lesser degree of variation in grey scale intensity), a higher grey scale mean (lighter) and higher Mean:SD ratio (either lighter or less varied grey scale intensities or a combination of both measures). Overall, the ratio between the mean grey scale intensity and SD explains the most variation in the data set (lies parallel with PC1, Fig. 4.4a and Fig. 4.5a) and separates inland *A. makarovi* from estuarine *A. makarovi* and *A. aestuarina*.

There is some degree of overlap between inland and estuarine groups but mainly in males, which are known to be highly polymorphic in *A. makarovi*. At the paired sites (Gower and Norfolk), where inland and estuarine *A. makarovi* exist in adjacent populations, there was some overlap in male colour morphs (Fig. 4.4b), explaining most of the overlap shown in Fig. 4.4a. Male polymorphism was also apparent between

inland *A. makarovi* populations. For example, inland males from Lisvane (LI) cluster together as most individuals have a low grey scale mean (darker). Norfolk inland males (NI) and a number of estuarine (NE) *A. makarovi* males show high grey scale SD values, as do individuals collected from Castle Hill (CI) and typically have higher degree of banding or variation in pigmentation. Norfolk estuarine *A. makaorvi* specimens also show higher grey scale intensities and some morphs tend to be lighter but with some banding. *Aphrodes aestuarina* populations and the Gower estuarine *A. makarovi* (GE) population show the lowest variation in grey scale intensity and highest Mean:SD ratio values, with a range of grey scale mean intensities seen. These individuals tend to be less banded, more uniform morphs, but showing varying degrees of mean grey scale intensities.

For females the separation of inland *A. makarovi* from estuarine *A. makarovi* and *A. aestuarina* was strong with very little overlap, although sample sizes were in some cases low and more specimens need to be analysed (e.g. no *A. aestuarina* female specimens were included from Norfolk). At the paired sites, Norfolk and Gower, separation of female inland *A. makarovi* from estuarine *A. makarovi* was complete, strongly suggesting lack of movement between these adjoining populations. Similarly female inland *A. makarovi* were separated completely from *A. aestuarina* at these paired sites, with *A. aestuarina* clustering with estuarine *A. makarovi*, strongly supporting the convergent evolution hypothesis. The Castle Hill females (CI) proved to be highly polymorphic and are entirely responsible for the overlap seen between inland *A. makarovi* and both estuarine *A. makarovi* and *A. aestuarina* (Fig. 4.5).

The data support the hypothesis that estuarine *A. makarovi* have evolved similar colour morphs to *A. aestuarina* and that movement between inland and estuarine populations of *A. makarovi* is limited. If the more mobile male *A. makarovi* are moving between adjoining inland and estuarine sites their female offspring would be expected to show intermediate colour morphs, but there was no evidence of this (Fig. 4.5).

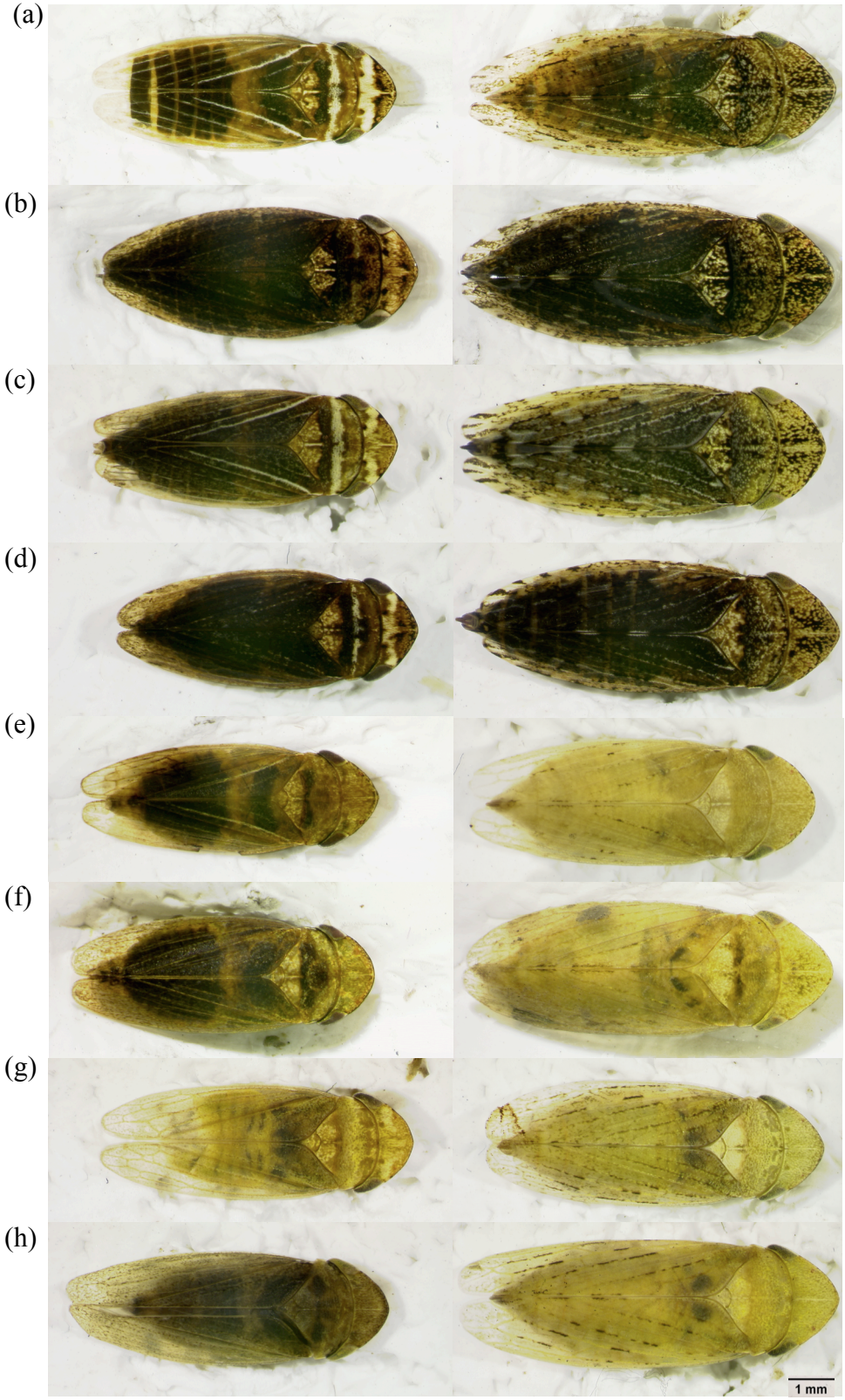


Figure 4.3. Examples of images obtained for *Aphrodes* species from inland and estuarine habitats (males = left, females = right). a-f) = *A. makarovi*; a) Castle Hill_I; b) Lisvane_I; c) Norfolk_I; d) Gower_I; e) Norfolk_E; f) Gower_E. g-h) = *A. aestuarina*; g) Norfolk_E; h) Sussex_E. Scale bar represents 1mm, images were taken at 1.6x magnification.

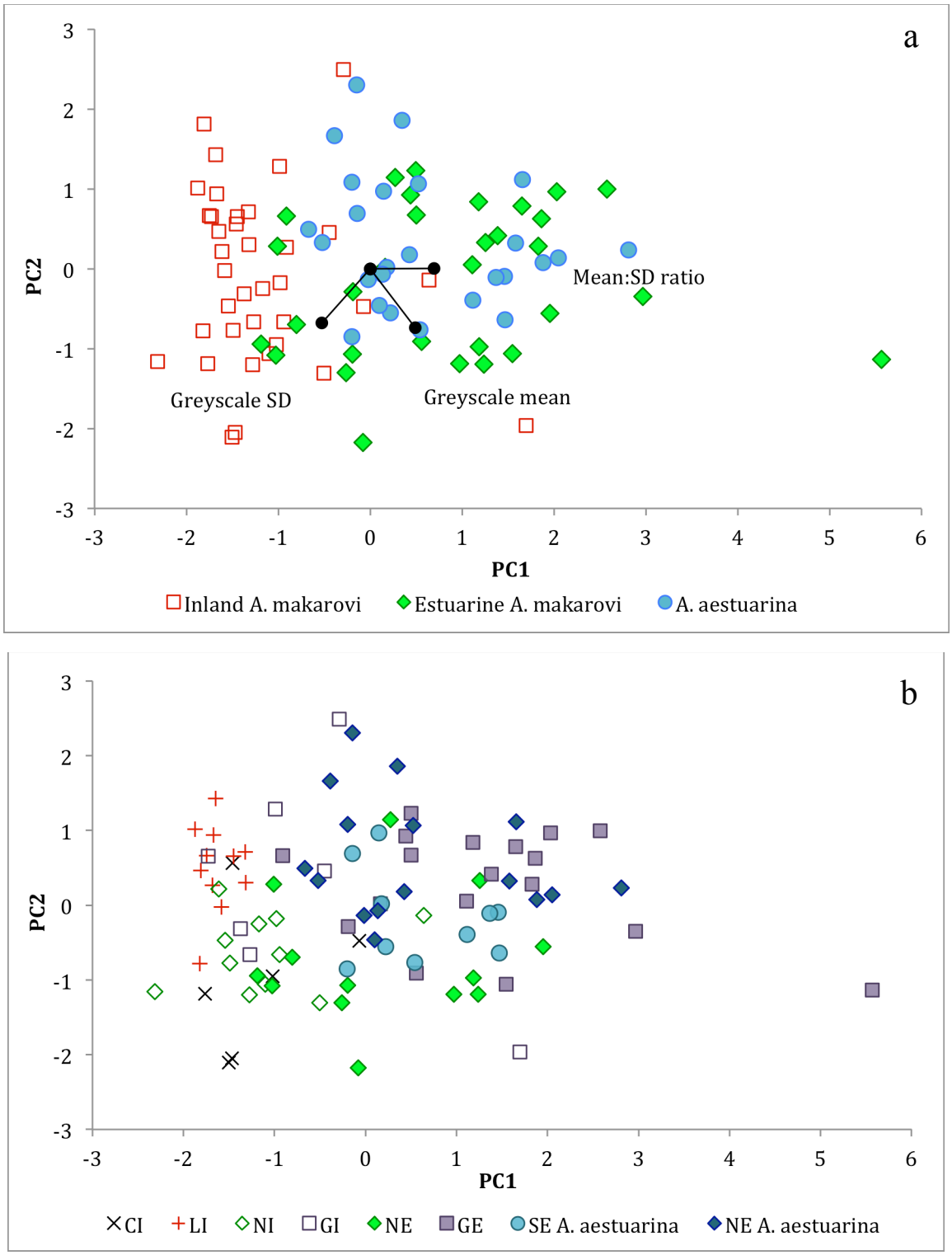


Figure 4.4. Principle component analysis, using a standardised correlation matrix of ratios taken for banding pattern and pigmentation analyses, for male *Aphrodes makarovi* and *A. aestuarina* specimens. Principle component (PC) 1 versus PC2 is shown explaining 66.6% and 31.4% of the variation, respectively. a) biplot illustrating relationship between inland and estuarine *A. makarovi* and *A. aestuarina* for PC1 versus PC2 and the corresponding ratios analysed. b) scatter plot of PC1 versus PC2 showing relationships between all sampling locations. See Table 4.1 for sampling site abbreviations.

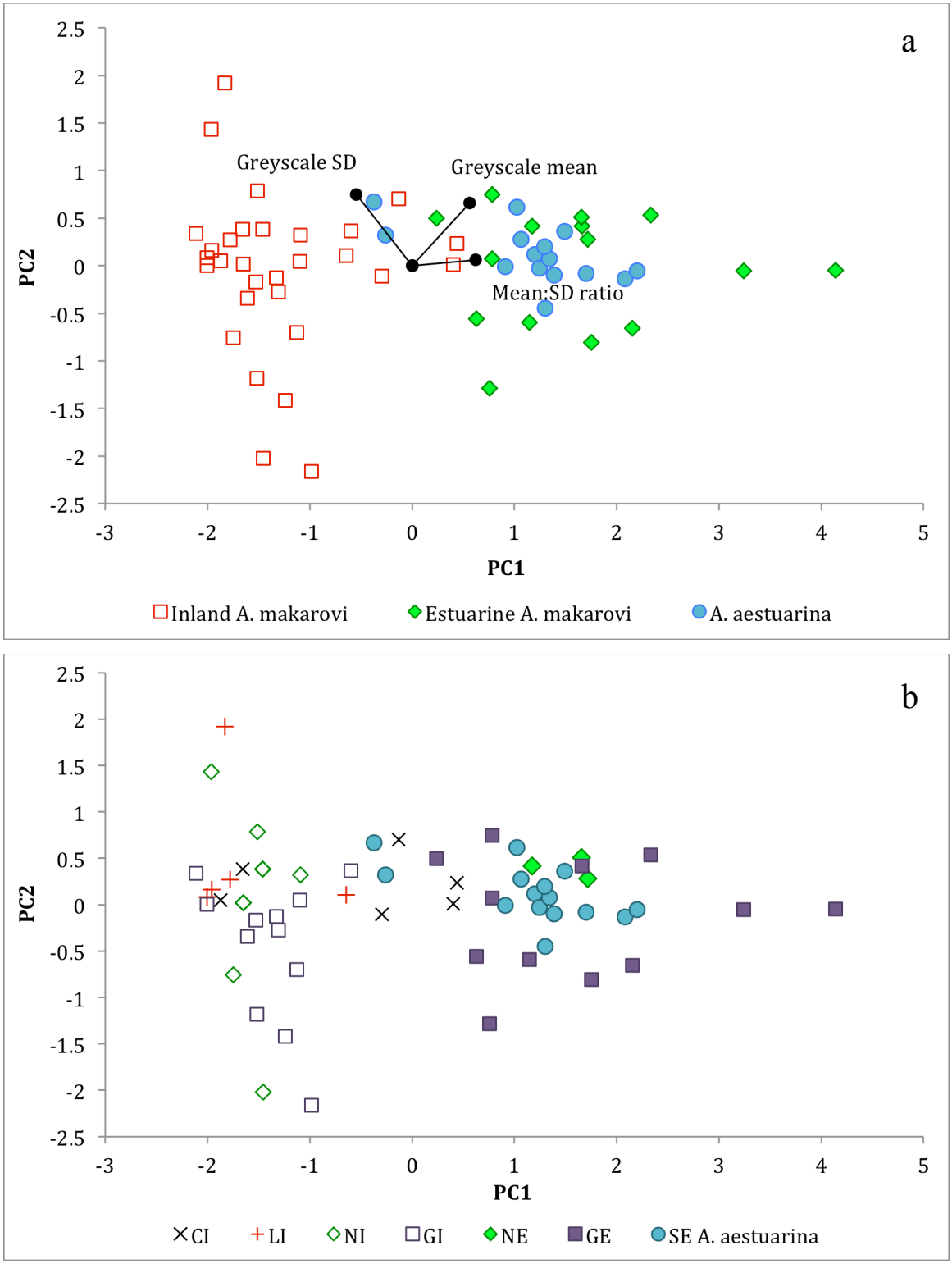


Figure 4.5. Principle component analysis using a standardised correlation matrix of ratios taken for banding pattern and pigmentation analyses, for female *Aphrodes makarovi* and *A. aestuarina* specimens. Principle component (PC) 1 versus PC2 is shown explaining 82.0% and 15.3% of the variation, respectively. a) biplot illustrating relationship between inland and estuarine *A. makarovi* and *A. aestuarina* for PC1 versus PC2 and the corresponding ratios analysed. b) scatter plot of PC1 versus PC2 showing relationships between all sampling locations. See Table 4.1 for sampling site abbreviations.

4.4.2. Mitochondrial data analysis

Fifteen unique *A. makarovi* mtDNA haplotypes for a 658 bp (inter primer length) sequence of the COI gene (GenBank Accession numbers; FR727167 – FR727170, HE587029 – HE587041) were identified from 186 specimens.

AMOVA results based on Euclidean distance between mtDNA haplotypes indicate a large proportion of genetic variation (62.35%) is attributed to differences within populations and 37.65% among populations (p-value < 0.0001, Table 4.2). When populations were grouped into six regions (Wales (GE, GI and LI), Devon (DE), Medway/Essex (RE and EE), Norfolk (NI and NE), Sussex (CI) and Kent (KE)) the highest Φ_{CT} value was obtained (Φ_{CT} = 0.065) although this was not significant (p-value = 0.316) and only 6.55% of the genetic variation was accounted for. Other regional groupings analysed were also non-significant and resulted in lower or negative percentages for the variation explained by these groupings (results not shown). Structure relating to habitat type resulted in negative values for the percentage of variation explained and this was also non significant (-5.32%, p-value = 0.931).

Network analysis of mtDNA haplotypes (Fig. 4.6) resulted in a star shaped phylogeny, a typical pattern expected for a population that has undergone recent population expansion (see also low nucleotide and haplotype diversity, shallow phylogenetic structure identified in Chapter 3). Only two haplotypes (H1 and H14) are shared between inland and estuarine *A. makarovi*, although the numbers of substitutions between haplotypes are low suggesting shallow phylogeographic structure relating to habitat association (or geographic locality) in the mtDNA of this species.

Table 4.2. Hierarchical analysis of molecular variance (AMOVA) results based on Euclidian distances between mitochondrial DNA cytochrome oxidase subunit I gene haplotypes, for 186 *Aphrodes makarovi* individuals sampled from ten populations (including Devon (DE) which was not included in amplified fragment length polymorphism analyses). Populations were analysed without (n=10 all populations separately) or with regional structuring (n=6, Wales, Devon, Medway/Essex, Norfolk, Sussex and Kent) and according to habitat type (n=2, inland and estuarine).

Source of variation	d.f.	Sum of squares	Variance	% Total	Φ statistic	*p*-value
Populations (n=10)						
Among populations	9	27.388	0.154	37.65		
Within populations	176	44.795	0.255	62.35		
Total	185	72.183	0.408		Φ_{ST} 0.377	< 0.0001
Region (n=6)						
Among groups	5	16.571	0.027	6.55	Φ_{CT} 0.065	0.316
Among populations within groups	4	10.817	0.130	31.61	Φ_{SC} 0.338	< 0.0001
Within populations	176	44.795	0.255	61.84		< 0.0001
Total	185	72.183	0.412		Φ_{ST} 0.382	
Habitat (n=2)						
Among groups	1	1.702	-0.021	-5.32	Φ_{CT} -0.053	0.931
Among populations within groups	8	25.686	0.164	41.31	Φ_{SC} 0.392	< 0.0001
Within populations	176	44.795	0.254	64.01		< 0.0001
Total	185	72.183	0.398		Φ_{ST} 0.360	

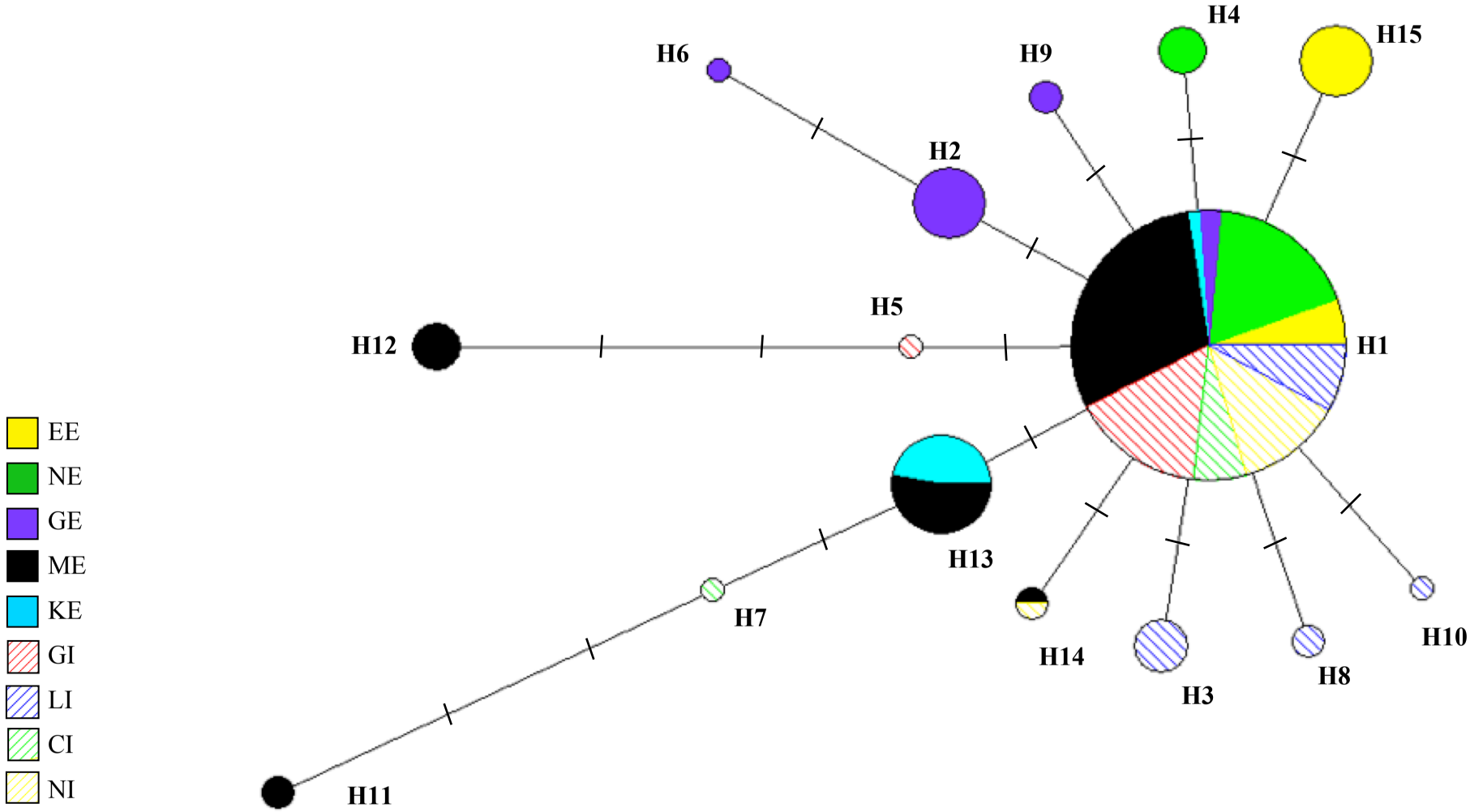


Figure 4.6. Mitochondrial haplotype network illustrating the relationship between 15 haplotypes identified for 658 base pairs of the mitochondrial cytochrome oxidase subunit I (COI) gene, which are associated with either estuarine (solid colours) or inland (diagonal stripes) sampling sites for *Aphrodes makarovi*, n = 186. H = haplotype number. See Table 4.1 for sampling site abbreviations.

4.4.3. AFLP data analysis

Eight samples were poorly amplified in AFLP analyses and were removed from the dataset leaving a total of 163 profiles (with an additional five positive control repeats) and 613 AFLP markers, amplified from three primer combinations, scored between 75-430 bp (PAAA+E44 and PAAT+E35) and 50-500 bp (PAAG+E42). Using the AFLP protocol that includes two enzymes that are both six base rare-cutter enzymes with sticky ends (*Eco*RI and *Pst*I) did not yield a significantly lower number of fragments than when a four base (*Taq*I) and six base enzyme (*Eco*RI) was used in AFLP protocol trials.

Mismatch error rates were calculated as 2.5% using AFLPSCORE. The locus selection threshold (< 400rfu) identified a total of 32 fragments to be removed due to having a low average peak height and therefore were harder to amplify reliably. After removal of fragments present in only one individual a total of 495 AFLP markers were retained.

4.4.3.1. Genetic diversity

The percentage of polymorphic AFLP markers ranged from 29.7 – 48.9% (for inland populations 34.7 – 48.9% and estuarine 29.7 – 39.4%) (Table 4.3). The inland population at Lisvane shows the highest level of polymorphism (48.9%) and the lowest was recorded at the estuarine site in Essex (29.7%.). Average expected heterozygosity (*He*) ranged between 0.11 – 0.12 (Table 4.3) and the mean number of bands per individual was 64.1 (range 41 – 86, SD 8.9). A small number of private AFLP bands were found in Wales (GE: 3 loci, GI: 1 locus), Norfolk (NI: 3 loci), Kent (KE: 1 locus) and Medway (ME: 4 loci).

Table 4.3. Data for the nine populations of *Aphrodes makarovi* sampled, including the genus of host plant present, number of specimens (*n*), percentage of polymorphic loci (*P*) and the mean expected heterozygosity (*He*) (calculated in AFLP-SURV, Vekemans *et al.* 2002). E = estuarine, I = inland.

Host	**Population**	**Abbreviation**	***n***	***He* ± SE**	***P* (%)**
Atriplex	Kent	KE	8	0.117 ± 0.007	39.4
Atriplex	Essex	EE	14	0.108 ± 0.006	29.7
Atriplex	Gower	GE	18	0.111 ± 0.006	35.8
Atriplex	Norfolk	NE	20	0.111 ± 0.007	35.6
Atriplex	Medway	ME	48	0.110 ± 0.006	34.1
Urtica	Lisvane	LI	10	0.117 ± 0.006	48.9
Urtica	Gower	GI	16	0.114 ± 0.006	34.7
Urtica	Norfolk	NI	20	0.113 ± 0.006	41.0
Urtica	Castle Hill	CI	9	0.106 ± 0.006	38.8

Mean genetic differentiation among populations, $\theta^{(\Pi)}$, calculated in HICKORY, was 0.07 (significant at the 95% credible interval, 0.0594 – 0.0820), suggesting moderate levels of differentiation among populations. Pairwise population $\theta^{(\Pi)}$ estimates (Table 4.4) that showed the highest levels of genetic differentiation were between Lisvane (LI) and other inland populations, although the sample size for Lisvane was low (n = 10). Differentiation between estuarine populations was highest when comparing Gower (GE) to other estuarine populations. The highest differentiation among different habitat comparisons is seen in comparisons involving Lisvane. A higher average $\theta^{(\Pi)}$ for different habitat comparisons (mean $\theta^{(\Pi)}$ = 0.08) is seen compared to same habitat comparisons (mean $\theta^{(\Pi)}$ = 0.07 and 0.06 for estuarine/estuarine and inland/inland comparisons respectively).

The deviance information criterion (DIC) statistic reported by HICKORY can be used as a model choice criterion (Holsinger *et al.* 2002). Values reported for the three models: full model, F_{IS} = 0 model and $\theta^{(\Pi)}$ = 0 model were 12902, 12959 and 16357 respectively. This clearly suggests preference for the full model compared to the $\theta^{(\Pi)}$ = 0, supporting the existence of a significant level of differentiation between populations. A difference of 57 between the DIC for the full model and F_{IS} = 0 model, which was not just due to the difference between the model dimensions (difference between model

dimension, pD, values is only 9), suggests some degree of departure from Hardy-Weinberg expectations.

Results for F_{IS} and F_{ST} estimates obtained from ABC4F were exceptionally high, average F_{IS} = 1 and F_{ST} = 0.76, which were not parsimonious with other estimates of differentiation among populations using various software available (mean F_{ST} = 0.02 in AFLP-SURV, F_{ST} = 0.05 in ARLEQUIN, F_{ST} = 0.07 in HICKORY) and therefore these values obtained using ABC4F were discounted. The particularly high F_{IS} estimates obtained suggest a similar problem as that previously described for HICKORY when attempting to obtain F_{IS} from dominant markers (a similar value of F_{IS} was obtained when using the full model in HICKORY which is advised to be discounted due to known problems, although HICKORY F_{ST} estimates (all models) were much lower than those obtained using ABC4F). Even when the ascertainment bias of marker selection and the biases identified by the use of non-informative flat priors were taken into consideration, it seems that unreasonable values of F_{IS} and F_{ST} were obtained with the ABC4F method.

Table 4.4. Pairwise population $\theta^{(n)}$ values using the f-free model (HICKORY, Holsinger *et al.* 2002) to estimate genetic differentiation among nine populations of *Aphrodes makarovi*. See Table 4.3 for location abbreviations. E = estuarine, I = inland populations. All values are significantly different from zero based on the 95% credible interval.

	KE	**EE**	**GE**	**NE**	**ME**	**LI**	**GI**	**NI**	**CI**
KE	/								
EE	0.0351	/							
GE	0.0874	0.0801	/						
NE	0.0547	0.0551	0.1423	/					
ME	0.0243	0.0293	0.1439	0.0416	/				
LI	0.1068	0.1329	0.0162	0.1603	0.1550	/			
GI	0.0458	0.0464	0.0897	0.0659	0.0642	0.0883	/		
NI	0.0658	0.0615	0.1228	0.0558	0.0581	0.1310	0.0240	/	
CI	0.0875	0.0551	0.0725	0.1598	0.0581	0.0765	0.0114	0.0262	/

4.4.3.2. Population structure

For population assignment analyses implemented by the program STRUCTURE no population structure (at any level tested) was identified when uncorrelated allele

frequencies were used (Appendix II (6)), thus only results for correlated allele frequencies are presented. Correlated allele frequencies allow for identification of subtle differences between closely related populations, which can be missed when allele frequencies are modelled without correlations (Falush *et al.* 2003). A lack of population structure occurs when all individuals have a high assignment to a single cluster or when all individuals show equal assignment to all clusters.

At the uppermost hierarchal level of genetic structure revealed by the analysis of all *A. makarovi* individuals, populations were separated into two groups (Appendix II (1) for the maximal log probability of the data, Pr($X|K$) reported by STRUCTURE (Table 4.10) and the ΔK results (Fig. 4.16)). When a conservative assignment threshold of ≥ 0.9 was applied to the data set only 33 (out of a total of 108) estuarine *A. makarovi* from a mixture of all geographic populations were assigned to cluster two with assignment probabilities ≥ 0.9 (Fig. 4.7). No inland *A. makarovi* were assigned to either cluster one or two using 0.9 threshold value with assignment probabilities ranging from 0.20 – 0.74 for cluster one. Using a less stringent assignment threshold of ≥ 0.7, 97 estuarine individuals were assigned to cluster two (11 not assigned) and for inland specimens, two individuals were assigned to cluster two, four individuals assigned to cluster one with the remainder showing a mixed assignment for each cluster. No structure was found relating to geographic locality. No sampling location showed a higher probability of belonging to either cluster one or two than any other, as individuals from each population were found across the range of assignment values for both inland and estuarine sites.

However, when comparing the adjacent inland and estuarine populations in Norfolk, strong structure relating to habitat type was identified (Appendix II (2), maximal Pr($X|K$) at $K = 2$, Table 4.11; $\Delta K = 2$ or 3, Fig. 4.17). Using an assignment threshold of ≥ 0.8 when $K = 2$, all estuarine individuals (n=20) were assigned to cluster two and 18 (out of 20) inland individuals were assigned to cluster one (Fig. 4.8). Using the less conservative threshold of ≥ 0.7 all inland and all estuarine specimens were assigned to cluster one and two respectively. However, at the Gower coupled inland and estuarine site no structure was identified (Appendix II (3)). No structure was identified when only inland populations or only estuarine populations were analysed by STRUCTURE in separate analyses (Appendix II (4) and (5), respectively).

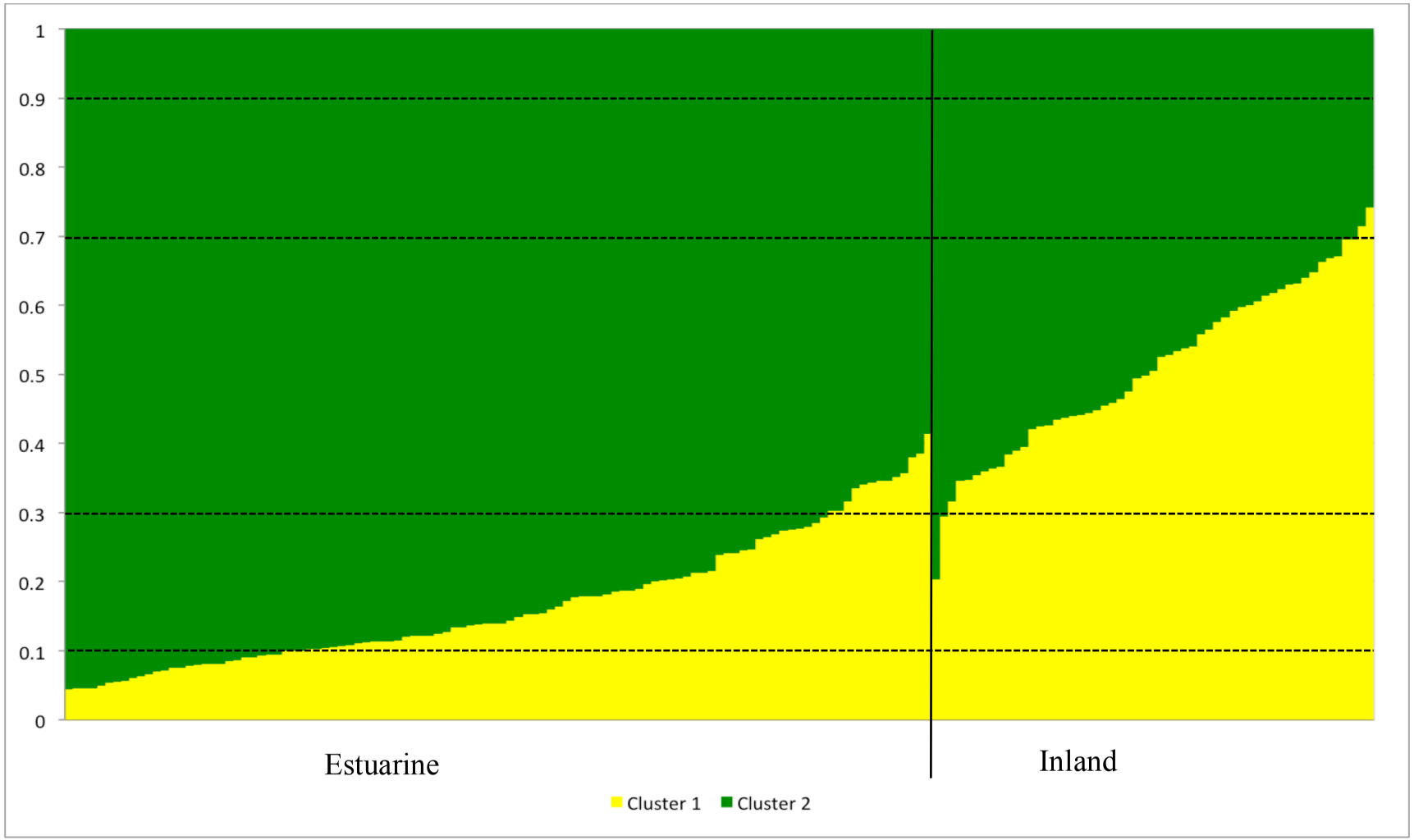


Figure 4.7. Results of STRUCTURE (Pritchard *et al.* 2000) clustering analysis showing the assignment of 163 *Aphrodes makarovi* specimens (each thin vertical bar) to either cluster one (yellow) or two (green) (*K*=2). Specimens were sampled from either estuarine or inland sites (solid black vertical line), based on 495 polymorphic amplified fragment length polymorphism (AFLP) loci. Horizontal dashed lines represent 0.9 and 0.7 probability thresholds for each cluster. Individuals were sorted according to their assignment values and sampling locations are mixed within the estuarine and inland groups.

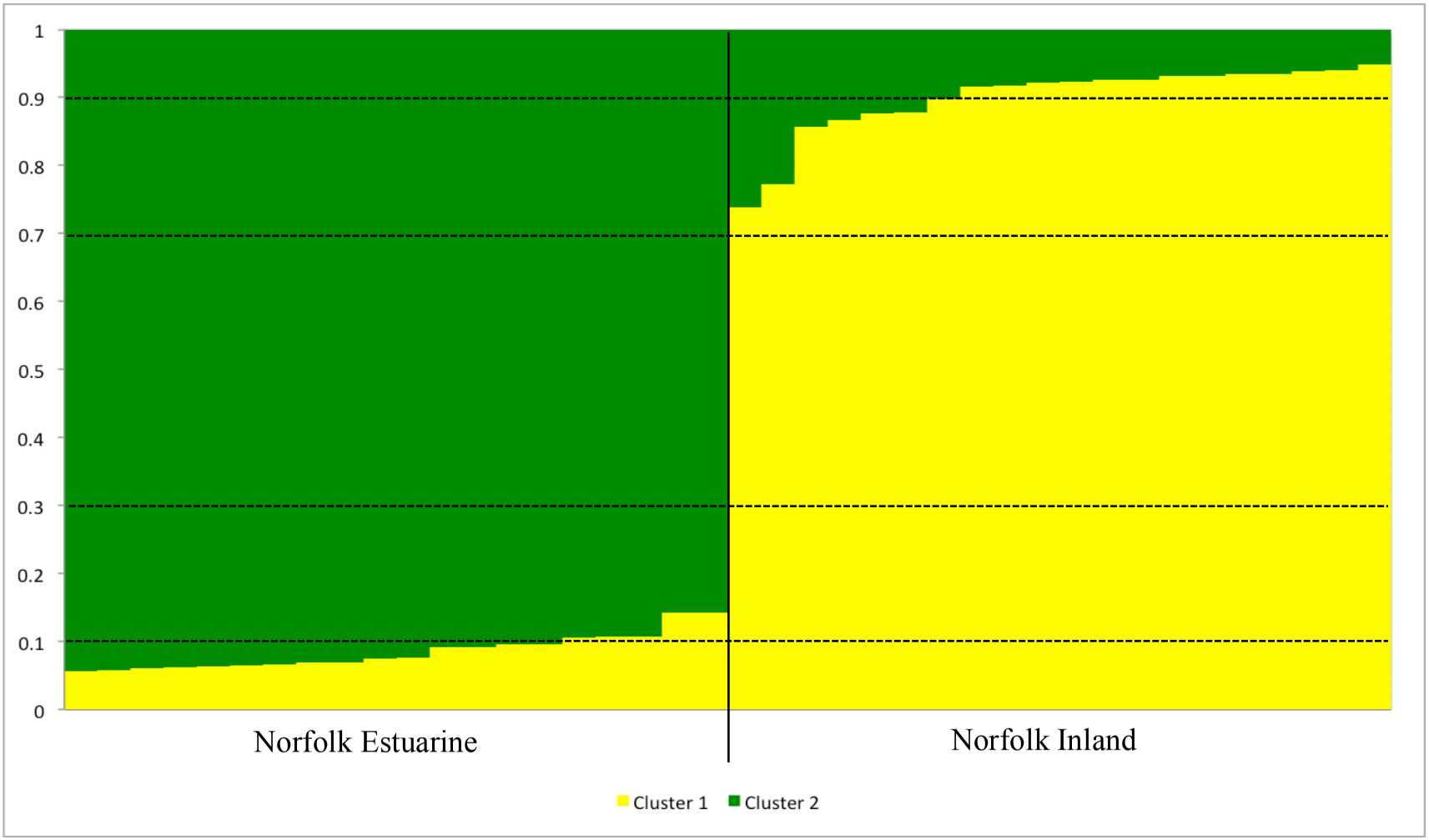


Figure 4.8. Results of STRUCTURE (Pritchard *et al.* 2000) clustering analysis showing the assignment of 40 *Aphrodes makarovi* specimens (each thin vertical bar) to either cluster one (yellow) or two (green) (*K*=2). Specimens were sampled from Norfolk estuarine and Norfolk inland sites (solid black vertical line), based on 266 polymorphic amplified fragment length polymorphism (AFLP) loci. Horizontal dashed lines represent 0.9 and 0.7 probability thresholds for each cluster.

Principle component analysis comparing all populations shows clustering into inland and estuarine groups with some overlap along the boundary of each cluster. Figure 4.9 illustrates a 3D plot for the first three principle coordinates, explaining 24.18%, 18.55% and 15.83% of the variation, respectively. Principle component analysis was carried out on inland and estuarine sites at the Gower and Norfolk (Fig. 4.10a, b). The percentage of variation explained by each principle component is comparable to those when comparing all nine populations. Clustering of coupled sites gives a similar pattern to when all populations were analysed together showing grouping into inland and estuarine clusters with little overlap suggesting clear population structure associated with habitat type. Both inland populations overlap significantly whereas the estuarine populations also cluster by location with some overlap. This suggests that the inland populations at Norfolk and Gower are more closely related to each other than they are to estuarine populations at either location. However, estuarine populations at the two sites are more different from one another than are the inland populations.

In reallocation tests carried out in AFLPOP for the whole data set, all but two individuals from estuarine *A. makarovi* populations were assigned back to the correct population of origin, with only two individuals (out of 108 total) with a log likelihood difference of < 1 (i.e. if the allocation to a certain population was less than 10 times more probable than to another one). These individuals with low likelihood values were present from RE and KE locations. Two individuals (1.85%) from the Gower_E (two females) were incorrectly allocated to the inland population with log likelihood differences of c. 7. For inland *A. makarovi* a much higher error in assignment to the correct population of origin was identified in reallocation tests. A total of two out of 55 inland specimens had log likelihood differences of < 1 (one assigned to estuarine and the other to the inland population). A total of 12 inland specimens (six males and six females, 21.82%) were incorrectly assigned to the estuarine population from NI, GI, CI and LI inland locations. Simulation results for 1000 simulated genotypes of parental populations resulted in 99.96% of individuals correctly assigned for both inland and estuarine populations (error rate of 0.04%).

Results for locally adjacent populations from the Gower, two estuarine specimens (one male and one female) were allocated to the inland population (11.11%, all with >1 log likelihood differences) and three individuals from the inland population (one male and two females) were assigned to the estuarine population (18.75%, all with > 1 log

likelihood differences). No individuals from either population were assigned with log likelihood values of < 1. Simulation results for 1000 simulated parental genotypes gave 100% successful assignment to the respective inland and estuarine populations.

For the Norfolk estuarine population, reallocation to source populations was 100% successful with no log likelihood scores of < 1. For the inland population, two individuals (one male and one female, 10%) were assigned to the estuarine population with no log likelihood scores of < 1. Simulation results for 1000 simulated parental genotypes gave 100% successful assignment to the respective inland and estuarine populations.

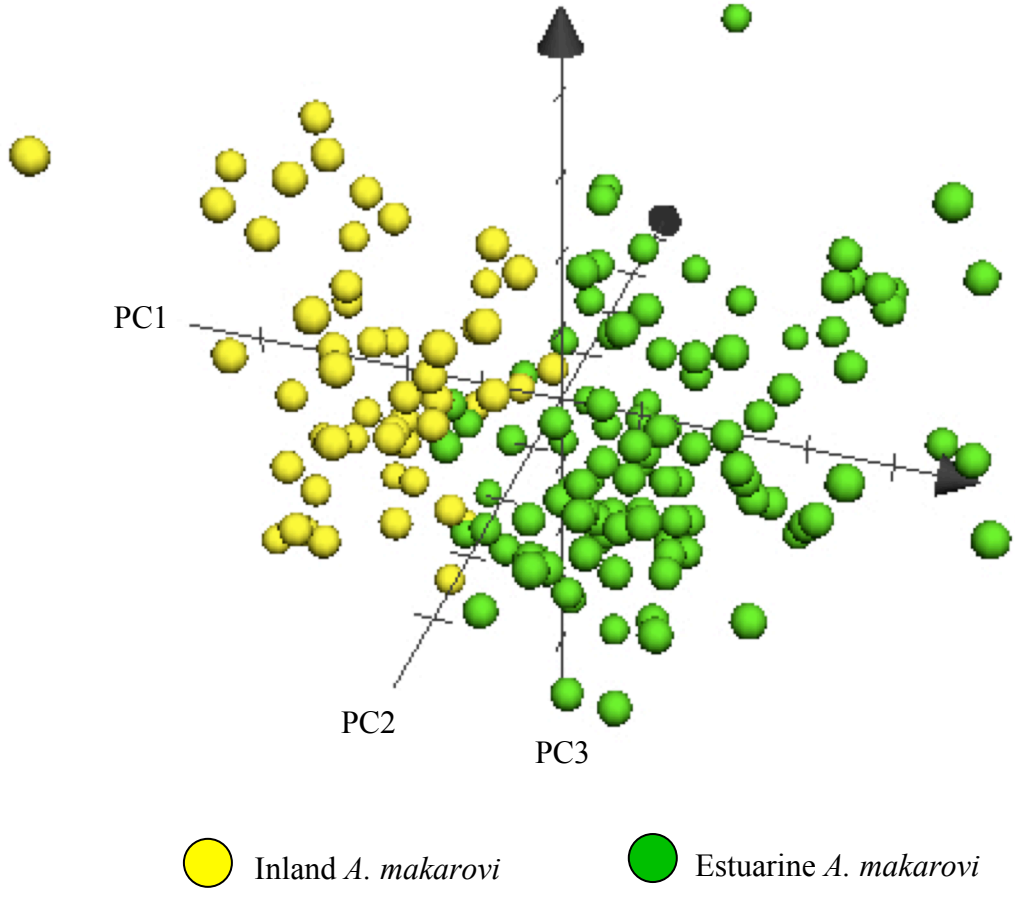


Figure 4.9. Principle coordinates analysis results based on Euclidean distances between amplified fragment length polymorphism (AFLP) multilocus phenotypes for nine *Aphrodes makarovi* populations from inland (yellow) or estuarine (green) habitat types. 3D plot illustrating the first three principle coordinates (PC1 = 24.18%, PC2 = 18.55%, PC3 = 15.83% variation explained) for 495 AFLP loci and 163 individuals.

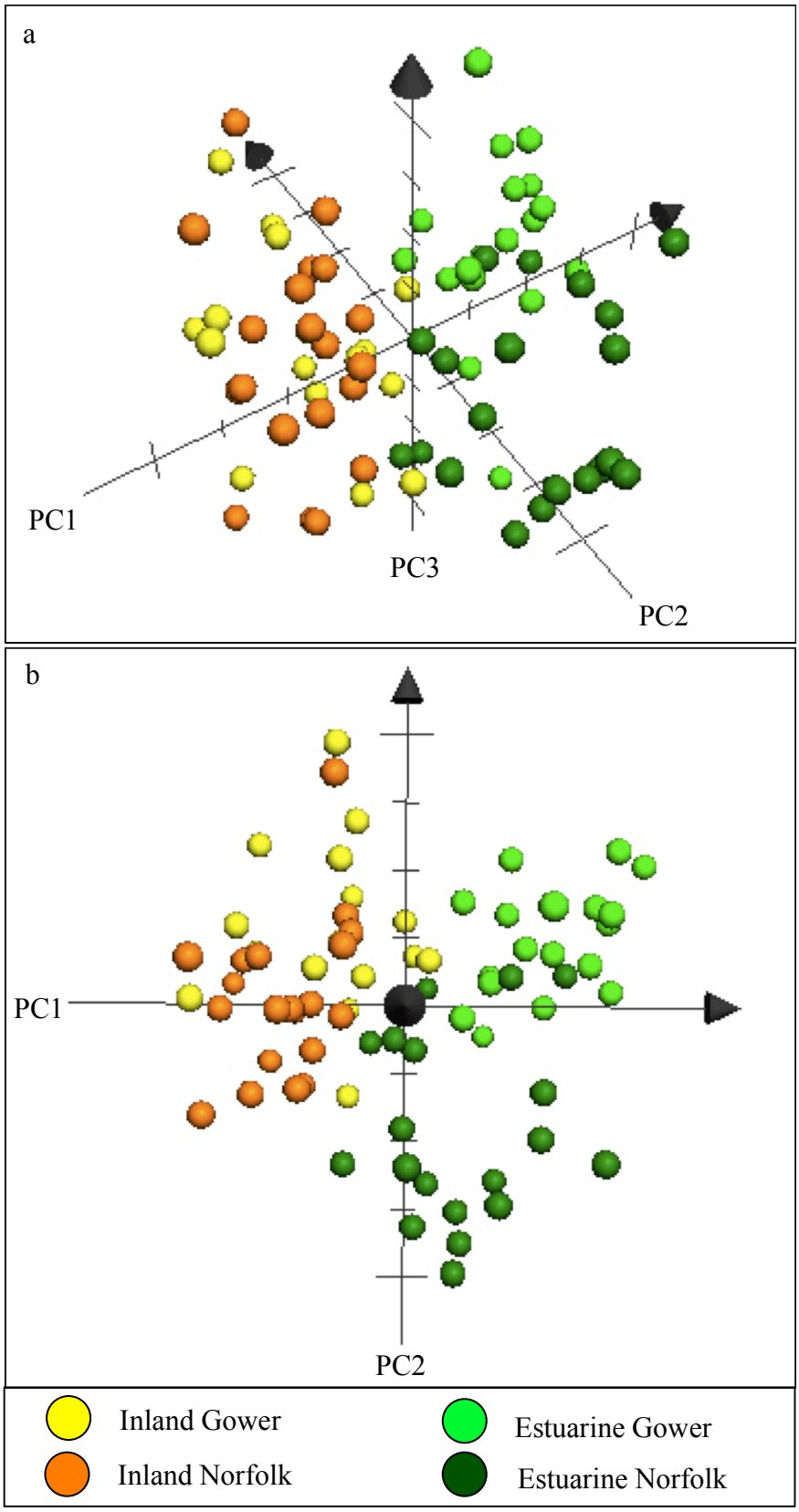


Figure 4.10. Principle coordinates analysis results, based on Euclidean distances between amplified fragment length polymorphism (AFLP) multilocus phenotypes. a) 3D plot illustrating the first three principle coordinates (PC1 = 22.09%, PC2 = 18.62%, PC3 = 15.57% variation explained), b) PC1 versus PC2 only, depicting the relationship between inland and estuarine *Aphrodes makarovi* from locally adjacent sympatric sites (Gower and Norfolk).

AMOVA results (Table 4.5) based on Euclidean distance indicate a large proportion of genetic variation (94.99%) is attributed to differences between individuals within populations and only 5.01% of variation is associated to differences among populations (*p*-value < 0.0001). When populations were grouped into four regions (Wales, Norfolk, Essex/Kent and Sussex) only 1.3% of the genetic variation was accounted for (*p*-value = 0.1), which is lower than the mismatch error rate calculated for the overall AFLP data set. Other regional groupings analysed were also non-significant and resulted in lower or negative percentages for the variation explained by these groupings (results not shown). Structure relating to habitat gave the highest Φ_{CT} value (0.032, *p*-value < 0.01), accounting for 3.21% of the total genetic variation.

Table 4.5. Hierarchical analysis of molecular variance (AMOVA) results based on Euclidian distances between amplified fragment length polymorphism multilocus phenotypes, for 163 *Aphrodes makarovi* individuals sampled from nine populations. Populations were analysed without (nine populations) or with (four populations) regional structuring and according to habitat type (two populations).

Source of variation	d.f.	Sum of squares	Variance	% Total	Φ statistic	*p*-value
Populations (n=9)						
Among populations	8	550.16	1.90	5.01		
Within populations	154	5553.64	36.06	94.99		
Total	162	6103.80	37.96		Φ_{ST} 0.050	< 0.0001
Region (n=4)						
Among groups	3	247.15	0.50	1.31	Φ_{CT} 0.013	0.10
Among populations within groups	5	303.00	1.50	3.94	Φ_{SC} 0.040	< 0.0001
Within populations	154	5553.64	36.06	94.75	Φ_{ST} 0.053	< 0.0001
Total	162	6103.80	38.06			
Habitat (n=2)						
Among groups	1	151.55	1.24	3.21	Φ_{CT} 0.032	< 0.01
Among populations within groups	7	398.60	1.25	3.23	Φ_{SC} 0.033	< 0.0001
Within populations	154	5553.64	36.06	93.56	Φ_{ST} 0.064	< 0.0001
Total	162	6103.80	38.54			

The MANTEL test carried out using Slatkin's linearised F_{ST} resulted in no significant correlation between genetic distances and geographic distances among *A. makarovi* populations (r = 0.042, *p*-value = 0.374), suggesting no significant pattern of isolation-by-distance between sampled populations (Fig. 4.11).

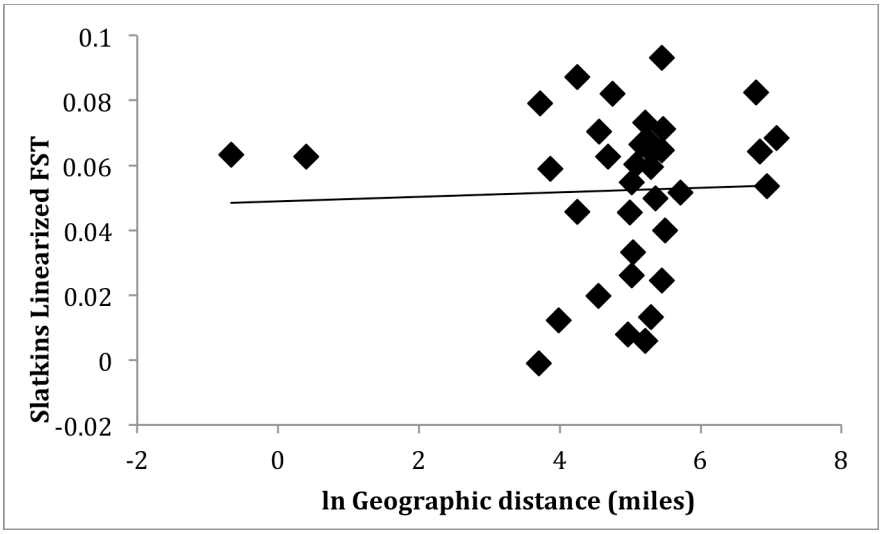


Figure 4.11. Scatter plot of Slatkin's pairwise linearized F_{ST}'s versus the natural logarithm of pairwise geographic distance (miles), for all *Aphrodes makarovi* populations. Straight line geographic distances were used for all inland population comparisons and coastal distances between coastal populations.

4.4.3.3. Phylogenetic structure

Using all loci, the NJ tree with 10,000 repetitions of Nei's pairwise genetic distance matrices was not well supported, but branched into inland and estuarine groups whereas the UPGMA tree revealed good support for both inland and estuarine clusters (not including GE) (Fig. 4.12 a, b).

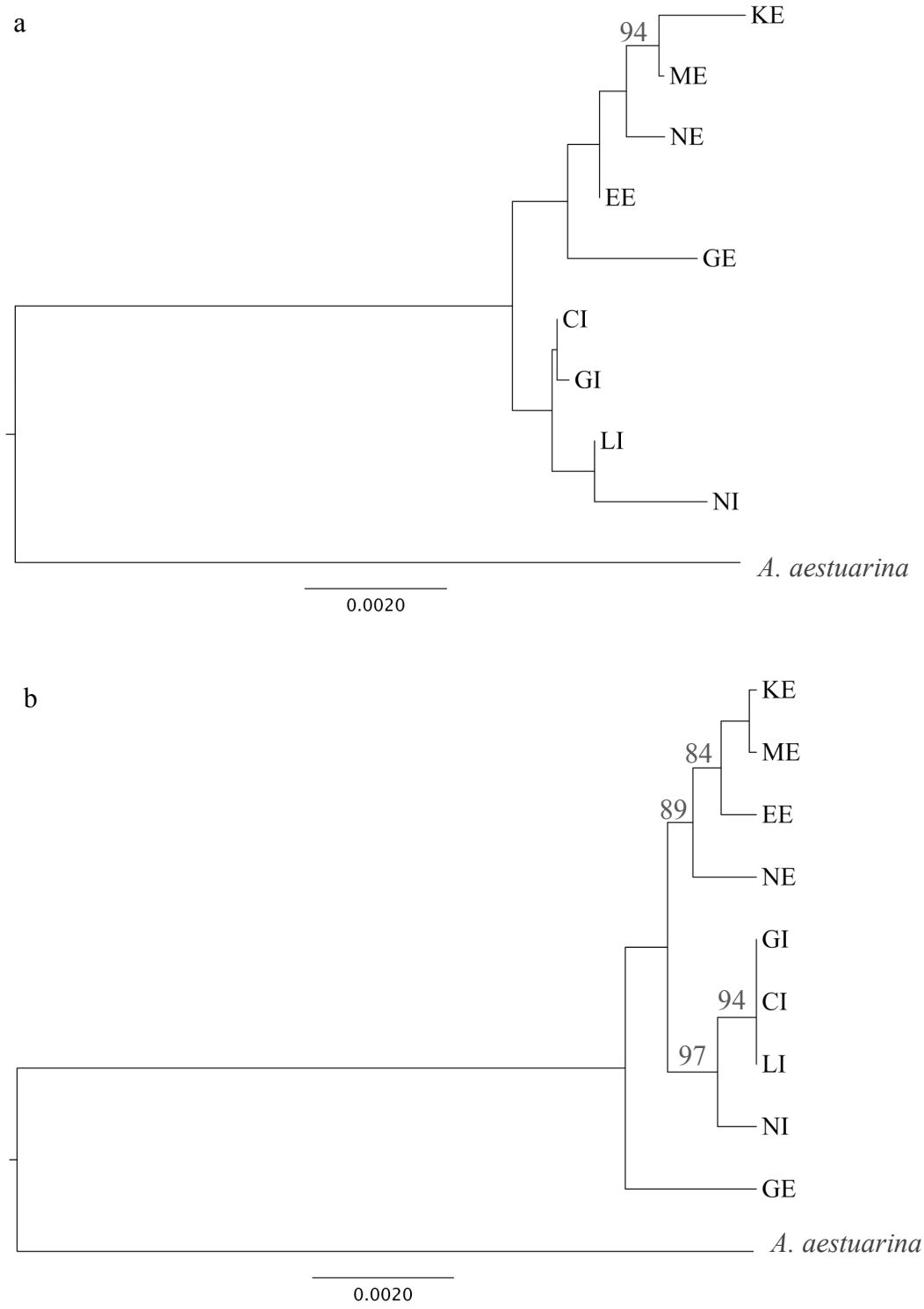


Figure 4.12. Neighbour-joining (a) and UPGMA (b) consensus phylogenies calculated in PHYLIP (Felsenstein 2005) based on 10,000 bootstrap replicates of Nei's genetic distance calculated in AFLP-SURV (Vekemans *et al.* 2002) between populations of *Aphrodes makarovi.* Bootstrap values (> 70%) for the 50% majority consensus trees are shown at the nodes of a representative non-consensus tree in order to retain branch length information. Trees are rooted using the related species, *Aphrodes aestuarina*. See Table 4.3 for location abbreviations.

4.4.3.4. Outlier loci

When all *A. makarovi* populations were grouped according to geographic region (four populations, Wales, Norfolk, Essex/Kent, and Sussex) MCHEZA identified five outlier loci (45, 270, 295, 387, 609) although two of which were not consistently obtained, occurring in less than 3/4 repeated runs (Table 4.6). BAYESCAN identified only one outlier (561), but this was not detected by MCHEZA. Using the structure relating to geographic origin and habitat type (all nine populations separately), MCHEZA identified seven outlier loci (45, 98, 132, 295, 304, 317, 609), only four of which were consistently detected. BAYESCAN detected eight outlier loci (45, 132, 295, 306, 326, 490, 538, 561), five of which were not identified by MCHEZA.

A similar structure relating to geographic region was tested that only included local pairwise populations (i.e. Norfolk versus Gower, two populations). MCHEZA identified one outlier (387), which was detected in 3/4 repeated runs. This was also one of the outlier loci previously identified in the overall regional structure (n = 4). BAYESCAN detected no outliers for this structure. When Norfolk and Gower coupled sites were grouped by region and habitat (four populations) three outlier loci were detected by MCHEZA (98, 295, 326), although, only two were consistently identified. These loci match those found when comparing geographic origin and habitat type with all nine populations. BAYESCAN detected only one outlier (326), but this was inconsistently detected by MCHEZA.

To identify habitat-specific outliers, local pairwise comparisons were performed within each region, structured according to habitat type (Norfolk and Gower separately, two populations each). No outliers were identified for Gower pairwise comparison by either method used. One outlier locus (556) was identified by MCHEZA for the Norfolk pairwise comparison although this was not recovered when using BAYESCAN. For population structure relating to habitat type, for all estuarine populations versus all inland populations (n = 2), five outlier loci were identified by MCHEZA (45, 132, 306, 538 and 605), two of which were consistently identified in both methods.

Table 4.6. Outlier loci results from the first round of outlier detection, using two methods, MCHEZA (Antao & Beaumont 2011) and BAYESCAN (Foll & Gaggiotti 2008). Repeat/4 = how many times the outlier loci were identified out of four runs. Average *p*-values for MCHEZA and average probability for BAYESCAN over four independent runs. * indicates loci retained after FDR correction.

Detection software			**MCHEZA**			**BAYESCAN**	
Data structure	**Locus**	**Repeat/4**	***p*-value > 0.99**	**F_{ST}**	**Repeat/4**	**Probability > 0.76**	**F_{ST}**
Region (n=4)	45	3/4	0.9965	0.1869	/	/	/
	270	3/4	0.9960	0.0920	/	/	/
	295	1/4	0.9913	0.0965	/	/	/
	387	2/4	0.9928	0.1023	/	/	/
	561	/	/	/	4/4	0.9754	0.0677
	609	4/4	0.9996	0.2060	/	/	/
Origin + habitat (n=9)	45	2/2	0.9950	0.1816	4/4	0.9993*	0.1463
	98	4/4	0.9998*	0.2024	/	/	/
	132	2/2	0.9949	0.1456	4/4	0.9814	0.1202
	295	4/4	0.9978	0.1721	4/4	0.8330	0.0716
	304	4/4	0.9999*	0.2542	/	/	/
	306	/	/	/	4/4	0.9359*	0.0982
	317	4/4	0.9970	0.1760	/	/	/
	326	/	/	/	4/4	0.9978*	0.1254
	490	/	/	/	4/4	0.9592*	0.0940
	538	/	/	/	4/4	0.8845	0.0935
	561	/	/	/	4/4	0.9786*	0.1012
	609	2/2	0.9952	0.1459	/	/	/
GE+I / NE+I (n=2)	387	3/4	0.9975	0.2240	/	/	/
GE / GI / NE / NI (n=4)	98	3/4	0.9957	0.2468	/	/	/
	295	4/4	0.9981	0.2793	/	/	/
	326	2/4	0.9956	0.2994	4/4	0.9480	0.1769
GE / GI – habitat (n=2)	/	/	/	/	/	/	/
NE / NI – habitat (n=2)	556		0.9995	0.4789	/	/	/
Habitat all (n=2)	45	4/4	0.9986	0.3064	4/4	0.8780	0.1348
	132	3/4	0.9968	0.1692	/	/	/
	306	4/4	0.9994	0.1594	/	/	/
	538	4/4	0.9995	0.2093	4/4	0.7731	0.1241
	605	2/4	0.9946	0.1556	/	/	/

After FDR corrections were carried out, very few loci were retained from either method used and were retained only for structure relating to geographic region and habitat (comparing all nine populations). BAYESCAN retained five outlier loci (loci 306, 326, 490, 538, 561) and MCHEZA retained two outlier loci (loci 98 and 304) after FDR correction (Table 4.6, outliers retained after FDR are denoted by an asterisk *). None of the detected outliers after FRD correction were found by both methods suggesting cautionary interpretation should be taken for all outliers detected.

Table 4.7. Outlier loci results from the second round of outlier detection, using two methods, MCHEZA (Antao & Beaumont 2011) and BAYESCAN (Foll & Gaggiotti 2008). Repeat/4 = how many times the outlier loci were identified out of four runs. Average *p*-values for MCHEZA and average probability for BAYESCAN over four independent runs.

Detection software			**MCHEZA**		**BAYESCAN**		
Data structure	**Locus**	**Repeat/4**	***p*-value > 0.99**	**F_{ST}**	**Repeat/4**	**Probability > 0.76**	**F_{ST}**
Region (n=4)	519	1/4	0.9924	0.0827	/	/	/
Origin + habitat (n=9)	/	/	/	/	/	/	/
GE+I / NE+I (n=2)	98	1/4	0.9913	0.1240	/	/	/
	295	2/4	0.9944	0.1566	/	/	/
GE / GI / NE / NI (n=4)	465	2/4	0.9934	0.2125	/	/	/
NE / NI – habitat (n=2)	295	1/4	0.9901	0.2766	/	/	/
	465	2/4	0.9948	0.3280	/	/	/
Habitat (n=2)	/	/	/	/	/	/	/

A second round of outlier analysis, after removing outlier loci identified in the first round (before FDR correction), found no additional outliers using BAYESCAN. MCHEZA identified some additional outlier loci in some of the population structures tested (Table 4.7), however, all were inconsistent occurring in no more than two of the four independent runs. None of the additional loci detected in the second round were retained after FDR correction. Runs were carried out only removing loci identified as outliers after FDR correction from the first round and results from the second round only identified loci from the first round that were not retained after FDR correction.

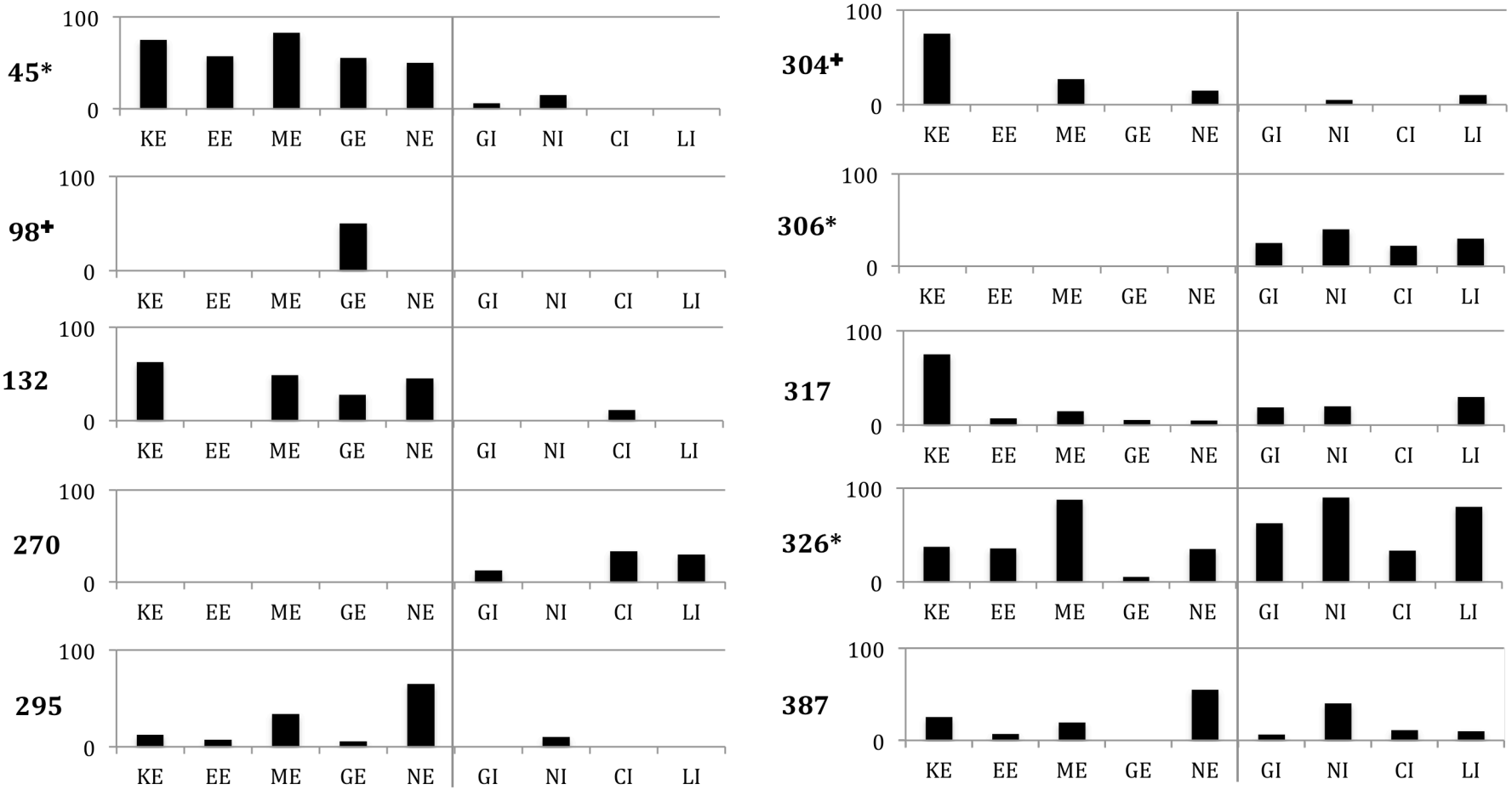


Figure 4.13. Frequency distributions (%) for all outlier loci obtained from all population structures tested. See Table 4.3 for location abbreviations. Estuarine locations = (~E), inland locations = (~I).

* loci retained after FDR correction in BAYESCAN (Foll & Gaggiotti 2008).

+ loci retained after FDR correction in MCHEZA (Antao & Beaumont 2011).

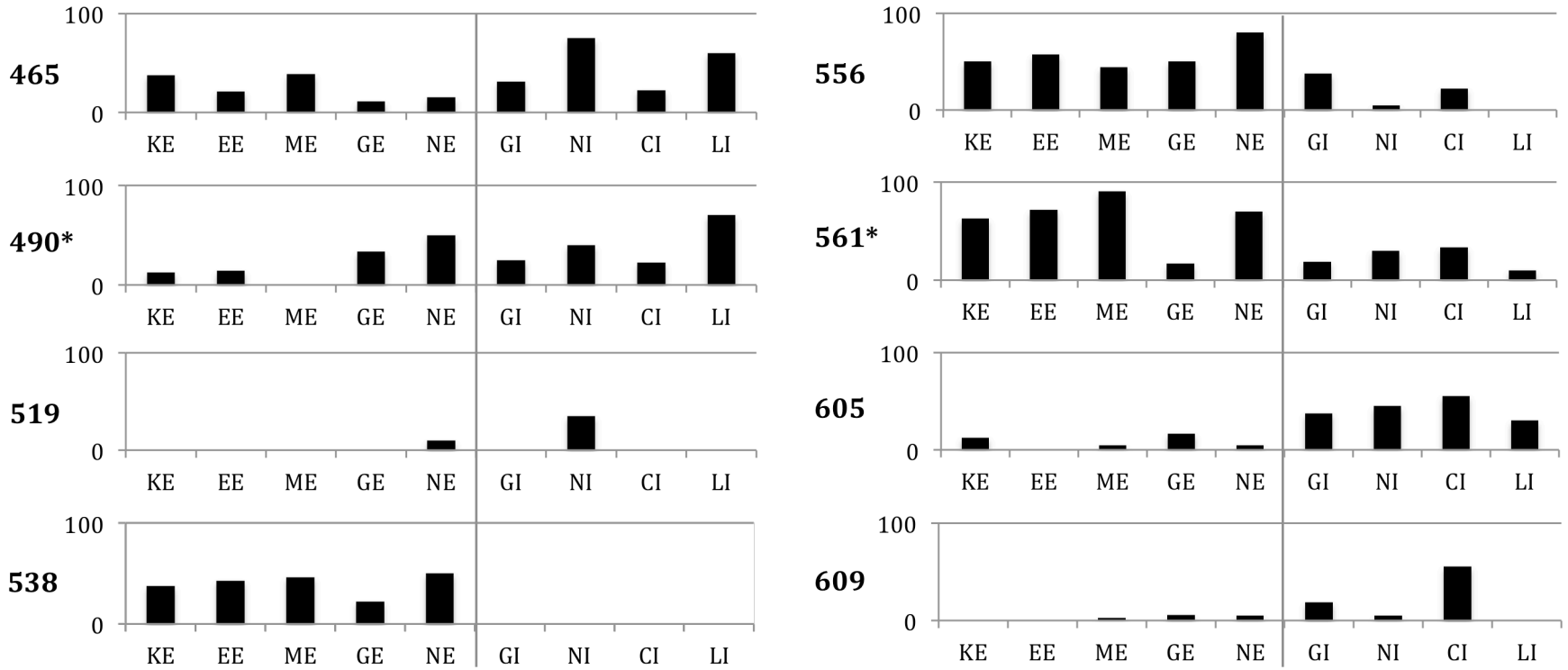


Figure 4.13 continued.

The frequency distributions (percentages) within populations of all outlier loci detected using both methods are shown in Fig. 4.13. Two outlier loci identified using BAYESCAN (retained after FDR) clearly show habitat related frequency differences (loci 45 present in all estuarine populations and loci 306 present only in inland populations). One other loci identified by MCHEZA and BAYESCAN is only present in estuarine populations (loci 538) but was not retained after FDR. Loci numbers 132, 295, 304, 326, 465, 490, 556, 561 and 605 were not so clear as to the extent that they are related with habitat type due to their presence in some but not all habitat associated populations or present in habitat and regional population structures tested Fig. 4.13). The remaining six loci not previously mentioned (loci 98, 270, 317, 387, 519, 609) show very little habitat association and are mainly present in high frequencies in particular geographic regions/populations or combinations of regions/populations with little relation to any habitat related population structure tested.

Table 4.8. Hierarchical analysis of molecular variance (AMOVA) results based on Euclidian distances between amplified fragment length polymorphism multilocus phenotypes, for 163 *Aphrodes makarovi* individuals sampled from nine populations, after removal of 12 possible habitat associated outlier loci. Populations were analysed without (nine populations) or with (four populations) regional structuring and according to habitat type (two populations).

Source of variation	d.f.	Sum of squares	Variance	% Total	Φ statistic	*p*-value
Populations (n=9)						
Among populations	8	466.60	1.39	3.90		
Within populations	154	5286.72	34.33	96.10		
Total	162	5753.33	35.72		Φ_{ST} 0.039	< 0.0001
Region (n=4)						
Among groups	3	211.42	0.47	1.30	Φ_{CT} 0.013	< 0.05
Among populations within groups	5	255.18	1.02	2.85	Φ_{SC} 0.029	< 0.0001
Within populations	154	5286.72	34.33	95.85	Φ_{ST} 0.041	< 0.0001
Total	162	5753.33	35.81			
Habitat (n=2)						
Among groups	1	105.39	0.69	1.91	Φ_{CT} 0.019	< 0.01
Among populations within groups	7	361.21	1.03	2.86	Φ_{SC} 0.029	< 0.0001
Within populations	154	5286.72	34.33	95.24	Φ_{ST} 0.048	< 0.0001
Total	162	5753.33	36.05			

The twelve loci retained as possible habitat associated outlier loci (45, 132, 295, 304, 306, 326, 465, 490, 538, 556, 561 and 605) and were removed from the data set before repeating AMOVA, isolation-by-distance (MANTEL test) and phylogenetic analyses to identify their effect on the habitat associated genetic structure between inland and estuarine populations of *A. makarovi*. Results of AMOVA analyses (Table 4.8) show a decrease in the percentage of variation explained by habitat type (3.21% before removal of outliers and 1.91% after removal), however, the *p*-value was still significant at the $p < 0.01$ significance level and retains the highest Φ_{CT} value (as found prior to outlier loci removal, Table 4.5). Removal of the 12 habitat associated outlier loci did not affect the percentage of variation explained by region (1.31% before and 1.30% after removal of outlier loci) or Φ_{CT} value (0.013 before and after), however, this result did become significant after removal of the outlier loci at < 0.05 significance level.

The MANTEL test carried out using Slatkin's linearised F_{ST} after removal of the 12 possible habitat associated outlier loci resulted in a greater correlation between geographic and genetic distance (regression coefficient, r, before = 0.042 and after removal of outlier loci, r = 0.174). Although, the *p*-value remained non-significant (*p*-value = 0.146), concordant with no significant pattern of isolation-by-distance (Fig. 4.14).

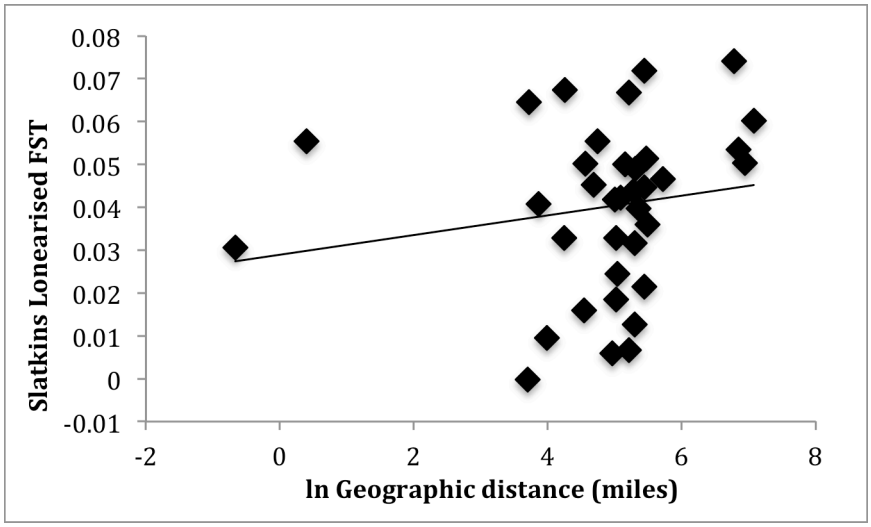


Figure 4.14. Scatter plot of Slatkin's pairwise Linearized F_{ST}'s versus the natural logarithm of pairwise geographic distance (miles) for all *A. makarovi* locations after removal of possible habitat associated outliers. Straight line geographic distances for all inland population comparisons and coastal distances between coastal populations.

Phylogenetic analyses were carried out omitting habitat associated outlier loci (Fig. 4.15). For NJ analysis omitting outlier loci, very low support for internal branches was seen (as before). However, overall tree topology changed, and Norfolk_E now forms part of the inland group and Gower_E is separate, although without good bootstrap support no conclusions can be made about the phylogenetic structure. Only good support is seen for the grouping of KE and RE in NJ analyses. The overall topology of the UPGMA tree did not change from that prior to outlier loci removal, but no bootstrap support was present leading to the estuarine group. However, there is good support for the split into inland and estuarine groups (except GE which is more basal), which was not seen prior to removal of outlier loci. This indicates that these outlier loci have little effect on the overall habitat related genetic structure seen in Fig. 4.7, but do have some effect on the bootstrap support for these groupings although overall genetic distance between all populations is weak

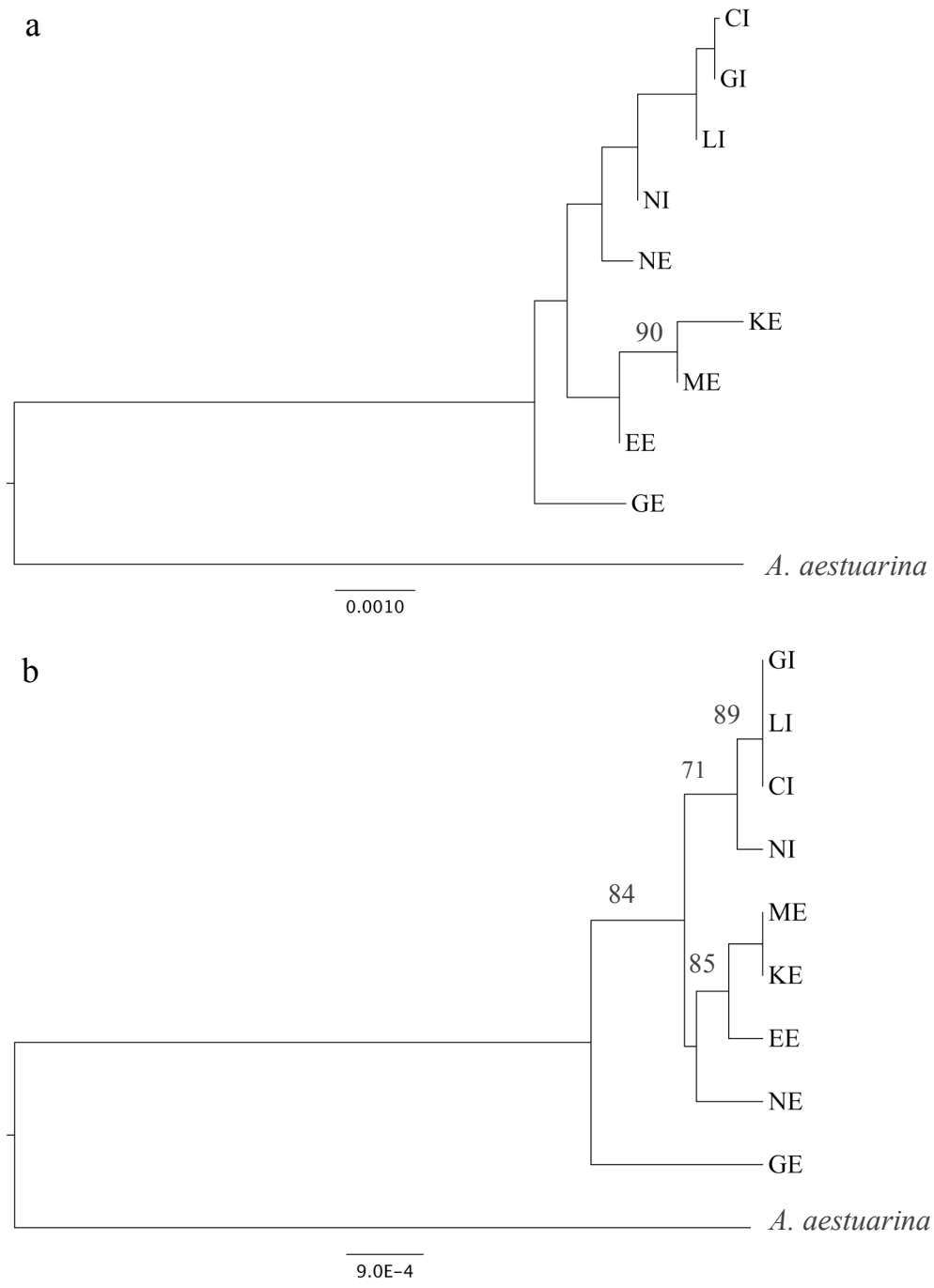


Figure 4.15. Neighbour-joining (a) and UPGMA (b) consensus phylogenies after removal of possible habitat associated outlier loci. Values above branches indicate bootstrap support >70% for the 50% majority consensus trees obtained, based on 10,000 bootstrap replicates of Nei's genetic distance between populations of *Aphrodes makarovi.* Trees are rooted using the related species, *Aphrodes aestuarina*. See Table 4.3 for location abbreviations.

4.5. Discussion

The role of habitat type (and associated host plants) in promoting genetic and morphological differentiation among *A. makarovi* populations was explored using morphological, mtDNA and AFLP marker analyses. Despite the likely generalist feeding habits of this species, adaptation to a more extreme saltmarsh environment is likely to be an important factor facilitating divergence in this species. Genetic and morphological differentiation exists between inland and estuarine populations of *A. makarovi*, although intermediate phenotypes and genotypes do occur suggesting that there is an appreciable degree of gene flow (> 1%) between populations in different habitats. Phylogenetic analyses resulted in near monophyletic habitat associated lineages, showing importance of habitat type in structuring the genetic diversity of this species. Mitochondrial DNA sequence data however, revealed no structure relating to habitat or geographic locality. The lack of fixed divergent AFLP loci or significant mtDNA structure argues that *A. makarovi* populations have diverged very recently and are in the earliest stages of ecotype formation. Further sampling of numerous locally adjacent populations over a larger geographic area, tests to determine the degree of habitat/host fidelity and fitness costs associated with respective environments are required to examine this hypothesis further. The evidence obtained here for the initial divergence of *A. makarovi* populations and ecological race formation is discussed and the contribution and insights into the process of ecological speciation.

4.5.1. Pigmentation and banding pattern polymorphism

Substantial clustering based on habitat type was identified due to differences in the degree of banding pattern and pigmentation on the head and thorax of inland and estuarine adapted *A. makarovi.* Estuarine populations (especially females) are clearly more similar to the estuarine species *A. aestuarina* compared with inland *A. makarovi,* although population sample sizes were low. This is concordant with morphological differences in whole body morphology for females (but less prominent in males) (Chapter 3). Female estuarine morphs tend to be lighter and typically show a more uniform colouration. Inland females tend to show darker and more spotted morphology, with a higher degree of contrast in pigmentation, although intermediates are found in

each habitat. This is similar for males where inland specimens tend to be darker or show a higher degree of banding than more uniform estuarine morphs, but again exceptions to this rule were found in both habitat types.

The degree of difference seen in females compared with males may be linked to dispersal as adult specimens were analysed and dispersal is likely to be linked to female movements, as they are the egg laying sex. Host fidelity is also an important factor in determining movements of phytophagous insects and thus the likelihood of host race formation (Drès & Mallet 2002); however, further studies exploring host fidelity of inland and estuarine adapted *A. makarovi* are needed. The observed differences in morphology between inland and estuarine populations could be explained by the effects of phenotypic plasticity on morphology (Drès & Mallet 2002), rather than due to an adaptive response driven by natural selection. This is unlikely though as the genetic structure identified in AFLP data indicates that there may be a genetic basis to the observed morphological variation rather than simply habitat associated phenotypic plasticity.

The importance of processes (see below) in contributing to the maintenance of colour polymorphism and the evolution of reproductive isolation is well acknowledged (Gray & McKinnon 2007). Sexual selection and sensory bias are important mechanisms; yet, whether body colouration influences mate choice in *Aphrodes* has not been tested, but is unlikely due to the overall morphological similarities between all four *Aphrodes* species. Although colour pattern is the most conspicuous phenotypic difference between inland and estuarine populations of *A. makarovi* its role in reproductive isolation cannot be assumed. Sexual communication (conspecific mate recognition) in leafhoppers (and all Auchenorrhyncha) is facilitated exclusively by species-specific vibrational mating signals (Claridge 1985; Čokl & Virant-Doberlet 2003), which do not differ significantly between *A. makarovi* populations in different habitats. Visual cues can potentially have an influence when partners are close but this requires further testing. Random genetic drift can maintain colour polymorphism, although may be difficult to detect due to the likely involvement of a number of processes (Gray & McKinnon 2007). Although, the lack of support for a role of geography in determining the nuclear AFLP differentiation means that this is unlikely (section 4.5.3).

Although the role of predation with respect to the cause of divergent selection in driving adaptive radiation remains controversial (Buckling & Rainey 2002) its importance may be under-appreciated (Meyer & Kassen 2007). Nosil and Crespi (2006) have found evidence showing that adaptive radiation can be driven by divergent selection from visual predators in *Timema* stick insects. Also, in lizards of the White Sand Ecotone, convergent evolution of light and dark dorsal colouration has evolved as predation has selected for crypsis in different light and dark habitats (Rosenblum 2006). In the peppered moth (*Biston betularia*), the increased occurrence of alleles producing melanic phenotypes is correlated with a rise in pollution levels (causing a darkening of environmental resting surfaces) during the nineteenth-century industrial revolution, in Britain (Berry 1990). Bird predation on less cryptic moth forms is thought to be a major factor causing an increase in the distribution of the melanic form, and subsequent decline as pollution levels were reduced (Cook *et al.* 2012).

Theridiid spiders (Theridiidae) are known predators of *Aphrodes* and have been shown to exploit vibrational mating signals of *A. makarovi* leafhoppers in inland habitats (Virant-Doberlet *et al.* 2011). No information is currently known about the effect of predation (by visual predators such as wolf spiders, Lycosidae) on *Aphrodes* colour morphology in different habitats. It is possible that predation pressure may have played a role in the movement of *A. makarovi* between inland and estuarine habitats as field observations during sample collections suggest that the density and diversity of spiders in inland populations is greater than in estuarine habitats, whereas *Aphrodes* densities in some estuarine habitats were higher than inland habitats (based on the considerable sampling effort required in inland habitats to obtain specimens). Visual predators may also have had an important role in the convergent evolution of colour morph of *A. makarovi* and *A. aestuarina* in estuarine habitats. Further work is required to test this hypothesis.

4.5.2. Mitochondrial DNA diversity and population structure

The proportion of mtDNA variation explained by regional structure or habitat type was low and insignificant with the majority of genetic variation present within and between sampled populations. A ‘star shaped’ network was recovered concordant with results from Chapter 3 (low nucleotide diversity, moderate haplotype diversity and shallow

phylogenetic structure), which is an indication of an abundant species that undergone a population bottleneck and recent population expansion from a small effective population size (Avise *et al.* 1987; Grant & Bowen 1998; Avise 2000). One mtDNA haplotype was widespread and likely to be ancestral (haplotype H1), while others are geographically local and are likely to be recent mutations that have not spread throughout populations (except for haplotype H1, only H14 is found in more than one sampling locality) (Avise *et al.* 1987; Avise 2000). Species with this distribution indicate phylogeographic continuity and life histories associated with weak long-term barriers to gene flow.

It is also important to consider that inference of phylogeographic history based solely on mtDNA can lead to incorrect conclusions due to other processes (introgression, selective sweeps and cytoplasmic infections) that can influence patterns of mtDNA variation (Ballard & Whitlock 2004 for an example of mtDNA introgression see Chapter 5). Reduced mitochondrial DNA variation and loss of geographical structure through selective sweeps of a single mtDNA variant has been linked to *Wolbachia* infections (Jiggins 2003). Thus it is important to employ the use of multiple gene markers when inferring phylogeographic history of species.

No significant mitochondrial structure was found but AFLP profiles showed differentiation associated with habitat type (see below). Such dis-concordance between marker types has previously been identified (Scheffer & Hawthorne 2007; Apple *et al.* 2010). Because mtDNA has a smaller effective populations size relative to nuclear genes means that mtDNA should reflect population divergence quicker than nuclear DNA (nDNA) (Rosenberg 2003). The AFLP technique samples predominantly neutral, genome-wide variation; however, some genomic regions may be linked to loci under divergent selection that approach fixation faster than neutral markers (mtDNA) that are diverging predominantly under the effects of genetic drift (Scheffer & Hawthorne 2007). A conflict between different neutral markers is expected during early stages of ecologically driven divergence and only genomic regions that are under ecological selection (or surrounding such regions) are likely to exhibit patterns of reduced gene flow (Via & West 2008), therefore possibly accounting for the lack of structure seen for mtDNA.

4.5.3. Nuclear AFLP diversity and habitat associated structure

The large number of nuclear AFLP markers used to assess genetic differentiation and population structure indicated that habitat association explained a significant proportion of genetic variance among populations sampled, compared to geographic region. Although much of the genetic variation is present within populations and heterozygosity estimates were similar across all populations sampled, genetic population structure correlating with habitat type was identified. Genetic similarity between population pairs was not significantly associated with the geographic distance between them, based on the non-significant isolation-by-distance pattern identified. This is concordant with result for mtDNA suggesting a lack of severe geographic barriers to gene flow.

Despite the lack of association with geography, the AFLP marker genotypes of adult *A. makarovi* are not randomly distributed with respect to habitat type. Clustering of the AFLP genotypes into nearly monophyletic habitat associated clusters supports the hypothesis of early-stage divergence, with some nodes showing good bootstrap support. This result means that movement of *A. makarovi* to saltmarsh environments from inland sites or vice-versa occurred probably only once and subsequently spread throughout the habitat type, rather than through a number of founder events from nearest inland or estuarine locations. As *A. makarovi* is an outbreeding and fairly mobile species, significance for internal branches within the UPGMA analysis was an encouraging result.

Overall pairwise population F_{ST} estimates reveal genetic differentiation between most inland and estuarine populations was higher than between populations present in the same habitat types. Bayesian clustering of all sampled populations also shows that inland populations have a higher degree of admixture from ancestral inland and estuarine populations compared to estuarine *A. makarovi* suggesting subtle population differences overall between different habitat types. When considering locally adjacent populations different patterns were identified in Bayesian clustering analyses, with Norfolk inland and estuarine populations showing considerable structure relating to habitat type but this was not identified at the Gower. PCA analyses and pairwise population F_{ST} estimates results reveal that estuarine populations are differentiated from other inland locations but also distinct from other estuarine populations at alternate geographic localities. Whereas inland populations are more similar to each other when

compared with estuarine populations, regardless of geographic locality. The Gower estuarine population showed significantly higher F_{ST} estimates compared to all populations regardless of habitat type (except for the inland Lisvane population) and did not cluster with either of the habitat associated clades in phylogenetic analyses.

Results of gene flow estimates from the two locally adjacent populations inferred from dominant AFLP markers using assignment and simulation tests also suggest that gene flow is appreciable (> 1%) between different habitats and is much higher at the Gower compared to Norfolk. Also, a higher number of inland *A. makarovi* specimens were incorrectly assigned to estuarine populations, so that higher rates of gene flow from estuarine to inland populations exist among the locally adjacent populations analysed here. Host races have been show to retain some ability to exploit their original hosts (Feder *et al.* 1995; Janz *et al.* 2001). It is not known whether inland or estuarine habitats (and associated host plants) are ancestral in this species. If inland habitats (and associated host plants) are ancestral to *A. makarovi,* this could explain increased gene flow from estuarine to inland habitats, as they may possess adaptations for both environments.

There are several explanations as to why different results were obtained from Gower and Norfolk sites. The age of the saltmarsh may be an important factor determining the degree of habitat-associated differentiation in *A. makarovi* and ecological differences such as tidal patterns relating to the degree of inundation and composition of plant communities. Based on field observations, Gower is less diverse in flora and fauna with extensive grass areas between suitable areas of host plant patches (*Atriplex*). The density of *Aphrodes* at the Gower is lower and only *A. makarovi* has been found there. Tidal inundations are typically high, covering most of the marsh during tidal progressions and in recent sampling years this site has received significant flooding during the months that adult *A. makarovi* are present. *Aphrodes makarovi* showed a patchy distribution at this location and was only present on the upper regions of this saltmarsh site and considerable sampling effort was required during collections (personal observation). Norfolk, however, supports both saltmarsh species in high densities with a more extensive distribution, including a more diverse array of plants and other insect groups. Certain ecological features of some saltmarshes probably make them less suitable/ more inhospitable habitats compared to well established and

extensive saltmarshes, which affect the ability and success of *Aphrodes* to colonise them.

A recent transect study exploring the distribution of saltmarsh adapted *Aphrodes* species (and other insect and plant species) across the Norfolk saltmarsh indicates that there may be some zonation in the distribution of the two *Aphrodes* species across extensive saltmarsh areas such as this. Preliminary data (Bluemel *et al.,* unpublished) reveal that *A. aestuarina* is found at the lower regions of this marsh that have a higher degree of inundation. Upper zones of the marsh support *A. makarovi* in higher densities than lower marsh zones, and are often found in sympatry with *A. aestuarina.* Upper zones receive a lower degree of inundation and are submerged completely only during high tidal progressions. Transects could not be obtained at the Gower (GE) due to the highly patchy and generally sparse distribution of *A. makarovi* at the Gower estuarine site and *A. makarovi* was not present in lower marsh zones making transect data collection impossible at this location. Probably then, where both species are present, one species affects the distribution of the other in complex ways that interact with environmental variables. Further attention to saltmarsh age, composition and diversity should be given during future investigations of habitat adaptation of *A. makarovi,* particularly when identifying suitable adjacent habitats for molecular comparison to enable detailed transects to be performed.

4.5.4. Evidence for selection and ecotype formation

Due to the non-significant isolation-by distance pattern identified and low genetic differentiation overall, migration appears to be greater than genetic drift in this species. Therefore, the frequency based method of DFDIST implemented in MCHEZA may not be appropriate for this study as the assumption of migration drift equilibrium is likely to be violated (see also Manel *et al.* 2009). This explains why different outlier loci were identified using different methods, as DFDIST (implemented in MCHEZA) does not account for population-specific demographic effects. Violations to the model can lead to higher number of false positives F_{ST} outliers (Foll & Gaggiotti 2008; Excoffier *et al.* 2009). False positives are generated using both methods (Beaumont & Balding 2004; Foll & Gaggiotti 2008) and thus incongruences between methods means that results should be taken cautiously. Both methods also have a high rate of false negatives (Foll

& Gaggiotti 2008) and further tests using an additional correlative approach using logistic regression are required to verify outlier loci (Manel *et al*. 2009) or including AFLP band intensity information may be beneficial (Fischer *et al.* 2011). BAYESCAN has been shown to be more efficient than DFDIST in certain demographic situations (Pérez-Figueroa *et al.* 2010) so BAYESCAN outlier loci are considered as more reliable here. A total of five loci (1%) were retained by BAYESCAN after FDR but only two showed significant association with habitat. These loci were not identified in regional groupings (after FDR correction) so they may possibly be directly under selection or linked to loci that are under selection. No outlier loci, however, were identified after FDR correction in any locally adjacent comparisons. Effects of selection between inland and estuarine *A. makarovi* may be diluted due to significant gene flow between populations of this fairly mobile species or *A. makarovi* populations have recently diverged.

The limitations of this study are appreciated, with only two locally adjacent populations analysed, low sample numbers in each habitat or location, and the dominant nature of AFLP markers. Even so this is the first study exploring the genetic architecture of the non-model species, *A. makarovi* inhabiting distinct habitats and results do indicate that the two sympatric ecotypes have diverged very recently. So recently in fact that mutual diagnostic variability has not evolved in either nuclear or mitochondrial markers utilised here. Despite the limitations, the loci retained as outliers would be good candidates for genetic linkage analysis to identify markers closely linked to quantitative trait loci (QTL) causing habitat specific adaptations.

Removal of all 12 possible habitat associated outlier loci found at the > 99% significance level in MCHEZA and > 0.76 posterior probability in either BAYESCAN (prior to FDR correction) had little effect on observed habitat associated phylogenetic structure (although slightly reduced the support for this association for some nodes but increased support for others in the UPGMA tree). Only marginally reduced variation explained by habitat type in AMOVA analyses was seen and habitat related structure remained significant. Removal of these loci increased the relationship seen between F_{ST} and geographic distance (MANTEL tests) although this relationship remained non-significant. This implies that there is a general barrier to gene exchange between inland and estuarine populations that is greater than would be expected from their spatial separation, and is robust against removal of these differentiated outlier loci.

Other studies have typically shown a significant reduction in support for ecological associated groupings after removal of outlier loci or changes in tree topology altogether (Wilding *et al.* 2001; Campbell & Bernatchez 2004; Bonin *et al.* 2006; Egan *et al.* 2008). However, this is not always the case as similar findings have been identified where removal of outlier loci do not affect such associations (Scheffer & Hawthorne 2007; Apple *et al.* 2010). Additional loci that show a smaller but non-negligible F_{ST} must be present that contribute to the habitat associated genetic structure seen, but whether these loci reflect divergence due to genetic drift or ecologically driven selection is not known. Divergence hitchhiking may be a reason for the more widespread nuclear differentiation identified, whereby genomic regions adjacent to outlier loci under divergent selection can experience reduced recombination due to selection against hybrids (Via & west 2008).

Size homoplasy of co-migrating AFLP fragments is also an important technical limitation to consider when identifying signatures of selection (Caballero *et al.* 2008). A large number of AFLP markers were recovered from each primer combination used, even after protocol modifications using six base restriction enzymes, although cautionary removal of all overlapping fragment regions was applied to the data set to reduce the effects of size homoplasy. An increased number of selective bases on selective primer sequences (Vekemans *et al.* 2002), to increase primer specificity and reduce the overall number of fragments, should be employed when carrying out further AFLP analyses of *Aphrodes* species. The large number of fragments recovered could be related to the size of leafhopper genomes as found in certain grasshopper species (Tatsuta & Butlin 2001). Fewer amplified AFLP fragments per primer combination may reduce the likelihood of size homoplasy, but it can still occur with low marker numbers (Whitlock *et al.* 2008). Error rate calculations were carried out to examine the robustness of the overall data set (2.5%), which was a value similar to those found in the literature (Ajmone-Marsan *et al*. 1997; Bonin *et al.* 2004).

Alternative reasons for the small number of outlier loci being retained after FDR correction argue that the pattern of selection identified is due to selection within a single generation. Under this scenario habitat associated populations of *A. makarovi* do not represent habitat/host races but rather a single panmictic population that has experienced strong habitat associated selection each generation. High mortality levels would be required to account for observed differences in AFLP frequencies as identified

here (Scheffer & Hawthorne 2007). This is a fairly limited survey of the genome and the number of AFLP loci involved may be more than realistically expected under this scenario (Scheffer & Hawthorne 2007). Identifying strong habitat associated selection within a single-generation in sympatry would be an important addition to the sympatric speciation literature.

There is evidence of introgression of *A. makarovi* mtDNA and nDNA into *A. aestuarina* thought to have occurred through a hybridisation event between these species (Chapter 5). If introgression of *A. aestuarina* AFLP nuclear loci into *A. makarovi* has occurred then this explains the colour similarities between the two species in estuarine habitats. If the gene(s) for body colouration are introgressing then only a small number of significant outlier loci may be detected, as found. However, no evidence for introgression of *A. aestuarina* AFLP loci into that of *A. makarovi* was found (Chapter 5).

An alternative reason for the occurrence of high F_{ST} outlier loci is that they occur due to intrinsic barriers created by incompatibilities between different genetic backgrounds, which often couple with ecological barriers, even when ecological selection is weak (Bierne *et al.* 2011). Genetic incompatibilities arise due to prezygotic isolation or selection that is habitat-independent (underdominance – homozygote advantage or epistasis – the effects of a gene that are modified by one or more additional genes). After contact occurs between incompatible genetic backgrounds they can form a tension zone or endogenous barriers to gene flow (Barton & Hewitt 1985; Bierne *et al.* 2011). Ecological barriers to gene flow (exogenous barriers) are groups of alleles adapted to specific environments. Because of the indirect effect of selection on neutral variation, both exogenous and endogenous barriers can significantly reduce gene flow. Tension zones can often coincide with exogenous barriers due to ecological selection (Barton & Hewitt 1985), although theory predicts that endogenous barriers are more effective barriers to neutral gene flow (Barton & Hewitt 1989; Bierne *et al.* 2011). Thus ecological barriers to gene exchange only show the position of genetic differentiation but are unlikely to explain its maintenance. Further analysis of fitness, heterozygote disadvantage, exploring the intrinsic barriers to gene flow between inland and estuarine *A. makarovi* should be undertaken.

Additional non-adaptive reasons for the presence of elevated F_{ST} outlier loci include loci from genomic regions with different inherent mutation rates or selective sweeps caused by universally favoured mutations (Akey *et al.* 2004; Bierne *et al.* 2011). Alleles in a population that coincide with the edge of range expansion (wave of advance) can also show increased frequencies (mimicking local selection at a particular locus) (Klopfstein *et al.* 2006; Excoffier & Ray 2008). This may be an important factor to consider if mtDNA data results, indicating possible population expansion, do represent the true demographic history of this species.

Predicted patterns of the early stages of divergence of sympatric ecotypes are expected to be similar to those outlined here. A number of significantly differentiated (but not fixed) AFLP loci were recovered that may themselves be (or closely linked to genomic regions) experiencing divergent selection with additional loci that also contributed to the observed patterns, but whether genetic drift or selection (via hitchhiking) is more important in determining their distributions is unclear. The lack of mtDNA structure and fixed differentiation of AFLP loci suggests recent divergence of *A. makarovi* ecotypes. The time since divergence is therefore unlikely to have been sufficient for populations to accumulate diagnostic habitat associated mtDNA variation.

Divergence of inland and estuarine *A. makarovi* populations through selection in allopatry and/or genetic drift due to historical climatic changes shifting the distributions of this species is also possible. The distribution of mtDNA and nDNA variation of presently sympatric *A. makarovi* ecotypes in the range studied show little association with geography or likelihood of a period of divergence in allopatry. However, historical patterns may have been different due to climatic changes, particularly during ice ages. Thus it is possible that initial divergence of *A. makarovi* inhabiting different environments occurred in allopatry. Inland adapted populations may have survived in glacial refugia, as it is likely that ice sheets and permafrost would have affected terrestrial species more severely than those inhabiting buffered estuarine habitats (Wilding *et al.* 2000). Due to the likely differences in distributions during recent glacial maxima, allopatric divergence followed by secondary contact and subsequent introgression cannot be ruled out (Coyne & Orr 2004). Deciphering between recent divergence in habitat use since the time of the most recent common ancestor or divergence in allopatry in a more recent glacial episode cannot be inferred from the current data. Uncovering further evidence for this refugium hypothesis requires

additional sampling of *A. makarovi* populations in different habitats, particularly from around mainland Europe (the site of potential refugia) to unravel the effects of historical and other factors on the distributions of genetic variation. However, there is currently no strong evidence for the existence of either *A. aestuarina* or estuarine *A. makarovi* anywhere else in Europe.

An important factor concerning ecotype formation is the possession of a genetically determined niche preference, such as feeding, oviposition or mating (Drès & Mallet 2002), which has not yet been addressed in detail in habitat associated *A. makarovi* populations. Host/habitat fidelity is predicted because nymphs are wingless and often found with adults collected on respective hosts. Furthermore, females are the egg laying sex and have been identified on the same host plant on for up to a week (inland *A. makarovi* found on *Urtica* at Lisvane, Virant-Doberlet personal communication). Further genetic studies should be based on wingless nymphs to ensure development occurs in/on their respective hosts/habitats, although the species identification of *Aphrodes* nymphs is difficult as they do not produce vibrational signals and are morphologically cryptic. Identification would rely on mtDNA sequence analysis (Chapter 3), or employing the use of mtDNA enzyme restriction digests to distinguish between *Aphrodes* species (Bluemel unpublished). But caution should be taken with this method when there is a possibility of hybridisation (Chapter 5).

Further divergence of presently sympatric *A. makarovi* ecotype populations is likely to depend on a combination of factors, including the degree of habitat/host plant fidelity (facilitating reproductive isolation), fitness costs associated with each habitat/host plant, intrinsic genetic incompatibilities, hybrid fitness and rates of gene flow. These are all questions that need to be addressed in detail to elucidate the cause and maintenance of differentiation identified among inland and estuarine ecotypes of *A. makarovi.*

4.6. Conclusion

Divergent natural selection related to habitat use seemingly plays an important role in structuring genetic variation in the polyphagous leafhopper, *Aphrodes makarovi*. Significant morphological and nuclear genetic differentiation identified between sympatric populations inhabiting inland and estuarine habitats, even in close geographical range, reveals that adaptation may be driven along such environmental gradients (and aided by possible habitat associated host plant specialisation). Due to the lack of fixed divergent loci or significant mtDNA structure means that *A. makarovi* populations are in all probability in the very earliest stages of ecotype formation. Further behavioural and genetic analyses of transects across locally adjacent populations of alternate habitat types and sampling across a larger geographic area are required.

Lastly, this study introduces a suitable model system of presently sympatric ecotypes that can be used for further exploration into the relative importance of geography, predation, ecological habitat and host specialisation driving divergence and speciation of phytophagous insects.

4.7. References

Ajmone-Marsan, P., Valentini, A., Cassandro, M., Vecchiotti-Antaldi, G., Bertoni, G., & Kuiper, M. (1997). AFLP^{TM} markers for DNA fingerprinting in cattle. *Animal Genetics, 28,* 418-426.

Akey, J. M., Eberle, M. A, Rieder, M. J., Carlson, C. S., Shriver, M. D., Nickerson, D. A, & Kruglyak, L. (2004). Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biology*, *2*, 1591-1599.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Bioinformatics, 9,* 323.

Antao, T., & Beaumont, M. A. (2011). MCHEZA: A workbench to detect selection using dominant markers. *Bioinformatics, doi:10.1093/bioinformatics/btr253*.

Apple, J. L., Grace, T., Joern, A., Amand, P. S. T., & Wisely, S. M. (2010). Comparative genome scan detects host-related divergent selection in the grasshopper *Hesperotettix viridis*. *Molecular Ecology*, *19*, 4012-4028.

Avise, J. C., Arnold, R. M., Ball, E., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., & Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review Ecology and Systematics*, *18*, 489-522.

Avise, J. C. (2000). *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*, 729-744.

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

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*, 113-148.

Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*, 497-503.

Beaumont, M. A., & Nichols, R. A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, *263*, 1619-1626.

Beaumont, M.A., & Balding, D.J. (2004). Identifying adaptive genetic divergence among populations from genome scans. *Molecular Ecology*, *13*, 969-980.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B*, *57*, 289-300.

Bensch, S., & Åkesson, M. (2005). Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology*, *14,* 2899-2914.

Berlocher, S. H., & Feder, J. L. (2002). Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Review of Entomology*, *47,* 773-815.

Berry, R. J. (1990). Industrial melanism and peppered moths (*Biston betularia* (L.)). *Biological Journal of the Linnean Society*, *39*, 301-322.

Biedermann, R., & Niedringhaus, R. (2004). *Die Zikaden Deutschlands*. WABV, Scheeßel. (pp. 214-215).

Bierne, N., Welch, J., Loire, E., Bonhomme, F., & David, P. (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, *20*, 2044-2072.

Black, W. C., Baer, C. F., Antolin, M. F., & DuTeau, N. M. (2001). Population genomics: genome-wide sampling of insect populations. *Annual Review of Entomology*, *46,* 441-469.

Bluemel, J. K., King, R. A., Virant-Doberlet, M., & Symondson, W. O. C. (2011). Primers for identification of type and other archived specimens of *Aphrodes* leafhoppers (Hemiptera, Cicadellidae). *Molecular Ecology Resources, 11*, 770-774.

Bonin, A., Bellemain, E., Eidesen, P. B., Pompanon, F., Brochmann, C., & Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, *13,* 3261-3273.

Bonin, A., Taberlet, P., Miaud, C., & Pompanon, F. (2006). Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Molecular Biology and Evolution, 23*, 773-783.

Bonin, A., Ehrich, D., & Manel, S. (2007). Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, *16*, 3737-3758.

Buckling, A. & Rainey, P. B. (2002). The role of parasites in sympatric and allopatric host diversification. *Nature*, *420*, 496-499.

Caballero, A., Quesada, H., & Rolán-Alvarez, E. (2008). Impact of amplified fragment length polymorphism size homoplasy on the estimation of population genetic diversity and the detection of selective loci. *Genetics*, *179*, 539-554.

Campbell, D., & Bernatchez, L. (2004). Generic scan using AFLP markers as a means to assess the role of directional selection in the divergence of sympatric whitefish ecotypes. *Molecular Biology and Evolution*, *21*, 945-956.

Claridge, M. F. (1985). Acoustic signals in the Homoptera: behaviour, taxonomy, and evolution. *Annual Review of Entomology, 30*, 297-317.

Čokl, A., & Virant-Doberlet, M. (2003). Communication with substrate-borne signals in small plant-dwelling insects. *Annual Review of Entomology, 48*, 29-50.

Cook, L. M., Grant, B. S., Saccheri, I. J., & Mallet, J. (2012). Selective bird predation on the peppered moth: the last experiment of Michael Majerus. *Biology Letters*, doi:10.1098/rsbl.2011.1136.

Coyne, J. A., & Orr, H. A. (2004). *Speciation.* Sinauer Associates, Sunderland, USA.

Dieckmann, U., & Doebeli, M. (1999). On the origin of species by sympatric speciation. *Nature*, *400*, 354-357.

Dietrich, C. H. (1999). The role of grasslands in the diversification of leafhoppers (Homoptera: Cicadellidae): a phylogenetic perspective. *Proceeding of the Fifteenth North American Prairie Conference* (pp. 44-49).

Drès, M., & Mallet, J. (2002). Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *357*, 471-492.

Duchesne, P., & Bernatchez, L. (2002). AFLPOP: a computer program for simulated and real population allocation, based on AFLP data. *Molecular Ecology Notes*, *2*, 380-383.

Edwards, J. (1908). On some British Homoptera hitherto undescribed or unrecorded. *Entomology Monthly Magazine, 44,* 56-59.

Egan, S. P., Nosil, P., & Funk, D. J. (2008). Selection and genomic differentiation during ecological speciation: isolating the contributions of host association via a comparative genome scan of *Neochlamisus bebbianae* leaf beetles. *Evolution*, *62*, 1162-1181.

Emelianov, I., Marec, F., & Mallet, J. (2004). Genomic evidence for divergence with gene flow in host races of the larch budmoth. *Proceedings of the Royal Society B: Biological Sciences*, *271,* 97-105.

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

Excoffier, L., & Ray, N. (2008). Surfing during population expansions promotes genetic revolutions and structuration. *Trends in Ecology & Evolution*, *23*, 347-351.

Excoffier, L., Hofer, T., & Foll, M. (2009). Detecting loci under selection in a hierarchically structured population. *Heredity*, *103*, 285-298.

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

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*, 1567-1587.

Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, *7*, 574-578.

Feder, J. L., Reynolds, K., Go, W., & Wang, E. C. (1995). Intra- and interspecific competition and host race formation in the apple maggot fly, *Rhagoletis pomonella* (Diptera: Tephritidae). *Oecologia*, *101*, 416-425.

Felsenstein, J. (2005). PHYLIP (Phylogeny Inference Package) version 3.6. *Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.*

Fischer, M. C., Foll, M., Excoffier, L., & Heckel, G. (2011). Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). *Molecular Ecology*, *20*, 1450-1462.

Foll, M., Beaumont, M. A., & Gaggiotti, O. (2008). An approximate Bayesian computation approach to overcome biases that arise when using amplified fragment length polymorphism markers to study population structure. *Genetics*, *179*, 927-939.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, *180*, 977-993.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & 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.

Funk, D. J., Filchak, K. E., & Feder, J. L. (2002). Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica*, *116*, 251-267.

Futuyma, D. J., & Moreno, G. (1988). The evolution of ecological specialization. *Annual Review of Ecology and Systematics*, *19*, 207-233.

Grant, S. W., & Bowen, B. W. (1998). Shallow population histories in deep evolutionary lineages of marine fishes: insights from sardines and anchovies and lessons for conservation. *Journal of Heredity*, *89*, 415-426.

Gray, S. M., & McKinnon, J. S. (2007). Linking color polymorphism maintenance and speciation. *Trends in ecology & evolution*, *22*, 71-9.

Hamilton, K. G. A. (1975). A review of the Northern hemisphere *Aphrodina* (Rhynchota: Homoptera: Cicadellidae) with special reference to the Nearctic fauna. *The Canadian Entomologist*, *107*, 1009–1027.

Hawthorne, D. J. (2001). AFLP-based genetic linkage map of the Colorado potato beetle *Leptinotarsa decemlineata*: sex chromosomes and a pyrethroid-resistance candidate gene. *Genetics*, *158*, 695-700.

Hedrick, P. W., Ginevan, M. E., & Ewing, E. P. (1976). Genetic polymorphism in heterogeneous environments. *Annual Review of Ecology and Systematics*, *7*, 1-32.

Holsinger, K. E., Lewis, P. O., & Dey, D. K. (2002). A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology*, *11*, 1157-1164.

Holsinger, K. E., & Wallace, L. E. (2004). Bayesian approaches for the analysis of population genetic structure: an example from *Platanthera leucophae* (Orchidaceae). *Molecular Evolution, 13*, 887-894.

Janz, N., Nyblom, K., & Nylin, S. (2001). Evolutionary dynamics of host-plant specialization: a case study of the tribe Nymphalini. *Evolution, 55*, 783-796.

Janz, N., & Nylin, S. (2008). The oscillation hypothesis of host-plant range and speciation. In: *Specialization, speciation, and radiation: the evolutionary biology of herbivorous insects* (ed. Tilmon KJ). University of California Press, Berkeley, CA. (pp. 203–215).

Jefferys, H. (1961). *Theory of probability*, Third edition. Oxford University press.

Jiggins, F. M. (2003). Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics*, *164*, 5-12.

Kirby, P. (1992). A review of the scarce and threatened Hemiptera of Great Britain. No. 2. *The Joint Nature Conservation Committee.* (pp. 154).

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. *Molecular Biology and Evolution*, *23*, 482-490.

Kondrashov, A. S., & Kondrashov, F. A. (1999). Interactions among quantitative traits in the course of sympatric speciation. *Nature*, *400*, 351-354.

Le Quesne, W. J. (1988). The *Aphrodes bicinctus* complex. *Proceedings of the 6th Auchenorryncha meeting, Turin, Italy* (pp. 413–415).

Litzkow, M. J., Livny, M., & Mutka. M. W. (1998). Condor – A hunter of idle workstations. *Proceedings of 8th IEEE International Conference on Distributed Computing Systems (ICDCS8), San Jose, California, USA,* (pp. 104–111).

Luikart, G., England, P. R., Tallmon, D., Jordan, S., & Taberlet, P. (2003). The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics*, *4,* 981-994.

Lynch, M., & Milligan, B. G. (1994). Analysis of population genetic structure with RADP markers. *Molecular Ecology*, *3,* 91-99.

Mallet, J. (2008). Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical transactions of the Royal Society B: Biological Sciences*, *363*, 2971-2986.

Manel, S., Gaggiotti, O. E., & Waples, R. S. (2005). Assignment methods: matching biological questions with appropriate techniques. *Trends in Ecology & Evolution*, *20*, 136-142.

Manel, S., Conord, C., & Després, L. (2009). Genome scan to assess the respective role of host-plant and environmental constraints on the adaptation of a widespread insect. *BMC Evolutionary Biology*, *9*, 1-10.

Mayr, E. (1963). *Animal species and evolution*. Cambridge, MA: Harvard University Press.

Meyer, J. R., & Kassen, R. (2007). The effects of competition and predation on diversification in a model adaptive radiation. *Nature*, *446*, 432-435.

Mitter, C., Farrell, B., & Wiegmann, B. (1988). The phylogenetic study of adaptive zones: has phytophagy promoted insect diversification? *American Naturalist*, *132*, 107-128.

Nickel, H. (2003). *The leafhoppers and planthoppers of Germany (Hemiptera, Auchenorrhyncha): patterns and strategies in a highly diverse group of phytophagous insects.* Pensoft Publishers, Sofia-Moscow. (pp. 103-105).

Nosil, P., & Crespi, B. J. (2006). Experimental evidence that predation promotes divergence in adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 9090-9095.

Nosil, P., Harmon, L. J., & Seehausen, O. (2009a). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, *24*, 145-156.

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009b). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*, 375-402.

Nylin, S., & Wahlberg, N. (2008). Does plasticity drive speciation? Host-plant shifts and diversification in nymphaline butterflies (Lepidoptera: Nymphalidae) during the tertiary. *Biological Journal of the Linnean Society*, *94*, 115-130.

Nylin, S., & Janz, N. (2009). Butterfly host plant range: an example of plasticity as a promoter of speciation? *Evolutionary Ecology*, *23*, 137-146.

Peakall, R., & Smouse P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, *6*, 288-295.

Pérez-Figueroa, A., García-Pereira, M. J., Saura, M., Rolán-Alvarez, E., & Caballero, A. (2010). Comparing three different methods to detect selective loci using dominant markers. *Journal of Evolutionary Biology*, *23*, 2267-2276.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*, 945-959.

Rasband, W. S. (1997-2007). ImageJ, U. S. National Institutes of Health, Bethesda, Maryland, USA, http://rsb.info.nih.gov/ij/.

Rice, W. R. (1987). Speciation via habitat specialization: the evolution of reproductive isolation as a correlated character. *Evolutionary Ecology*, *1*, 301-314.

Rosenberg, N. A. (2003). The shapes of neutral gene genealogies in two species: probabilities of monophyly, paraphyly, and polyphyly in a coalescent model. *Evolution*, *57*, 1465-1477.

Rosenblum, E. B. (2006). Convergent evolution and divergent selection: lizards at the white sands ecotone. *The American Naturalist*, *167*, 1-15.

Rousset, F. (1997). Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, *145,* 1219-1228.

Rundle, H. D., & Nosil, P. (2005). Ecological speciation. *Ecology Letters*, *8*, 336-352.

Scheffer, S. J., & Hawthorne, D. J. (2007). Molecular evidence of host-associated genetic divergence in the holly leafminer *Phytomyza glabricola* (Diptera: Agromyzidae): apparent discordance among marker systems. *Molecular Ecology*, *16*, 2627-2637.

Schlötterer, C. (2003). Hitchhiking mapping-functional genomics from the population genetics perspective. *Trends in Genetics*, *19*, 32-38.

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology and Evolution*, *16*, 372-380.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, *139*, 457-462.

Stinchcombe, J. R., & Hoekstra, H. E. (2008). Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. *Heredity*, *100*, 158-170.

Storz, J. F. (2005). Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology*, *14,* 671- 688.

Tatsuta, H., & Butlin, R. K. (2001). Amplified fragment length polymorphism for the analysis of genetic structure in grasshopper populations. *Journal of Orthopteran Research*, *10*, 203-207.

Tishechkin, D. Y. (1998). Acoustic signals and morphological characters of leafhoppers belonging to *Aphrodes bicinctus* group (Homoptera, Cicadellidae) from European Russia. *Entomological Review, 78*, 370-377.

Tregenza, T., & Butlin, R. K. (1999). Speciation without isolation. *Nature*, *400*, 311-312.

Vekemans, X., Beauwens, T., Lemaire, M., & Roldán-Ruiz, I. (2002). Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, *11,* 139-151.

Via, S. (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution, 16*, 381-390.

Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, *17*, 4334-4345.

Virant-Doberlet, M., Prešern, J., Stewart, A. J. A., & Wilson, M. R. (2005). The genus *Aphrodes*: bioacoustic approach to solve taxonomic problems. *Congress of International BioAcoustic Council (IBAC)*. (pp. 35).

Virant-Doberlet, M., King, R. A., Polajnar, J., & Symondson, W. O. C. (2011). Molecular diagnostics reveal spiders that exploit prey vibrational signals used in sexual communication. *Molecular Ecology*, *20*, 2204-2216.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Fritjers, A., Pot, J., Peleman, J., Kuiper, M., & Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, *23*, 4407-4414.

Weir, B. S., & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, *38*, 1358-1370.

Whitlock, R., Hipperson, H., Mannarelli, M., Butlin, R. K., & Burke, T. (2008). An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Molecular Ecology Resources*, *8,* 725-735.

Wilding, C. S., Grahame, J., & Mill, P. J. (2000). Mitochondrial DNA COI haplotype variation in sibling species of rough periwinkles. *Heredity*, *85*, 62-74.

Wilding, C. S., Butlin, R. K., & Grahame, J. (2001). Differential gene exchange between parapatric morphs of *Littorina saxatilis* detected using AFLP markers. *Journal of Evolutionary Biology*, *14,* 611-619.

Zhivotovsky, L. A. (1999). Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, *8,* 907-913.

4.8. Appendix

Appendix I – Sampling locations and geographic coordinates

Table 4.9. Geographic coordinates (GPS) for all sampling locations across the UK.

Location	Host plant/ habitat type	Country	Geographic co-ordinate
Kent_E			
Pegwell Bay	*Atriplex* - saltmarsh	England	N 51° 18.954' E 001° 21.590'
Essex_E			
Canvey Bridge	*Atriplex* - saltmarsh	England	N 51° 32.551' E 000° 33.834'
Canvey Island	*Atriplex* - saltmarsh	England	N 51° 31.343' E 000° 37.025'
Mersea Island	*Atriplex* - saltmarsh	England	N 51° 47.728' E 000° 55.322'
Horsey Island	*Atriplex* - saltmarsh	England	N 51° 51.552' E 001° 14.649'
Gower_E			
Penclaudd	*Atriplex* - saltmarsh	Wales	N 51° 38.646' W 004° 06.659'
Norfolk_E			
Stiffkey Marsh	*Atriplex* - saltmarsh	England	N 52° 57.621' E 000° 55.389'
Warham Marsh	*Atriplex* - saltmarsh	England	N 52° 57.367' E 000° 54.505'
Medway_E			
R (site destroyed)	*Atriplex* - saltmarsh	England	N 51° 23.460' E 000° 30.560
R5-Baty's Marsh	*Atriplex* - saltmarsh	England	N 51° 22.588' E 000° 28.986'
R6-Riverside Walk	*Atriplex* - saltmarsh	England	N 51° 23.348' E 000° 30.617'
R7-Gillingham Pier	*Atriplex* - saltmarsh	England	N 51° 23.849' E 000° 33.327'
R9-Hoo St Werbergs	*Atriplex* - saltmarsh	England	N 51° 24.677' E 000° 33.727'
R10-Stoke/Grain	*Atriplex* - saltmarsh	England	N 51° 26.959' E 000° 39.356'
Lower Twydall	*Atriplex* - saltmarsh	England	N 51° 23.198' E 000° 35.610'
Funton Creek	*Atriplex* - saltmarsh	England	N 51° 22.807' E 000° 41.825'
Gower_I			
Penclaudd	*Urtica* - Grassland	Wales	N 51° 38.609' W 004° 05.742'
Church Lane	*Urtica* - Grassland	Wales	N 51° 38.020' W 004° 05.992'
Graveyard	*Urtica* - Grassland	Wales	N 51° 38.262' W 004° 05.925'
Near Ilston	*Urtica* - Grassland	Wales	N 51° 35.482' W 004° 05.429'
Norfolk_I			
Warham Marsh coastal footpath	*Urtica* - Grassland	England	N 52° 57.367' E 000° 54.505'
Stiffkey Marsh coastal footpath	*Urtica* - Grassland	England	N 52° 57.409' E 000° 55.384'
Stiffkey Marsh Road	*Urtica* - Grassland	England	N 52° 57.272' E 000° 55.431'
Stiffkey-Wells Road	*Urtica* - Grassland	England	N 52° 56.946' E 000° 54.404'
Lisvane_I			
Lisvane	*Urtica* - Grassland	Wales	N 51° 32.160' W 003° 10.173'
Castle Hill_I			
Castle Hill	*Urtica* - Grassland	England	N 50° 50.473' W 000° 04.400'

Appendix II – Determining the number of genetic clusters (*K*) in analyses using STRUCTURE version 2.2 (Pritchard *et al.* 2000).

1. Inland vs Estuarine *A. makarovi* populations (correlated allele frequencies).

Both the maximal log probability of the data, Pr(*X*|*K*) reported by STRUCTURE (see value highlighted in Table 4.10) and the Δ*K* method of Evanno *et al.* (2005) that calculates the rate of change in the log probability of data between successive *K*-values (Fig. 4.16), identified two genetic clusters (*K* = 2) for all inland and estuarine *A. makarovi* populations. For *K* = 2 assignment results see Figure 4.7, Section 4.4.3.2.

Table 4.10. Log probability values given by STRUCTURE, averaged over 10 replicate runs for each value of *K* tested (*K* = 1-10).

K	Log probability	*K*	Log probability
1	-28046.5125	6	-28615.8375
2	**-27763.025**	7	-28338.0875
3	-35325.4125	8	-28266.225
4	-31924.5	8	-29478.6625
5	-29094.6875	10	-30197.95

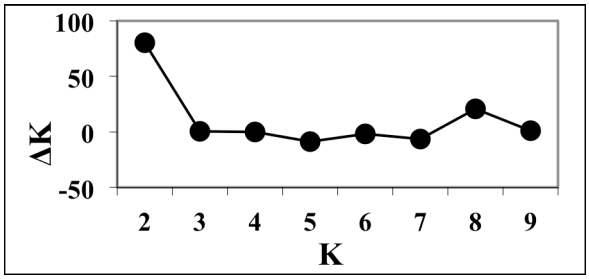


Fig. 4.16. Δ*K* values for *K* genetic clusters (*K* = 2-9).

2. Norfolk Inland vs Norfolk Estuarine (correlated allele frequencies).

The maximum log probability (see value highlighted in Table 4.11) is shown at *K* = 2 for Norfolk inland and Norfolk estuarine populations. The Δ*K* method (Fig. 4.17) identified two or three genetic clusters (*K* = 2 or 3). For *K* = 2 assignment results see Fig. 4.8, Section 4.4.3.2. For *K* = 3 assignment results see Fig. 4.18 below. Results for *K* = 3 are similar to that shown in Fig. 4.8, except that four inland Norfolk individuals show a high probability (> 0.9) of belonging to the third cluster.

Table 4.11. Log probability values given by STRUCTURE, averaged over 10 replicate runs for each value of *K* tested (*K* = 1-5).

K	Log probability
1	-5779.18
2	**-5753.38**
3	-5864.71
4	-6688.22
5	-6482.41

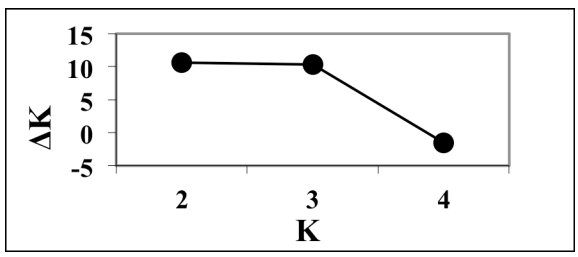


Fig. 4.17. Δ*K* values for *K* genetic clusters (*K* = 2-4).

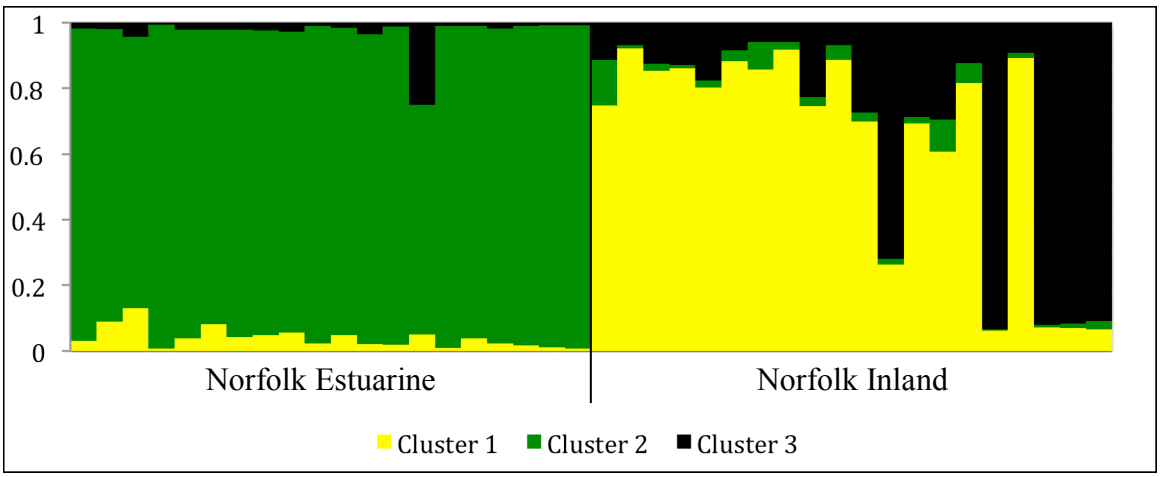


Fig. 4.18. Assignment probabilities for Norfolk inland versus Norfolk estuarine populations for three clusters, $K = 3$.

3. Gower Estuarine vs Gower Inland (correlated allele frequencies).

Both the maximum log probability (see value highlighted in Table 4.12) and the ΔK (Fig. 4.19) reveal two genetic clusters ($K = 2$) when comparing the Gower estuarine and inland populations. For $K = 2$ assignment results see Fig. 4.20 below. No genetic structure was identified with all individuals from the Gower inland and estuarine locations showing similar assignment values for both clusters.

Table 4.12. Log probability values given by STRUCTURE, averaged over 10 replicate runs for each value of K tested (K = 1-5).

K	Log probability
1	-5034.85
2	**-5023.4**
3	-5098.13
4	-5318.23
5	-5355.47

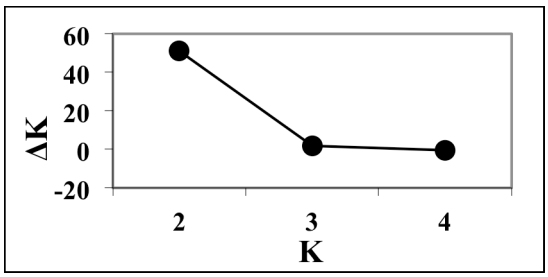


Fig. 4.19. ΔK values for K genetic clusters (K = 2-4).

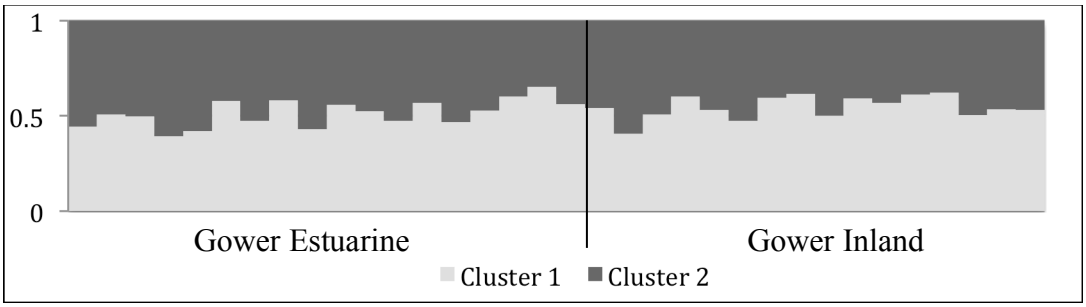


Fig. 4.20. Assignment probabilities for Gower estuarine versus Gower inland populations for two clusters, $K = 2$, showing no genetic structure.

4. All inland *A. makarovi* populations (correlated allele frequencies).

Both the maximum log probability (see value highlighted in Table 4.13) and the Δ*K* (Fig. 4.20) identified three genetic clusters (*K* = 3) when comparing all inland *A. makarovi* populations. For *K* = 3 assignment results see Fig. 4.22 below. No genetic structure was identified with all individuals from all inland populations showing similar assignments values for the three clusters.

Table 4.13. Log probability values given by STRUCTURE, averaged over 10 replicate runs for each value of *K* tested (*K* = 1-5).

K	Log probability
1	-8775.71
2	-8779.58
3	**-8747.48**
4	-8805.41
5	-8868.43

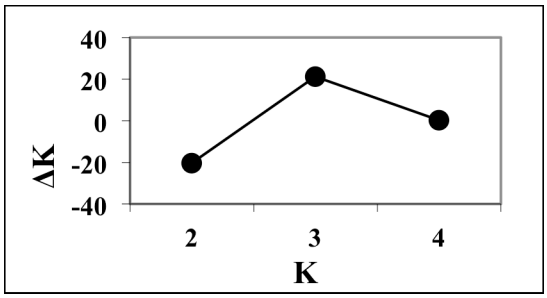


Fig. 4.21. Δ*K* values for *K* genetic clusters (*K* = 2-4).

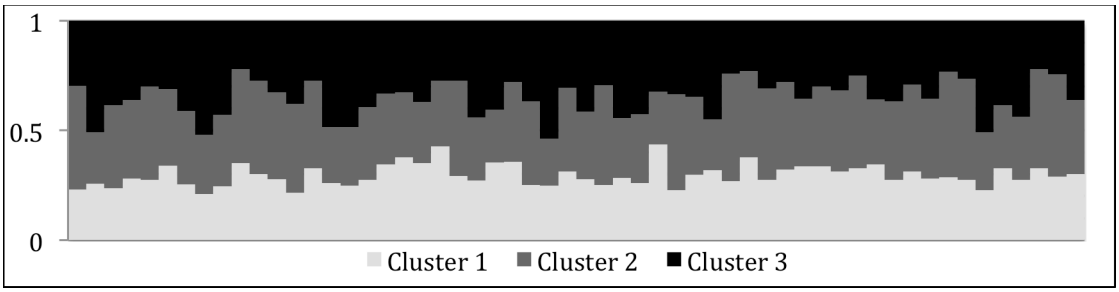


Fig. 4.22. Assignment probabilities for all inland *A. makarovi* specimens for three clusters, *K* = 3, showing no genetic structure.

5. All estuarine *A. makarovi* populations (correlated allele frequencies).

Both the maximum log probability (see value highlighted in Table 4.14) and the Δ*K* (Fig. 4.23) imply two genetic clusters (*K* = 2) when comparing all estuarine populations. For *K* = 2 assignment results see Fig. 4.24 below. No genetic structure was identified with all inland populations showing similar assignments values for both clusters.

Table 4.14. Log probability values given by STRUCTURE, averaged over 10 replicate runs for each value of *K* tested (*K* = 1-5).

K	Log probability
1	-17178.54
2	**-17159.96**
3	-20362.38
4	-19440.68
5	-20065.34

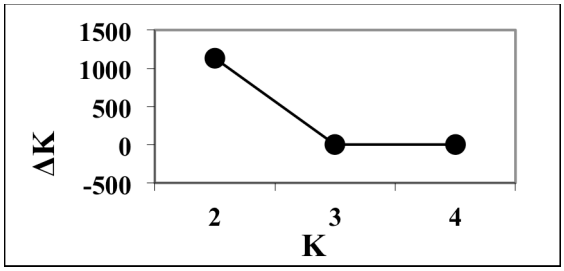


Fig. 4.23. Δ*K* values for *K* genetic clusters (*K* = 2-4).

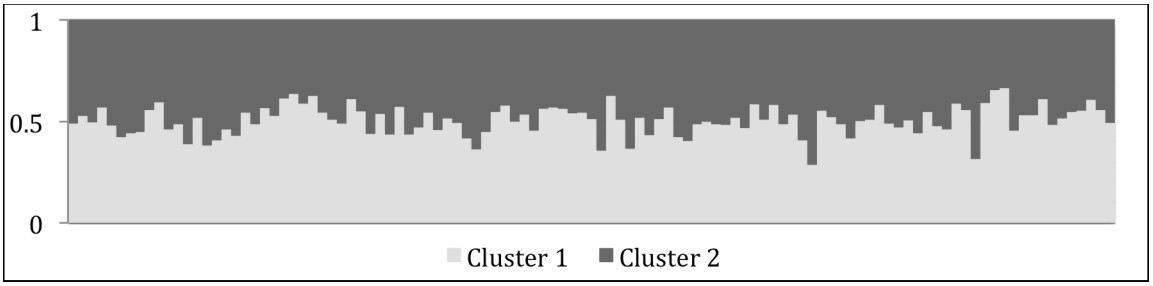


Fig. 4.24. Assignment probabilities of all estuarine *A. makarovi* for two clusters, *K* = 2, showing no genetic structure.

6. Results using un-correlated allele frequencies.

For all STRUCTURE results using un-correlated allele frequencies, no genetic clustering was seen, with the highest log probability typically shown at *K* = 1 and very little variation in Δ*K* values (results not shown). For all values of *K* tested using un-correlated allele frequencies, all individuals were assigned to a single cluster with high probability (≥ 0.9) and are therefore not reported here.

Chapter 5:

Introgressive hybridisation in *Aphrodes* leafhoppers (Hemiptera, Cicadellidae): exploring the mismatches between vibrational signals, mitochondrial DNA and AFLP genotypes.

5.1. Abstract

A recent study uncovered discordance between species-specific vibrational mating signals and mitochondrial DNA (mtDNA) in a single population of *Aphrodes* leafhoppers (Hemiptera, Cicadellidae) in the Medway estuary, UK (Chapter 3). A combined approach was taken, employing the use of mtDNA cytochrome oxidase subunit I (COI) gene sequences and 554 bi-parentally inherited nuclear amplified fragment length polymorphism (AFLP) loci to address the hypothesis that the discordance between mating signal and mtDNA is the product of hybridisation and introgression between *Aphrodes makarovi* and *A. aestuarina*. Unambiguous distinction between *A. makarovi* and *A. aestuarina* was recovered in nuclear AFLP Bayesian clustering analyses, concordant with mating signal data for all populations including the Medway estuary. Complete fixation of mtDNA from *A. makarovi* was observed in the mismatched Medway estuary *A. aestuarina* population, which is 6.9% (K2P distance) divergent from that of *A. aestuarina* mtDNA found at other localities. Of the 42 mismatched specimens identified, 95.6% were found to possess the most common *A. makarovi* haplotype (mH1), also present among sympatric Medway *A. makarovi*. Together, these results suggest that interspecific mtDNA exchange is likely to explain the reticulate evolutionary pathway for this mismatched population, rather than retention of an ancient ancestral polymorphism or convergence. Low levels of uni-directional nuclear introgression were observed, so the hybridisation event is likely to be of historical origin, followed by repeated backcrossing of hybrids with *A. aestuarina.* A number of private AFLP loci (and two private single base pair divergent mtDNA haplotypes based on current sampling effort) were recovered from the mismatched population, providing additional support for a lack of on-going or recent hybridisation. Further evidence is also required to determine among several possible reasons for the fixation of *A. makarovi* mtDNA, including chance (drift) or selection for *A. makarovi* mtDNA and/or linked nuclear genes in the genetic background of *A. aestuarina.* These results demonstrate the potential for introgressive hybridisation to have substantial and possible long-term effects on the genetic configuration of species and can produce considerable discrepancies among speciation histories based on nuclear and mitochondrial markers.

5.2. Introduction

The role of introgressive hybridisation as an important source of evolutionary innovation is becoming increasingly well acknowledged due to the mounting evidence that it can promote heterozygosity, adaptive potential and can even lead to speciation (Arnold 1992; Harrison 1993; Buerkle *et al* 2000; Coyne & Orr 2004; Seehausen 2004; Mallet 2005, 2007). Introgression therefore has important consequences for studies exploring biodiversity, speciation and conservation (Mallet 2005; Ryan 2006).

Introgression is defined as the infiltration of the genes of one species into the gene pool of another through repeated backcrossing of an interspecific hybrid with one of the parental populations. This phenomenon is common in plants, but is becoming increasingly well recognised in animals (Mallet 2005; Schwenk *et al.* 2008), including birds (Grant *et al.* 2004; Vallender *et al.* 2007; Grant & Grant 2010), butterflies (Bull *et al.* 2006; Kronfrost *et al.* 2006), termites (Lefebvre *et al.* 2008), cichlid fish (Egger *et al.* 2007; Koblmüller *et al.* 2007), eels (Albert *et al.* 2006), wildcats (Beaumont *et al.* 2001) and many more. Such evidence comes from widespread studies using genetic markers to reveal patterns of reticulate evolution (natural hybridisation between different evolutionary lineages) at the molecular level.

Asymmetrical barriers to gene flow are often identified in studies of hybrid zones where introgression of genes is greater in one direction than the other (Barton & Hewitt 1985; Kronfrost *et al.* 2006; Gomes *et al.* 2009) and thus may give clues to recent movements in a hybrid zone (Harrison 1993). Introgression of mitochondrial DNA (mtDNA) and nuclear DNA (nDNA) does occur, but many studies of natural populations have reported significant mitochondrial introgression with little or no nuclear introgression (Ferris *et al.* 1983; Powell 1983; Dowling & DeMarais 1993; Bernatchez *et al.* 1995; Wilson & Bernatchez 1998; Shaw 2002; Linnen & Farrell 2007; Gompert *et al.* 2008; Renoult *et al.* 2009; reviewed in Chan and Levin 2005), suggesting that introgression of mtDNA may be more common (Ballard & Whitlock 2004). However, because differentiation of mtDNA typically occurs more rapidly than nuclear divergence, mtDNA introgression may just be easier to identify (Gompert *et al.* 2008). Alternatively, if low numbers of nuclear markers are used to address the likelihood of introgression there may be a lower chance of identifying introgressed markers.

Introgression of mtDNA can result in complete replacement in some species (Llopart *et al.* 2005) and can have considerable affects on mtDNA phylogenies (Funk & Omland 2003; Ballard & Whitlock 2004).

One possible explanation for biased mtDNA introgression is that foreign mtDNA alleles are relatively neutral and not linked to alleles with associated fitness costs in novel environments and/or genetic backgrounds in comparison to foreign nuclear alleles (Martinsen *et al.* 2001; Funk & Omland 2003). Maladapted recombinant genotypes are unlikely to survive long enough to reproduce (hybrid breakdown), therefore nuclear introgression is reduced due to indirect selection on nuclear genes, but the low linkage between mtDNA and nDNA would reduce the effects of hybrid breakdown (Ballard & Whitlock 2004). A theoretical study by Chan & Levin (2005) suggests that frequency-dependant assortative mating can explain elevated rates of mtDNA introgression (when females are the choosy sex, heterospecific males are more likely to be accepted when conspecific males are rare). Alternatively, if certain mtDNA variants result in fitness advantages and improved survival, then a selective sweep can drive the fixation of particular haplotypes associated with higher fitness, which are under direct positive selection (Ballard & Whitlock 2004). Selective sweeps can also occur through indirect positive selection for mtDNA haplotypes linked to maternally inherited infections, such as *Wolbachia* (Jiggins 2003; Gompert 2008).

The phylogenetic pattern produced by mitochondrial introgression is similar to those that arise due to retention of ancestral polymorphism, typically prevalent in young species radiations, as time since speciation may be insufficient for mitochondrial differences to accumulate (incomplete lineage sorting) (Moritz *et al.* 1987). Additionally, convergence would result in a similar pattern and may be expected if divergent lineages were under similar selection pressures (Funk & Omland 2003). All of these processes typically show non-monophyly of mitochondrial gene trees and discordance between nuclear and mitochondrial phylogenies. Incomplete lineage sorting potentially affect any single-locus gene tree and thus identifying mitochondrial introgression requires comparison of mtDNA phylogenetic patterns against a nuclear background (consistent phenotypic or genotypic differences) that can discriminate among the parent species in sympatry (Funk & Omland 2003). Conflicting genealogies can be obtained depending on the marker type because each gene has a distinctive history that is determined by selection and mutation (Ballard & Whitlock 2004). It is

therefore important to use a range of marker types, including nDNA markers, to distinguish between the possible mechanisms that lead to reticulate evolutionary patterns (Funk & Omland 2003; Ballard & Whitlock 2004).

A recent study examining the four morphologically cryptic species of the leafhopper genus *Aphrodes* revealed congruence between male vibrational mating signals and phylogenetic analysis of mitochondrial (mtDNA) cytochrome oxidase subunit I gene (COI) sequence data, supporting the existence of four distinct, well-supported, monophyletic species (Chapter 3). Sexual communication (conspecific mate recognition) in leafhoppers (and all Auchenorrhyncha, except cicadas) is facilitated exclusively by species-specific vibrational mating signals (Claridge 1985; Cŏkl & Virant-Doberlet 2003). In the *Aphrodes* genus, substrate borne male mating signals have been shown to discriminate among species (Tishechkin 1998; Virant-Doberlet *et al.* 2005; 2006; Chapter 3), as found in other phytophagous insect taxa (Virant-Doberlet & Čokl 2004). However, at a number of saltmarsh sites around Rochester in the Medway estuary (UK), individuals were identified showing a mismatch between vibrational signal and mtDNA. These mismatched specimens produced the male mating signal of *A. aestuarina* (or females responded to it in signal playback tests, Virant-Doberlet unpublished data) but clustered with *A. makarovi* in the phylogeny based on mtDNA (Chapter 3). The majority of mismatched specimens shared the most common haplotype present in sympatric and allopatric populations of *A. makarovi* (Chapter 3) with either one or two individuals possessing one of two related haplotypes (both differing by one base pair from the commonest *A. makarovi* haplotype).

Mismatched individuals were identified either in sympatry with *A. makarovi* or in populations in the Medway estuary made up solely of individuals showing a mismatch between mating signal and mtDNA, but never in sympatry with non-mismatched *A. aestuarina*. Vibrational mating signals were recorded for a subset of individuals and larger numbers of specimens were analysed using molecular methods (Chapter 3). It is possible that other unrecorded males (or females not analysed for their response in signal playback experiments) from other sampled populations of estuarine *A. makarovi* and *A. aestuarina* may also in fact be misclassified when based solely on mitochondrial data. However, to my knowledge the mismatched specimens are limited to sites around the Medway estuary.

The closest location to the Medway estuary where *A. aestuarina* have been identified is across the Themes estuary along the Essex coast identified by mating signal and mtDNA analyses. Beyond this *A. aestuarina* have been identified extensively along the Sussex and Norfolk coastlines (Chapter 3). Based on current knowledge, *A. aestuarina* are only found inhabiting estuarine locations on the host plant *Atriplex portulacoides* (Sea Purslane, Chapter 3, but have also been documented on *Sueda vera*, Shrubby Seablight, Edwards 1908; Kirby 1992). *Aphrodes makarovi* is found ubiquitously at inland sites (on a range of plants including *Urtica, Taraxcum* and *Cirsium* sp., Tishechkin 1998; Nickel 2003; Biedermann & Niedringhaus 2004; Chapter 4), but are also frequently found, sometimes in sympatry with *A. aestuarina,* at estuarine sites inhabiting the same host plants (Chapter 2 – Bluemel *et al.* 2011; Chapter 3). Populations of *A. makarovi* inhabiting such ecologically different habitats may represent the earliest stages of ecotype formation (Chapter 4).

Hybridisation and subsequent introgression of *A. makarovi* mtDNA to *A. aestuarina* may be a possible explanation for the mismatch observed between mating signal and mtDNA sequences in the Medway estuary. Alternatively, the mismatch identified may be due to the retention of an ancient ancestral polymorphism or convergent evolution within the mtDNA. To distinguish among these hypotheses, the amplified fragment length polymorphism (AFLP) technique (Vos *et al.* 1995) was employed to examine the genomic divergence among *A. makarovi, A. aestuarina* and the Medway estuary populations. A large number of genome-wide nuclear markers can be assayed using this technique without the need for designing specific primers (Ajmone-Marsan *et al.* 1997; Bensch & Åkesson 2005), and is therefore particularly useful for studies on non-model organisms (Bensch & Åkesson 2005; Meudt & Clarke 2007). This technique has been increasingly used for natural populations of a variety of organisms to examine genetic diversity, population structure and identify cases of introgressive hybridisation (Gompert *et al.* 2006; Kronfrost *et al.* 2006; Bonin *et al.* 2007; Egger *et al.* 2007; Koblmüller *et al.* 2007; Vallender *et al.* 2007; denHartog *et al.* 2010; McKinnon *et al.* 2010; Colbeck *et al.* 2011).

If the reticulate evolutionary pattern identified in the mtDNA gene tree for the Medway estuary population is a product of mitochondrial introgression, then it would be expected that the patterns of relatedness shown in the nuclear genome should be more similar to that of *A. aestuarina* compared to *A. makarovi* (Funk & Omland 2003;

Gompert *et al.* 2006). Alternatively, if the nuclear genotypes identified in the mismatched Medway population are more similar to *A. makarovi* than to *A. aestuarina* then both mtDNA and nDNA marker patterns would not agree with current taxonomy based on vibrational mating signals. Additionally, as *A. makarovi* is found in sympatry in the Medway estuary (and with *A. aestuarina* elsewhere), recent hybridisation and nuclear introgression may also be prevalent, in which case a range of intermediate, highly admixed genotypes would be expected.

5.2.1. Aims and hypotheses

The null hypothesis, that the mismatch identified between mating signal and mtDNA found in the Medway estuary is the result of hybridisation between *A. aestuarina* and *A. makarovi,* was tested. Mitochondrial DNA sequence data and a substantial number of nuclear AFLP markers were analysed with the following aims: 1) to identify whether the mismatch identified between mating signals and mtDNA is also present in analyses of nDNA, 2) whether there are any intermediate AFLP genotypes indicating recent hybridisation or introgression of nDNA in addition to mtDNA, and 3) how widespread across the UK this phenomenon is.

5.3. Materials and Methods

5.3.1. Sample collection, species identification and DNA extraction

Specimens were collected across England and Wales from June-August during the years 2005-2009, using a converted leaf blower (D-vac suction sampler, Electrolux, BVM 250). Specimens were stored in absolute ethanol at -80°C either directly from the field or after male vibrational mating signals had been recorded (Chapter 3). Legs of *A. makarovi* and *A. aestuarina* specimens were dissected and genomic DNA was extracted using DNeasy Tissue Kits (Qiagen). A sample of 265 *A. makarovi* and *A. aestuarina* individuals, including 18 known mismatched specimens collected from the Medway estuary (Fig. 5.1, Table 5.1), were used for AFLP analyses (with an additional five positive control repeats). Mismatched specimens were those that emitted the male mating signal of *A. aestuarina* (or females responded to in signal playback tests) but possessed *A. makarovi* mtDNA (Chapter 3). A map of the Medway estuary is shown in Fig. 5.2. Species identification was carried out using either COI or bioacoustic methods, or both (Chapter 3). Geographical coordinates (GPS) can be found in section 5.8, Appendix I).

The samples sizes for AFLP analysis included between 8-74 individuals per location (Table 5.1). As sample sizes for Essex populations were low these were treated as one regional group. Due to the genetic differentiation between sympatric inland and estuarine *A. makarovi* populations (Chapter 4), specimens from Gower and Norfolk populations inhabiting these different environments were treated as separate populations.

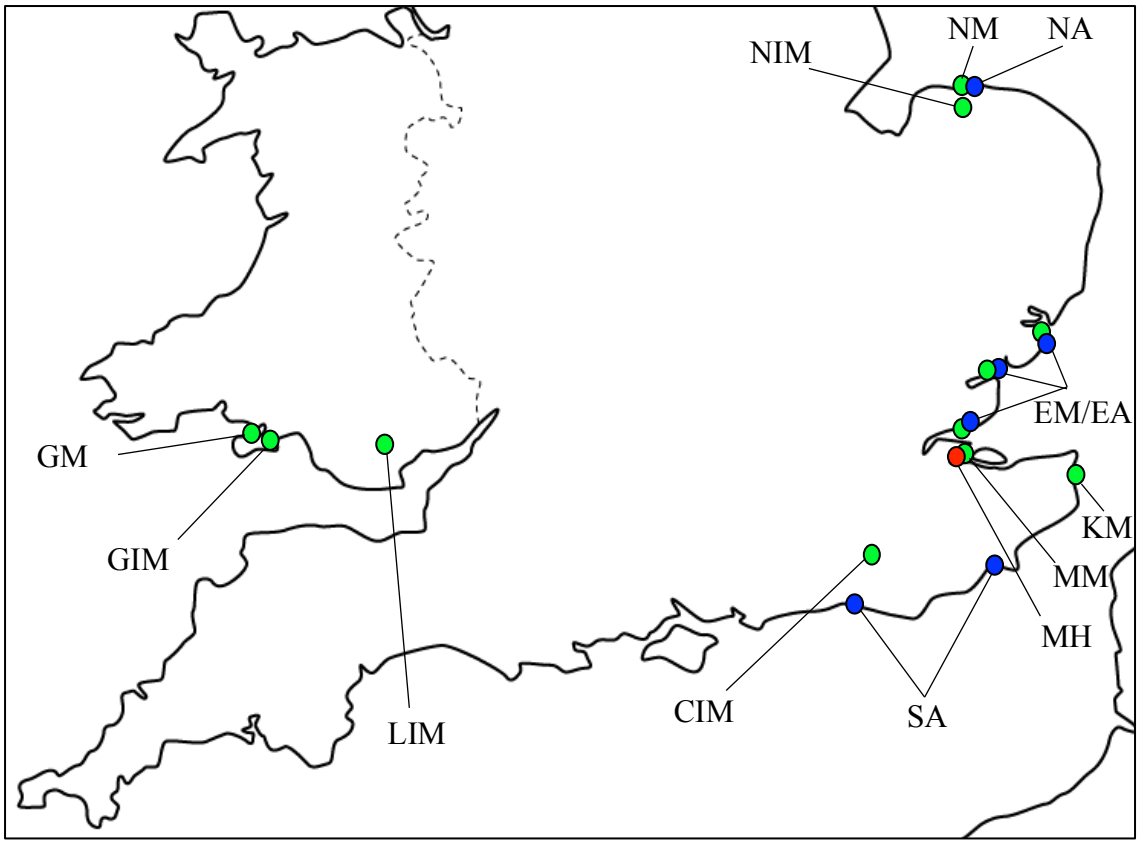


Figure 5.1. Sampling locations for *Aphrodes makarovi* and *Aphrodes aestuarina* in Wales and England used for amplified fragment length polymorphism analyses. Green circles indicate *A. makarovi* populations (~M) and blue circles represent *A. aestuarina* populations (~A) and the red circle represents where mismatched specimens were identified in the Medway estuary (MH) that emitted the male mating signal of *A. aestuarina* (or females responded to in signal playback tests) but possessed *A. makarovi* mitochondrial DNA (Chapter 3). Marginally overlapping circles indicate where more than one species is found in sympatry (see Table 5.1 for sampling site abbreviations). All locations are saltmarsh habitats (primary host plant *Atriplex portulacoides*) except for four inland *A. makarovi* sites (~I~) (primary host plant *Urtica* sp.).

Figure 5.2. Map of the Medway estuary sampling localities. Green circles are locations where *Aphrodes makarovi* was identified based on vibrational mating signal and mitochondrial DNA (mtDNA) analyses (Chapter 3). Red circles are where specimens were recorded for *A. aestuarina* mating signal but possessed *A. makarovi* mtDNA. Red and green indicated where both *A. makarovi* and mismatched specimens were identified in sympatry. White circles show where *A. makarovi* was identified based on mtDNA only. See Appendix I for GPS coordinates for locations R-R12.

Table 5.1. Information for each of the 10 sampled populations, including host plant, habitat type, location name, species present based on mating signal and/or mitochondrial cytochrome oxidase subunit I gene sequence data (Chapter 3), the number of male and female (M/F) specimens from each location of each species used in amplified fragment length polymorphism (AFLP) analyses. * *mismatch* = specimens collected from the Medway estuary that emitted the male mating signal of *A. aestuarina* (or females responded to in signal playback tests) but possessed *A. makarovi* mitochondrial DNA (Chapter 3).

Host plant / Habitat type	**Location**	**Species**	**Abbreviation**	***n* AFLP (M/F)**
Atriplex - saltmarsh	Medway	* *mismatch*	MH	18 (11/7)
		A. makarovi	MM	74 (47/27)
	Essex	*A. makarovi*	EM	15 (8/7)
		A. aestuarina	EA	11 (7/4)
	Norfolk	*A. makarovi*	NM	20 (13/7)
		A. aestuarina	NA	20 (17/3)
	Sussex	*A. aestuarina*	SA	20 (9/11)
	Kent	*A. makarovi*	KM	8 (6/2)
	Gower	*A. makarovi*	GM	20 (10/10)
Urtica - grassland	Gower	*A. makarovi*	GIM	19 (7/12)
	Lisvane	*A. makarovi*	LIM	10 (8/2)
	Norfolk	*A. makarovi*	NIM	20 (11/9)
	Castle Hill	*A. makarovi*	CIM	10 (5/5)
				265 (143/122)

5.3.2. Mitochondrial DNA

5.3.2.1. Mitochondrial DNA sequencing protocol

The mitochondrial COI sequence dataset used in this chapter (GenBank Accession numbers; FR727167 – FR727170, FR727173 – FR727175, HE587029 – HE587045) was obtained using methods described in Chapter 3, consisting of 321 *A. makarovi* (n = 218, including 18 mismatched specimens from the Medway) and *A. aestuarina* (n = 103) individuals in total. The 710 base pair (bp) COI sequences were amplified using polymerase chain reaction (PCR) and universal invertebrate primers LCO1490 and HCO2198 (Folmer *et al.*1994). Products of PCR reactions were sequenced directly using the original PCR primers (Chapter 3) and BigDye (version 3.1) sequencing chemistry in both forward and reverse directions. Reactions were run by the Cardiff

University Molecular Biology Support Unit and analysed on an Applied Biosystems 3130x1 Genetic Analyser.

5.3.2.2. Mitochondrial DNA sequence alignment and analysis

Sequences were aligned using SEQUENCHER version 4.9 (Gene Codes). Mitochondrial sequence variation among *A. makarovi* and *A. aestuarina* and the Medway estuary mismatched specimens was visualised by constructing a median-joining network using NETWORK version 4.6 (Bandelt *et al.* 1999).

Population demographic expansion and decline can affect the pattern of genetic polymorphism, leaving characteristic signatures in the distribution of nucleotide site differences between individuals (Rogers & Harpending 1992). Patterns of distribution are usually multimodal for populations exhibiting equilibrium and unimodal for lineages that have undergone recent population expansions or a bottleneck (Rogers & Harpending 1992). Inference of population demographic history was assessed using mismatch distribution analysis under spatial expansion and sudden expansion models assuming a constant deme size (Rogers & Harpending 1992) as implemented in ARLEQUIN version 3.5.1.2 (Excoffier & Lischer 2010). The sum of squared deviation (Schneider & Excoffier 1999) and raggedness index (r, quantifies the smoothness of the observed mismatch distribution) tests of significance (Harpending 1994) were used to test the null hypothesis of population expansion. The fit of the observed mismatch distribution to models of expansion was tested with 10,000 permutations. Parameters estimated include tau (τ), the time to the expansion and the effective population sizes before and after the expansion, θ_0 and θ_1, respectively. Populations that have remained constant in size typically generate distributions with more numerous and ragged peaks whereas a population that has undergone recent population expansion will produce a smooth, unimodal distribution (Harpending 1994). Each species was tested as a single population due to the high level of genetic divergence between species and to avoid autocorrelation within the analysis by splitting each species by geographic populations.

Tests of selective neutrality were carried out in ARLEQUIN using 10,000 permutations. Tajima's D (Tajima 1989) and Fu's F_S (Fu 1997) tests are based on an infinite-site

model without recombination and thus suitable for short DNA sequences. Other than effects of selection, significantly negative D values can indicate signatures of population expansion (Aris-Brosou & Excoffier 1996) and the F_S statistic is known to be sensitive to population demographic expansions generally resulting in large negative F_S values (Fu 1997).

5.3.3. Amplified fragment length polymorphism (AFLP)

5.3.3.1. AFLP protocol

AFLP analysis was performed according to Vos *et al.* (1995) with modifications described in Chapter 4. The two restriction enzymes used were *Pst*I and *Eco*RI and three primer combinations PAAA/E44, PAAG/E42 and PAAT/E35 were amplified (Chapter 4). *Eco*RI primers were labelled with a 6-FAM fluorescent tag and reactions were run by the Cardiff University Molecular Biology Support Unit and analysed on an Applied Biosystems 3130x1 fragment analyser.

5.3.3.2. Scoring and error rates

Electropherogram trace files were imported into GENEMARKER version 1.95 (SoftGenetics). GeneScan ROX-500 size standards were applied to the project and manually checked for quality and edited where required. Poor quality profiles (failed amplification) were removed from subsequent analyses. All peaks above 150 rfu (peak height identified as a suitable background noise threshold) and between 50-500 bp were scored using GENEMARKER. A panel was created automatically using all samples. Bin positions were manually checked to identify incorrect bin positioning and low quality or noise peaks (irregular shape or pull-ups). Overlapping bin positions were deleted from the data set to avoid ambiguous scoring due to possible size homoplasmy of co-migrating fragments (Vekemans *et al.* 2002). PCR negatives were checked for possible contaminants and any peaks above the background noise threshold were deleted from the respective primer combination.

AFLPSCORE version 1.4b (Whitlock *et al.* 2008) was used to identify thresholds (relating to average locus peak height and relative peak height across all loci) that resulted in acceptable mismatch error rates (< 5%) but maximised the number of AFLP markers retained for further analysis. Mismatch error rates (Bonin *et al.* 2004), based on five repeated genotype profiles were calculated using the data filtering option, a locus selection threshold of 400 rfu (18% of the total mean normalised peak height across all loci) and a relative phenotype calling threshold of 150 rfu (7% of the total mean normalised peak height across all loci). A binary matrix of retained AFLP markers was created in AFLPSCORE for the three primer combinations and a subset were compared to the original electropherograms to check for computational copying errors.

5.3.3.3. Identification of individual hybrids

The nuclear genetic diversity was calculated assuming Hardy-Weinberg equilibrium (HWE) as the average expected heterozygosity, *He* (Nei's gene diversity) for each population. AFLP allele frequency estimates were obtained assuming HWE, using a Bayesian method with non-uniform prior distribution of allele frequencies (Zhivotovsky 1999) and genetic diversity statistics were computed following the approach of Lynch & Milligan (1994) with 10,000 bootstraps. Deviations from Hardy-Weinberg equilibrium assumption may be apparent due to presence of hybrids and thus the expected heterozygosity should be interpreted cautiously. All analyses were performed using AFLP-SURV version 1.0 (Vekemans *et al.* 2002).

Principal coordinate analysis (PCA) was performed in GENALEX 6.4 (Peakall & Smouse 2006) to visualise the AFLP nuclear data set comparing genotypes of Medway estuary specimens to those of *A. makarovi* and *A. aestuarina*. A standardised distance method was used based on Euclidean distances (calculated in ARLEQUIN). Population structure was inferred using two Bayesian model-based clustering methods as implemented in STRUCTURE version 2.2 (Pritchard *et al.* 2000) and NEWHYBRIDS version 1.1 Beta3 (Anderson & Thompson 2002) in an attempt to identify hybrid individuals of mixed ancestry (Bonin *et al.* 2007). An advantage of Bayesian methods is that they do not require reference to pure genotypes. AFLPOP version 1.1 (Duchesne & Bernatchez 2002) was also used which is a frequentist assignment simulation method but unlike those mentioned above, does require knowledge of reference specimens of pure genotypes,

for which information may not always be obtained easily. Another advantage of the Bayesian methods is that they accommodate uncertainties relating to the true genotypes within the model when using dominant data.

STRUCTURE assigns individuals into *K* genetic clusters using multi-locus data, without using any prior information regarding population origin and has been adapted to accommodate dominant data (Falush *et al.* 2007). The admixture model was used to estimate the proportion of each individual's genome that has descended from each source population (the proportion of ancestry). Ten independent runs for each value of *K* ranging from one to ten were performed with 1,000,000 Markov Chain Monte Carlo (MCMC) repetitions, following a burn-in of 250,000 and run on CONDOR (Litzkow *et al.* 1998) computational facility (ARCCA, Cardiff University). Runs were performed using both correlated and uncorrelated allele frequencies (Pritchard *et al.* 2000; Falush *et al.* 2003). The true number of clusters was determined using both the maximal log probability of the data, Pr($X|K$) reported by STRUCTURE and the ΔK method of Evanno *et al.* (2005) that calculates the rate of change in the log probability of data between successive *K*-values. Results were visualised using individual assignment values averaged over the 10 replicate runs. A > 0.9 threshold was chosen to identify pure parental specimens (when centred around 1 or 0). If likelihood values were centred around 0.5, then individuals are likely to represent F_1 generation hybrids and if likelihood values do not lie around 1, 0 or 0.5 they are most likely later generation backcrossed hybrids.

STRUCTURE results were then compared to those obtained using NEWHYBRIDS (Anderson & Thompson 2002). NEWHYBRIDS is a Bayesian model based method designed specifically to identify hybrids and computes the posterior probability that an individual belongs to one of six hybrid classes: pure *A. makarovi*, pure *A. aestuarina*, F_1, F_2, backcross (BC) to pure *A. makarovi* and BC to pure *A. aestuarina* categories. Jeffrey's and Uniform priors for the mixing proportion and the allelic frequencies were tested using a burn-in of 20,000 iterations followed by a further 100,000 iterations.

Training samples (pure genotypes) were introduced into the data set for each parent species as those with matching COI mtDNA and mating signals identified in previous analyses (Chapter 3). These training individuals were included as NEWHYBRIDS has a reduced ability to obtain reasonable results using AFLPs without training samples

(Anderson & Thompson 2002), due to the longer burn-in times required for convergence MCMC sampler and thus affecting the true allele estimates for each species. Due to the geographic and genetic distinctiveness of Wales inland and estuarine *A. makarovi* populations (Chapter 4), these populations were not included in the hybrid analyses as it is unlikely that if hybrids are present in the Medway estuary, that they were formed between a cross involving Welsh *A. makarovi.* However, Welsh specimens were included in the initial computation of allele frequencies, as advised (Anderson & Thompson 2002). All inland *A. makarovi* populations were also excluded from the hybrid analyses (but not computation of allele frequencies). Tests were also done including Welsh and inland populations in the hybrid analysis but this did not affect the outcome. Final analyses were run by including all specimens in the computation of allele frequencies and excluding all Wales and inland populations from hybrid analyses.

AFLPOP was used to simulate 1000 pure parental genotypes from reference samples (identified as pure individuals from Bayesian clustering analyses using NEWHYBRIDS and STRUCTURE) and to simulate genotypes relating to the six hybrid classes (pure *A. makarovi,* pure *A. aestuarina,* F_1, F_2, BC to *A. makarovi* and BC to *A. aestuarina*). Ten replicated simulations were conducted and a log likelihood difference of zero was used so that all individuals were assigned to a class. The success rate of assignment of the simulated populations to the six classes can be used as a measure of how well the data set can discriminate between different categories of hybrids. A total of 80 pure *A. aestuarina* genotypes (44 *A. aestuarina* and 33 mismatched Medway estuary specimens) and 90 pure *A. makarovi* genotypes were chosen to represent pure parental populations. These individuals all showed a > 0.9 probability of either being pure *A. makarovi* or *A. aestuarina* in all Bayesian clustering analyses (section 5.4.2.1). Due to the large number of *A. makarovi* specimens in the total data set compared to *A. aestuarina* the 90 pure *A. makarovi* were chosen from estuarine populations along the east coast of England (omitting those from all inland (55 specimens) and Gower estuarine populations (18 specimens)). Simulation tests were also conducted using the total data set (162 pure *A. makarovi* specimens), which gave concordant results. Re-allocation of reference genotypes was also performed, in which each genotype is removed from the computation of frequencies within its known population and assigned as an unknown. The log likelihood differences were then assessed to identify individuals showing a low log likelihood difference and thus questionable assignment.

5.3.3.4. Population structure of Aphrodes aestuarina *and the mismatched Medway estuary population*

Further analysis was done to identify genetic differentiation of nuclear AFLP markers among *A. aestuarina* populations and the mismatched Medway estuary specimens using STRUCTURE. The population structure among *A. makarovi* populations was explored in Chapter 4. PCA analysis was carried out as described in section 5.3.3.3 using Euclidian distances calculated in ARLEQUIN comparing *A. aestuarina* populations with the mismatched Medway estuary population.

To identify further levels of genetic structuring in the dataset using STRUCTURE, all *A. aestuarina* and Medway estuary mismatched specimens, which share the same mating signal and cluster in previous Bayesian analyses, were analysed as a single data set using correlated and uncorrelated allele frequencies for K=1-5 using methods described in section 5.3.3.3. Subsequently all non-mismatched *A. aestuarina* specimens and the mismatched Medway population were analysed separately based on the prior knowledge of a mismatch (or lack of) between AFLP markers and mtDNA to identify any further genetic differentiation within these groups.

The number of private AFLP bands for each population was calculated using GENALEX based on a Euclidean distance matrix. Due to the possibility of nDNA introgression (as well as mtDNA), private bands were identified by comparing the mismatched Medway estuary population to other *A. aestuarina* populations, and to the whole data set including all *A. makarovi* populations. This was necessary to identify the proportion of private alleles in the Medway estuary population that are shared with *A. makarovi* but not with *A. aestuarina*.

Problems arise with traditional methods to evaluate genetic differentiation among populations using dominant markers, compared with codominant markers. Therefore the hierarchical Bayesian approach implemented in HICKORY version 1.1 (Holsinger *et al.* 2002) was used, which does not assume any prior knowledge of the degree of inbreeding within populations. The $\theta^{(\Pi)}$ parameter, analogous to F_{ST} (based on Weir & Cockerham's approach, 1984), was estimated for each pairwise population comparison. Results were averaged over three runs using default settings. The method most applicable to dominant data is the f-free model that does not attempt to estimate F_{IS} due

to its unreliability when estimated using other models (Holsinger & Wallace 2004). Values derived for the deviance information criterion (DIC) calculated from the full model (which attempts to estimate F_{IS}), $\theta^{B} = 0$ (model that assumes no differentiation among populations) and the $F_{IS} = 0$ model (assuming no inbreeding among populations) were used to identify how well each model fitted the data (a better fit results in smaller DIC values).

To evaluate how genetic variation was distributed a hierarchical analysis of molecular variance (AMOVA) was performed in ARLEQUIN. AMOVA analyses were carried out using Euclidian distances calculated in ARLEQUIN, specifying either no population structure (which relates to the four regional populations, namely Norfolk, Sussex, Essex and Medway) or structure relating to non-mismatched *A. aestuarina* populations versus the mismatched Medway population (two populations), or based on Bayesian clustering results down to the regional scale (Norfolk, Medway, Sussex/Essex, three populations).

The correlation between genetic and geographical distance was examined with a MANTEL test in ARLEQUIN. Pairwise F_{ST} estimates from HICKORY were used (5.4.2.2) and compared to pairwise population geographic distances (calculated as average coastal distances between localities).

5.4. Results

5.4.1. Mitochondrial DNA

Variation within the 658 bp (inter primer length) fragment of the COI gene revealed twenty-four unique mtDNA haplotypes for a total sample of 321 *Aphrodes* specimens (GenBank Accession numbers; FR727167 – FR727170, FR727173 – FR727175, HE587029 – HE587045). Seventeen haplotypes were identified for *A. makarovi* including specimens from the Medway estuary, and seven haplotypes were found in *A. aestuarina* (Chapter 3). For descriptive statistics and molecular diversity indices for each species see Chapter 3 (3.5.3). The percentage sequence divergence between *A. makarovi* and *A. aestuarina* lineages calculated in Chapter 3 was 6.93%, based on the Kimura-2-parameter (Kimura 1980) distance measure. Due to the significant sequence divergence between mtDNA lineages two separate networks were drawn (Fig. 5.3), corresponding to *A. makarovi* (including the Medway estuary specimens) and *A. aestuarina* lineages identified in previous phylogenetic analyses (Chapter 3).

At all localities except the Medway estuary the species mating signal that was recorded corresponded with the mtDNA haplotypes found (and AFLP clustering analyses, section 5.4.2.1). A 'star shaped' network was recovered for *A. makarovi* (Fig. 5.3) with the majority of specimens found to possess the most common *A. makarovi* haplotype (mH1), which was found at all *A. makarovi* sampling sites. A number of location-specific *A. makarovi* haplotypes (except haplotype mH14) differing by one or two bp from haplotype mH1 were found at low frequency. All *A. aestuarina* haplotypes from Norfolk, Essex and Sussex populations (possessing both *A. aestuarina* mating signals and mtDNA) also resulted in a 'star shaped' network, similar to that of *A. makarovi* but containing fewer haplotypes (Fig. 5.3). The main haplotypes are *A. aestuarina* haplotype aH1 (Sussex and some Essex specimens) and aH2 (found only in the Norfolk *A. aestuarina* population). Five relatively uncommon haplotypes were found, differing by between one and thee base pairs from *A. aestuarina* haplotype aH1. No mtDNA haplotypes were shared between Norfolk and the other *A. aestuarina* populations. Sampling maps showing the distribution of haplotypes for both species can be found in Chapter 3 (Fig. 3.10 c, d).

All specimens found at sites in the Medway estuary possessed *A. makarovi* mtDNA sequences (90 individuals in total, eight haplotypes). Of these, 18 mismatched specimens (11 males and seven females) were shown to produce or respond to *A. aestuarina* signals. Eight individuals were found to possess, or respond to, the *A. makarovi* mating signal (seven males and one female responded in playback tests). Mating signals were not recorded for a total of 64 individuals (38 males and 26 females). When results for the unrecorded specimens were compared to AFLP clustering results for specimens where the signal data was known, discrimination between *A. makarovi* and those showing a mismatch was possible (section 5.4.2.1). Of these unrecorded individuals, 24 grouped with those showing a mismatch between vibrational signal and mtDNA and the remainder clustered with *A. makarovi* (section 5.4.2.1). Based on these results, the mismatched specimens (42 individuals in total) either possessed the common *A. makarovi* haplotype mH1 (38 individuals) or the less frequent haplotypes mH7 (one individual), mH16 (one individual) and mH17 (two individuals). Haplotypes mH16 and mH17 were only found in the mismatched specimens in the Medway estuary region (Fig. 5.3).

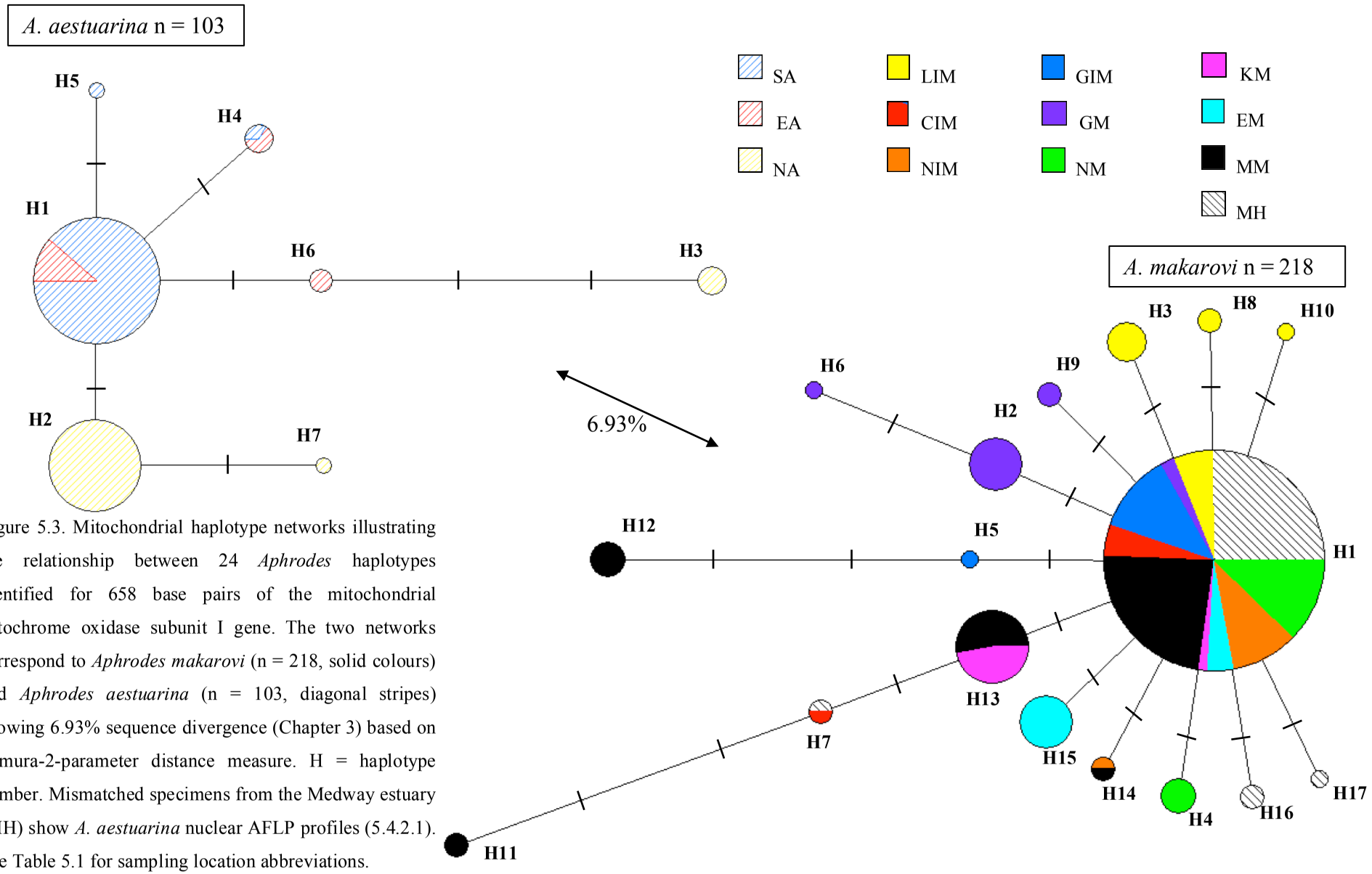


Figure 5.3. Mitochondrial haplotype networks illustrating the relationship between 24 *Aphrodes* haplotypes identified for 658 base pairs of the mitochondrial cytochrome oxidase subunit I gene. The two networks correspond to *Aphrodes makarovi* (n = 218, solid colours) and *Aphrodes aestuarina* (n = 103, diagonal stripes) showing 6.93% sequence divergence (Chapter 3) based on Kimura-2-parameter distance measure. H = haplotype number. Mismatched specimens from the Medway estuary (MH) show *A. aestuarina* nuclear AFLP profiles (5.4.2.1). See Table 5.1 for sampling location abbreviations.

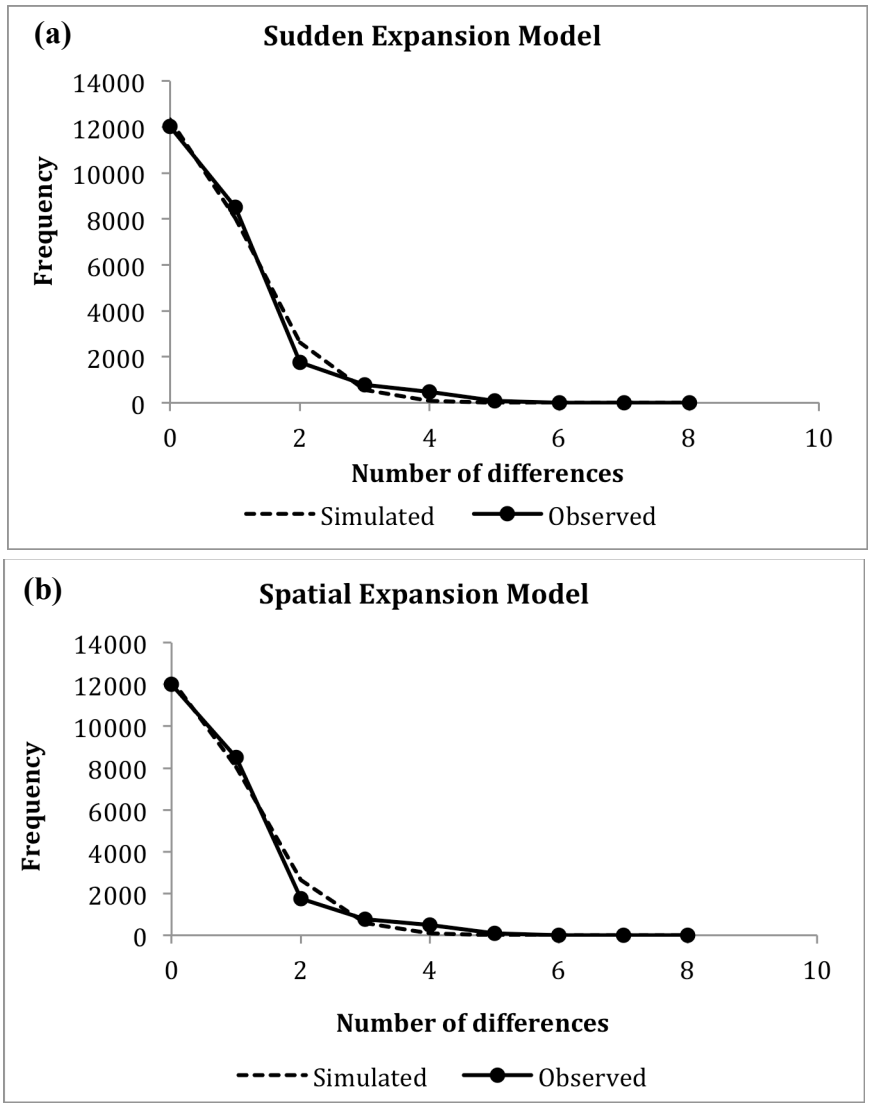


Figure 5.4. Frequency distributions of pairwise nucleotide differences between mitochondrial DNA sequences for individuals of *Aphrodes makarovi* across the UK, for (a) parameters estimated under the sudden expansion model and (b) parameters estimated under the spatial expansion model assuming constant deme size. Values shown are for Tau (τ), Theta*0* (θ_0), Theta*1* (θ_1), the model sum of squared deviation (SSD) and *p*-value and finally the Harpending's raggedness index and *p*-value for each model.

The frequency distribution estimates of pairwise nucleotide differences and parameter estimates from the mismatch analyses for *A. makarovi* and *A. aestuarina* are shown in Fig. 5.4 and 5.5, respectively. For *A. makarovi,* the frequency distribution does not show a signature of population growth (bell-shaped). The sum of squared deviation (SSD) and raggedness index under the sudden expansion model (Fig. 5.4a) were not statistically significant (*p*-value = 0.16 and 0.41 respectively) and thus the null hypothesis of a sudden expansion could not be rejected. Similarly, the null hypothesis of spatial expansion assuming constant deme size (Fig. 5.4b) could not be rejected (SSD *p*-

value = 0.07 and raggedness index *p*-value = 0.43). The opposite is seen for *A. aestuarina* where both sudden expansion and spatial expansion assuming a constant deme size the *p*-values for both models (SSD *p*-value < 0.05 and < 0.0001 respectively) and raggedness indexes (*p*-value < 0.01 and < 0.05 respectively) were significant (Fig. 5.5a, b), thus rejecting the null hypothesis of expansion for both models. For both species values θ_1 and M values are exceptionally large suggesting that there were no recent coalescent events and thus the effective population size cannot be estimated accurately with the current available data (L. Excoffier personal communication).

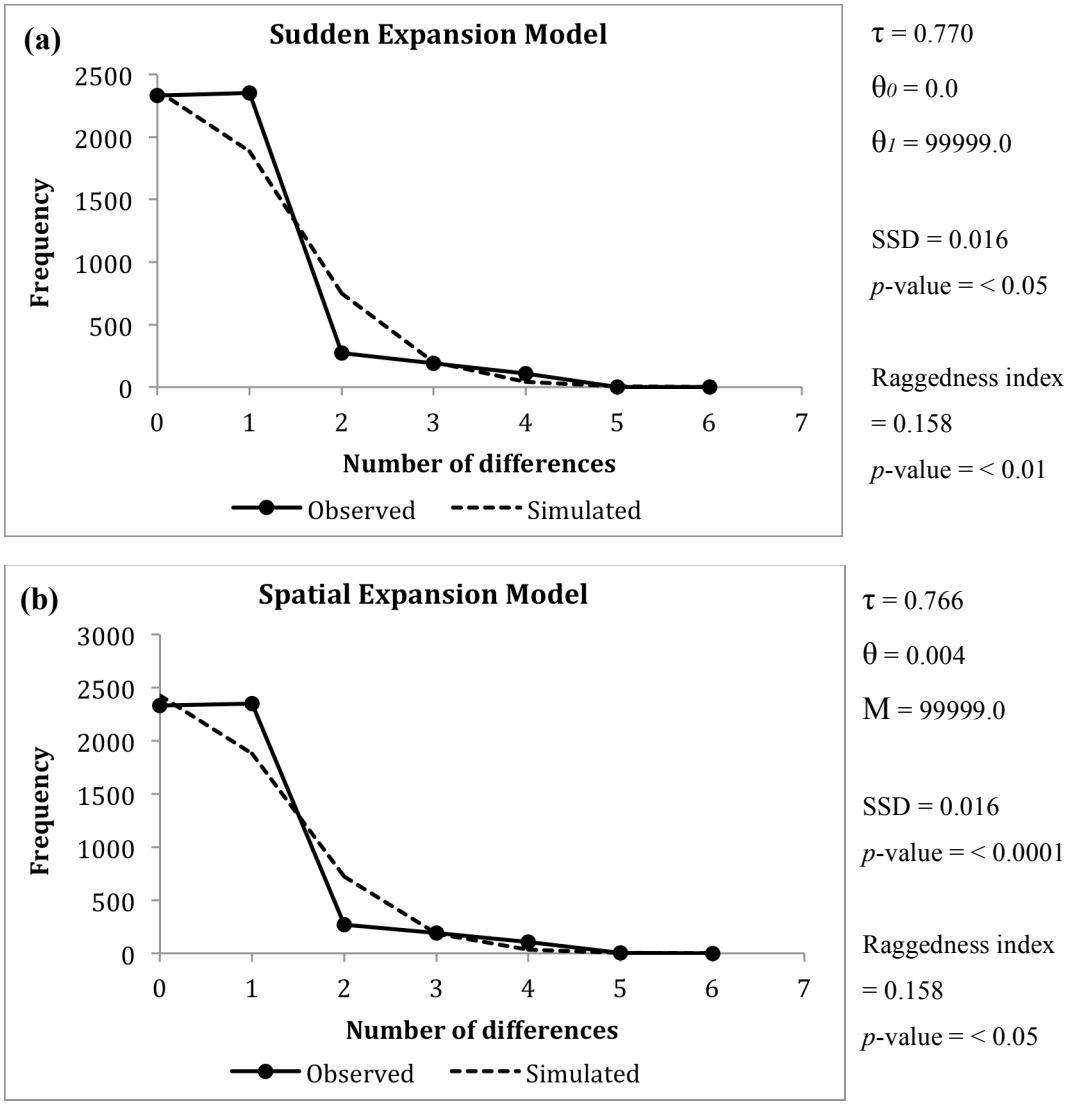


Figure 5.5. Frequency distributions of pairwise nucleotide differences between mitochondrial DNA sequences for individuals of *Aphrodes aestuarina* across the UK, for (a) parameters estimated under the sudden expansion model and (b) parameters estimated under the spatial expansion model assuming constant deme size. Values shown are for Tau (τ), Theta*0* (θ_0), Theta*1* (θ_1), the model sum of squared deviation (SSD) and *p*-value and finally the Harpending's raggedness index and *p*-value for each model.

For *A. makarovi,* tests of selective neutrality, both Tajima's D and Fu's F_S were significant (*p*-value < 0.05 and < 0.0001, respectively) (Table 5.2) and the test statistic values were negative suggesting a signature of population expansion or selection in this species which is concordant with mismatch distributions, network shape and shallow phylogenetic structure seen in this species. For *A. aestuarina* both the test statistics were negative. However, both *p*-values were non-significant and so signatures of population expansion or selection cannot be inferred from this result. This is also concordant with mismatch distribution results in this species.

Table 5.2. Tajima's D (Tajima 1989) and Fu's F_S (Fu 1997) selective neutrality test results, including the test statistic and *p*-value for *Aphrodes makarovi* and *A. aestuarina*, based on 658 base pairs of the mitochondrial cytochrome oxidase subunit I (COI) gene.

	Tajima's D	***p*-value**	**Fu's F_S**	***p*-value**
A. makarovi	-1.938	< 0.05	-15.198	< 0.0001
A. aestuarina	-1.043	0.155	-2.214	0.148

5.4.2. Amplified fragment length polymorphism (AFLP)

A total of 12 samples amplified poorly and were removed from the AFLP dataset giving a total of 253 individuals (with an additional five positive control repeats). A total of 613 AFLP markers, amplified from three primer combinations, scored between 75-430 bp (PAAA+E44 and PAAT+E35) and 50-500 bp (PAAG+E42). Mismatch error rates were calculated as 2.5% using AFLPSCORE, a value similar to those found in the literature (Ajmone-Marsan *et al*. 1997; Bonin *et al.* 2004), and the locus selection threshold (< 400rfu) identified a total of 32 fragments to be removed due to having a low average peak height. After removal of unreliable amplifying fragments and those present in only one individual a total of 554 AFLP markers were retained for analysis. The numbers of polymorphic loci were calculated overall (between 22.7% and 44.2% polymorphic loci) for each species for each population analysed (for *A. makarovi* between 26.5 – 44.2% and *A. aestuarina* 22.7 – 31.4% polymorphic loci and for the Medway estuary 24.5%, Table 5.3). The heterozygosity estimates were similar across

all populations ranging between 0.08 and 0.105 (Table 5.3) and the mean number of bands per individual was 63 (range 39-88, SD 8.7).

Table 5.3. Data for populations of *Aphrodes makarovi* (~M saltmarsh habitat, ~IM inland habitat) and *A. aestuarina* (~A, all estuarine habitat) and the Medway estuary mismatched population (MH), including the population, number of specimens (*n*), the mean expected heterozygosity (*He*) and percentage of polymorphic loci (*P*) (calculated in AFLP-SURV, Vekemans *et al.* 2002). Mismatched specimens were identified based on comparison between nuclear amplified fragment length polymorphism (AFLP) clustering results and mitochondrial DNA. For sampling location abbreviations see Table 5.1.

Species	**Population**	***n***	***He* ± SE**	***P* (%)**
A. makarovi	KM	8	0.105 ± 0.006	35.4
	EM	14	0.097 ± 0.006	26.5
	GM	18	0.099 ± 0.006	31.9
	NM	20	0.099 ± 0.006	31.8
	MM	48	0.098 ± 0.006	30.5
	GIM	16	0.103 ± 0.005	31.0
	NIM	20	0.101 ± 0.006	36.6
	LIM	10	0.105 ± 0.006	44.2
	CIM	9	0.095 ± 0.006	34.8
A. aestuarina	NA	20	0.080 ± 0.006	22.7
	SA	18	0.085 ± 0.006	23.8
	EA	10	0.087 ± 0.006	31.4
mismatched	MH	42	0.089 ± 0.006	24.5

5.4.2.1. Identification of individual hybrids

Results of PCA clustering analysis indicate two genetic clusters corresponding to *A. makarovi* and *A. aestuarina* that are separated along PC1, which explains 56.99% of the total variation found in the AFLP data (Fig. 5.6). At all locations except the Medway estuary, the mtDNA haplotypes found correspond to the AFLP PCA results. This suggests that the mismatch identified between analyses, and thus the likelihood of hybridisation is restricted to the Medway estuary population. A total of 42 individuals from the Medway estuary show a mismatch between mtDNA and nDNA suggesting that these represent the total number of mismatched specimens from the sample tested (shown in black in Fig. 5.6). All mismatched specimens possessing *A. makarovi* mtDNA cluster with *A. aestuarina* along PC1 (Fig. 5.6). There are no intermediate genotypes in the centre of the PCA plot between the two clusters showing that the presence of early generation hybrids (F_1/F_2) in the sample is unlikely. The majority of specimens showing a mismatch between mtDNA and nDNA separate from non-mismatched *A. aestuarina* along PC3 (explaining 9.18% of the total variation), with some overlap, mainly at the centre of these two groups. The remaining 48 specimens from the Medway estuary cluster with *A. makarovi,* concordant with mtDNA results. PC2 explained 11.28% of the variation in the data and separated inland and estuarine *A. makarovi* into clusters (Chapter 4), but *A. aestuarina* and the mismatched Medway specimens show no clustering along this axis. A map of the Medway estuary is shown in Fig. 5.7 showing the distribution of *A. makarovi* and mismatched specimens in this region.

Results for Bayesian clustering analyses in STRUCTURE for correlated and uncorrelated allele frequencies for $K = 2$ are shown in Fig. 5.8 and 5.9, respectively (Appendix II (1) for the maximal log probability of the data, $\Pr(X|K)$, and the ΔK results). All *A. makarovi* individuals (except one from the estuarine Gower population) cluster as a single group at the > 0.9 probability threshold for correlated and uncorrelated allele frequencies, suggesting that admixture of *A. aestuarina* nDNA into *A. makarovi* is highly improbable. Of the 42 individuals that show a mismatch between mtDNA and AFLP PCA results using correlated allele frequencies, eight individuals were admixed using the > 0.9 probability threshold (Fig. 5.8). The proportions of admixture from the *A. makarovi* population are low, ranging between 0.12 and 0.34. Also, using correlated allele frequencies, four typical *A. aestuarina* individuals were not classified as pure *A.*

aestuarina at the > 0.9 probability threshold with low probabilities of admixture from *A. makarovi* ranging from 0.13 to 0.23. These four individuals were not found at a single sampling location but were collected from Norfolk (n= 1), Essex (n = 2) and Sussex (n = 1) *A. aestuarina* populations. Using uncorrelated allele frequencies all *A. aestuarina* and mismatched Medway specimens show a high probability (> 0.9) of belonging to a single cluster (Fig. 5.9).

Results from the Bayesian analyses using NEWHYBRIDS with Jeffreys priors (Fig. 5.10) and Uniform priors (Fig. 5.11) suggests that all typical *A. makarovi* are pure at the > 0.9 probability threshold (except for one individual from the Gower estuarine population, Fig. 5.11). Results for Jeffreys priors show nine possible hybrid individuals with a > 0.1 likelihood (between 0.14 and 0.83) of belonging to 1x backcross hybrid class (Fig. 5.10), six of which show > 0.5 probability of being a 1x backcross to *A. aestuarina.* The remaining 33 individuals showing a mismatch between mtDNA and nDNA give a > 0.9 probability of being pure *A. aestuarina*. Results for typical *A. aestuarina* specimens detected the same four individuals identified previously as admixed by STRUCTURE, with a probability of between 0.22 and 0.66 of belonging to the 1x backcross to *A. aestuarina* hybrid class. Fewer, but the same admixed specimens were identified using Uniform priors (Fig. 5.11).

Both STRUCTURE and NEWHYBRIDS results suggest that significant introgression is unlikely in the mismatched Medway estuary population (or any other). Overall, STRUCTURE identified a lower number of admixed individuals compared to NEWHYBRIDS, which has been shown previously when dealing with later generation backcrossed hybrids (Vähä & Primmer 2006). Regardless of this fact, NEWHYBRIDS still did not identify a large proportion of the mismatched specimens as having a high probability of admixture. Using correlated allele frequencies in STRUCTURE or Jeffrey's priors in NEWHYBRIDS yielded a higher likelihood of admixture compared with uncorrelated allele frequencies or Uniform priors respectively. The same individuals were identified as admixed using both of the methods tested, but reveal varying degrees of admixture from the *A. makarovi* gene pool in the mismatched specimens.

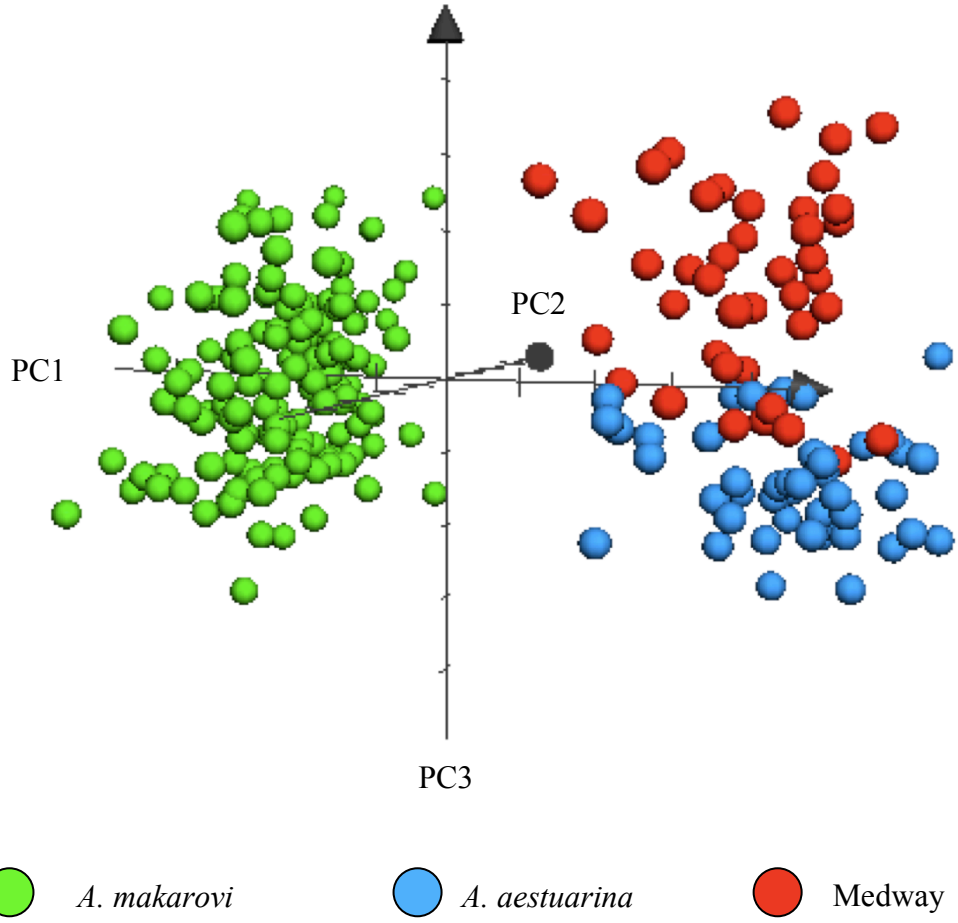


Figure 5.6. Principle coordinates analysis results based on Euclidean distances between amplified fragment length polymorphism (AFLP) multilocus phenotypes for 163 *Aphrodes makarovi* (green), 48 *A. aestuarina* (blue) and 42 mismatched specimens (red; possessing *A. makarovi* mitochondrial DNA) from the Medway estuary. 3D plot illustrating the first three principle coordinates (PC1 = 56.99%, PC2 = 11.28%, PC3 = 9.18% variation explained) for 554 AFLP loci.

Figure 5.7. Map of the Medway estuary (Kent, UK) sampling localities. Green circles are locations where *Aphrodes makarovi* was identified based on mitochondrial DNA (mtDNA) and amplified fragment length polymorphism (AFLP) data analyses. Red circles are where mismatched specimens were identified, which possess *A. makarovi* mtDNA but cluster with *A. aestuarina* AFLP genotypes. Red and green indicated where both *A. makarovi* and mismatched specimens were identified in sympatry.

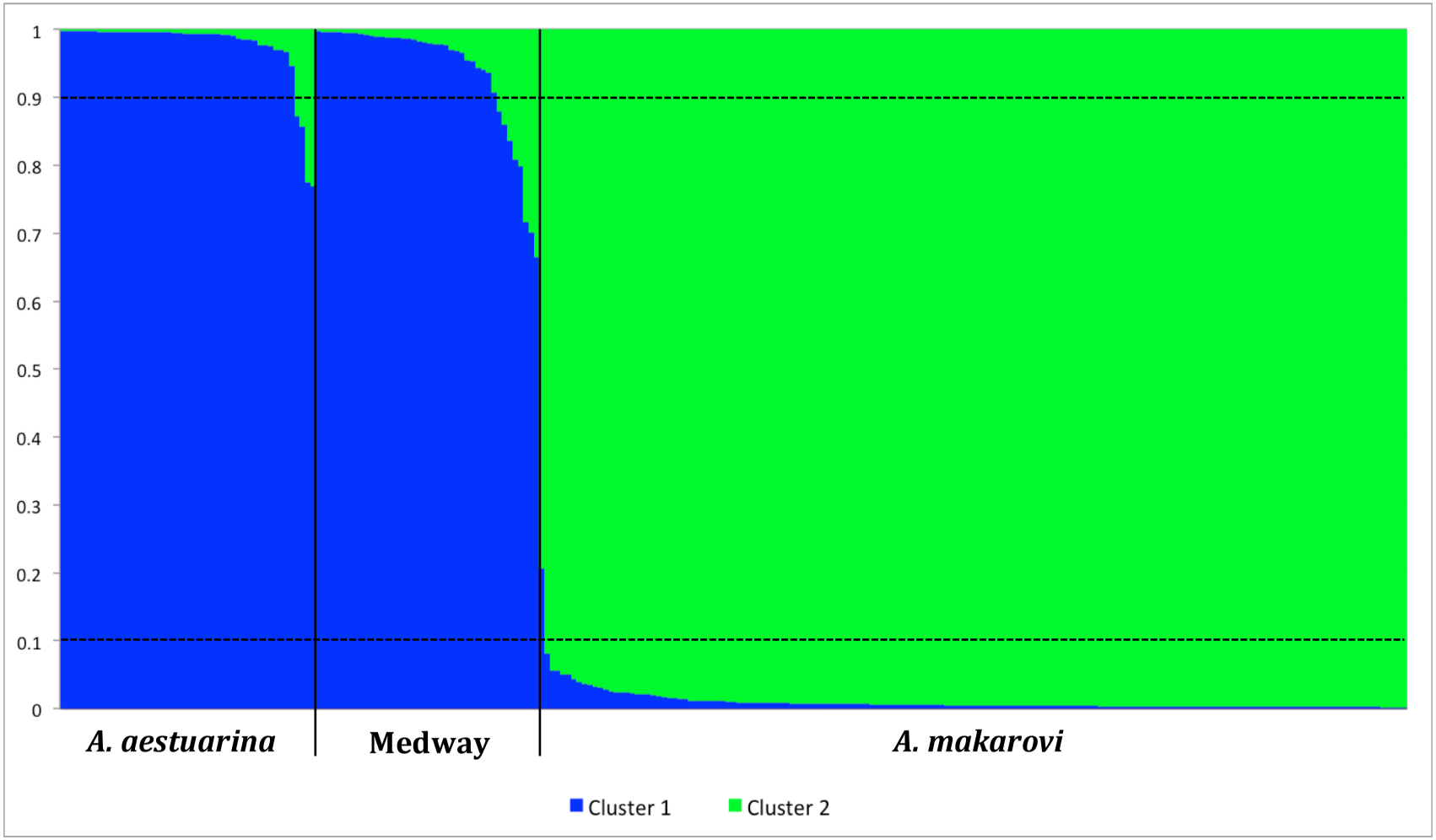


Figure 5.8. Results of STRUCTURE (Pritchard *et al.* 2000) clustering analysis using correlated allele frequencies, showing the assignment of 253 *Aphrodes makarovi* and *A. aestuarina* specimens (each thin vertical bar) to either cluster one (blue) or two (green) (*K*=2). Mismatched individuals from the Medway estuary were shown to possess *A. makarovi* mitochondrial DNA (5.4.1). Results are based on 554 polymorphic amplified fragment length polymorphism loci. Horizontal dashed lines represent 0.9 probability threshold for each cluster. Individuals were sorted according to their assignment values.

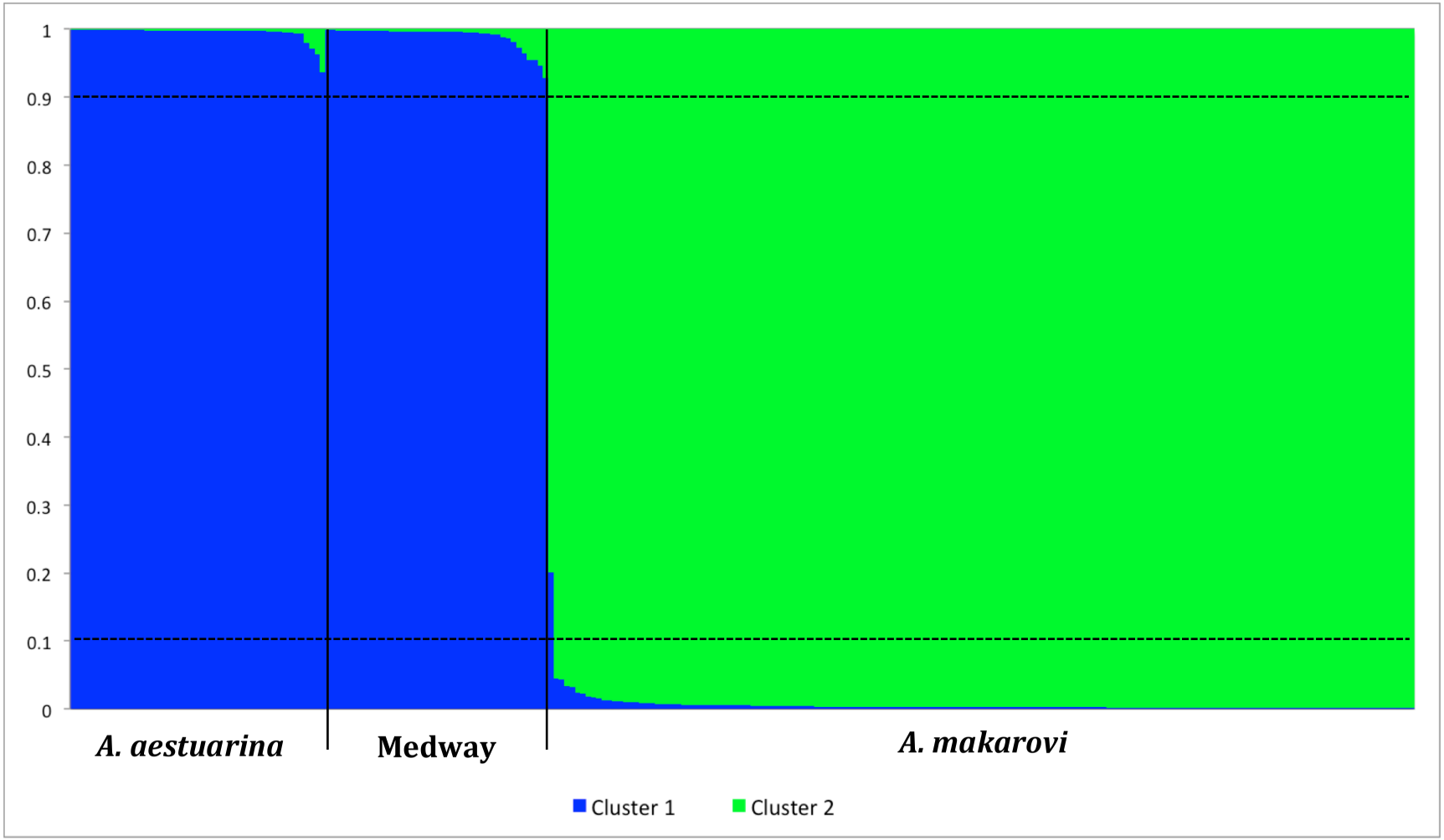


Figure 5.9. Results of STRUCTURE (Pritchard *et al.* 2000) clustering analysis using uncorrelated allele frequencies, showing the assignment of 253 *Aphrodes makarovi* and *A. aestuarina* specimens (each thin vertical bar) to either cluster one (blue) or two (green) (*K*=2). Mismatched individuals from the Medway estuary were shown to possess *A. makarovi* mitochondrial DNA (5.4.1). Results are based on 554 polymorphic amplified fragment length polymorphism loci. Horizontal dashed lines represent 0.9 probability thresholds for each cluster. Individuals were sorted according to their assignment values.

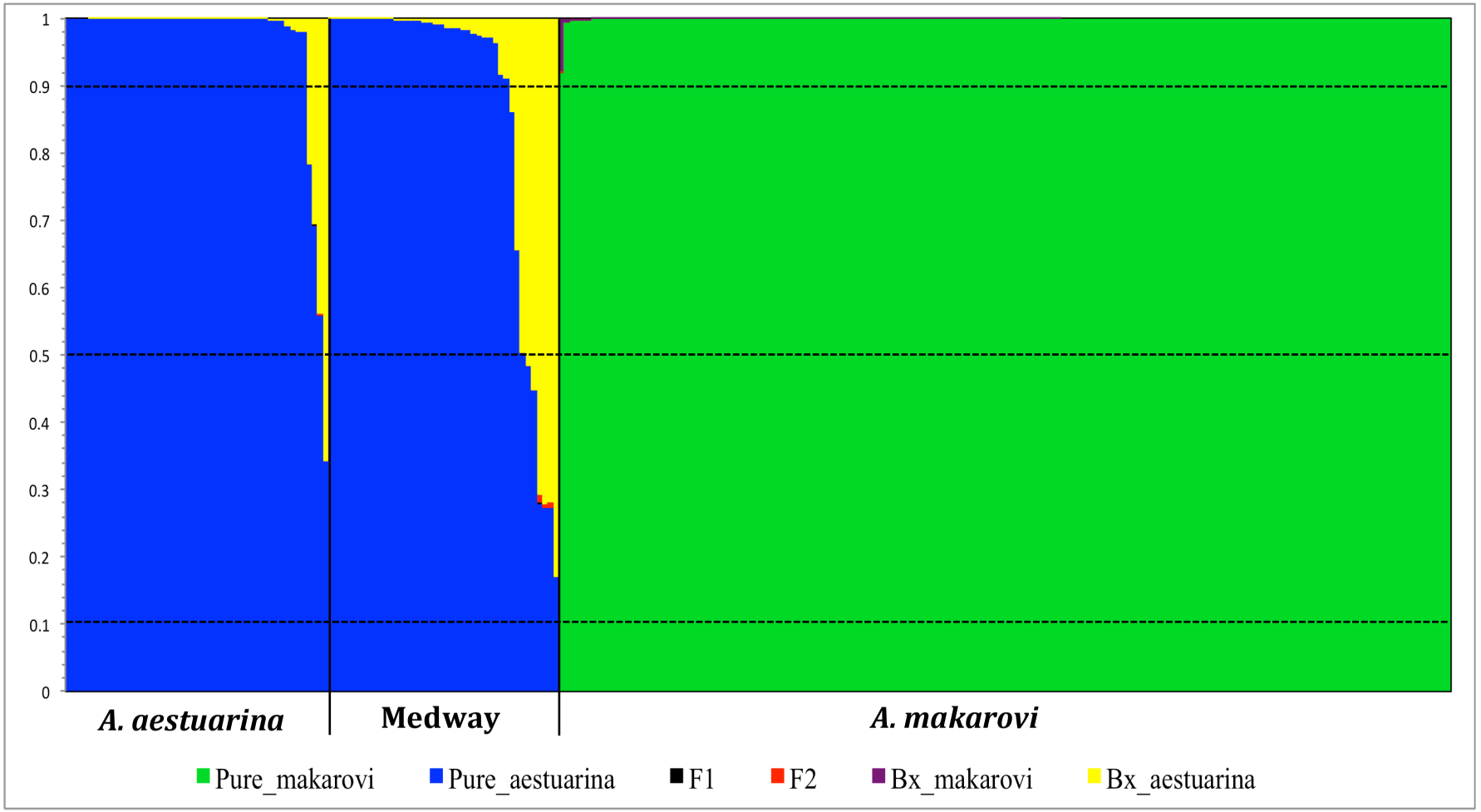


Figure 5.10. Results of NEWHYBRIDS (Anderson & Thompson 2002) clustering analysis using Jeffreys priors, showing the assignment of 253 *Aphrodes makarovi* and *A. aestuarina* specimens (each thin vertical bar) to one of six assignment classes (pure *A. makarovi*, pure *A. aestuarina,* F_1, F_2, first generation hybrid backcross (bx) to *A. makarovi* and bx to *A. aestuarina*). Mismatched individuals from the Medway estuary were shown to possess *A. makarovi* mitochondrial DNA (5.4.1). Results are based on 554 polymorphic amplified fragment length polymorphism loci. Horizontal dashed lines represent 0.9 and 05 probability thresholds for each cluster. Individuals were sorted according to their assignment values.

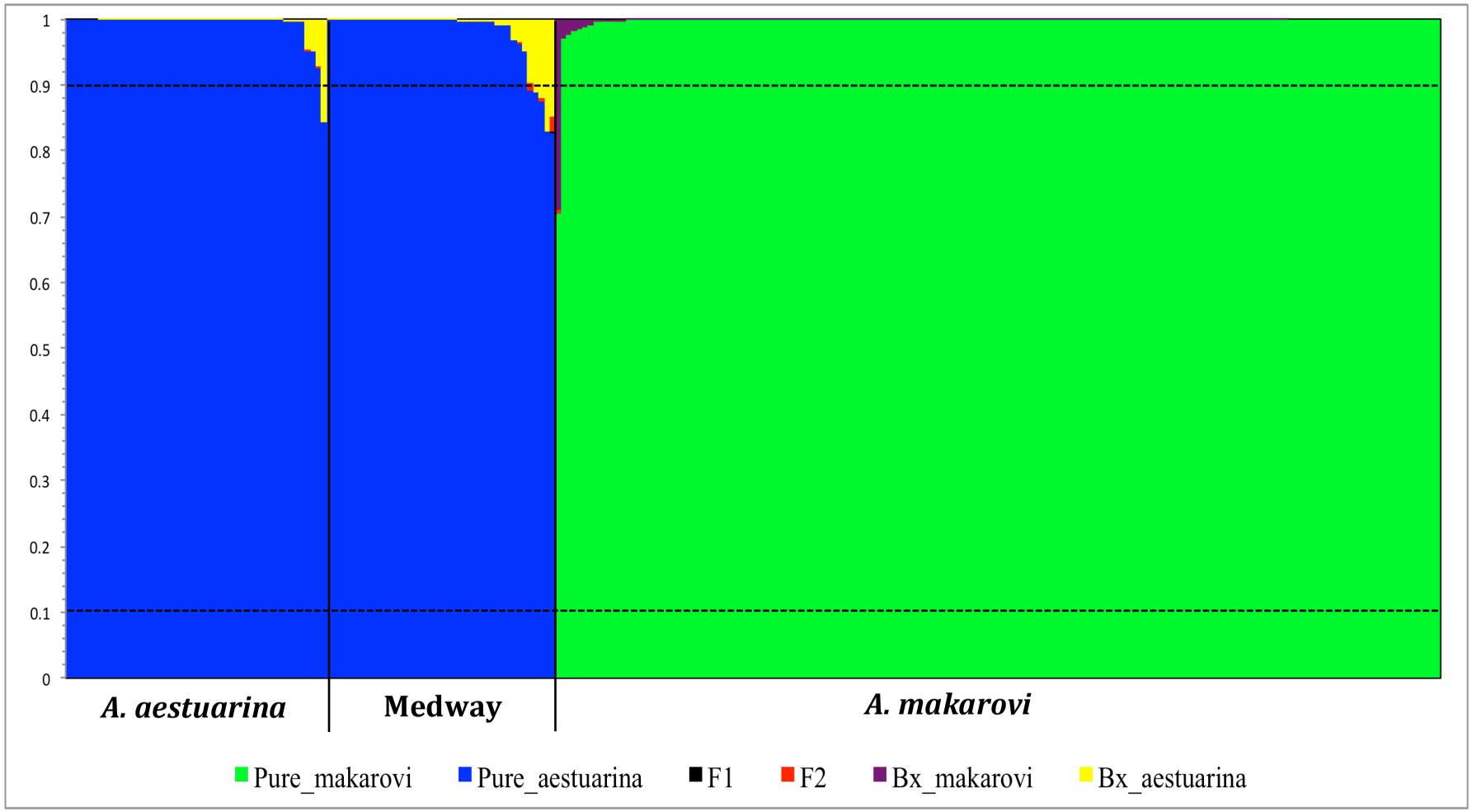


Figure 5.11. Results of NEWHYBRIDS (Anderson & Thompson 2002) clustering analysis using Uniform priors, showing the assignment of 253 *Aphrodes makarovi* and *A. aestuarina* specimens (each thin vertical bar) to one of six assignment classes (pure *A. makarovi*, pure *A. aestuarina,* F_1, F_2, first generation hybrid backcross (bx) to *A. makarovi* and bx to *A. aestuarina*). Mismatched individuals from the Medway estuary were shown to possess *A. makarovi* mitochondrial DNA (5.4.1). Results are based on 554 polymorphic amplified fragment length polymorphism loci. Horizontal dashed lines represent 0.9 probability thresholds for each cluster. Individuals were sorted according to their assignment values.

AFLPOP was used to simulate and assign 1000 genotypes of the six genotypic classes generated from reference genotypes for each parent species (80 *A. aestuarina* specimens from Norfolk, Essex, Sussex and mismatched Medway estuary population and 90 *A. makarovi* specimens), identified as pure in Bayesian clustering analyses. Results for the assignment of pure genotypes were ~99% accurate for both parent species (Table 5.4). Simulated hybrid genotypes were assigned as having hybrid ancestry in > 98% of cases but there was poor discrimination between F_1 and F_2 generation hybrids (Table 5.4). Assignment of simulated backcross hybrid genotypes was > 88% accurate. Re-assignment of actual pure source population genotypes was 100% accurate which gives an indication of the precision of allele frequency estimates and all log likelihood difference values were greater than 20 (log likelihood value of 1 indicates that the allocation to a certain population was 10 times more probable than to another one).

All results indicate very low levels of *A. makarovi* nuclear introgression in the mismatched Medway estuary population thus hybridisation and the subsequent introgression of *A. makarovi* mtDNA into this population is unlikely to be very recent or on-going, as indicated by the lack of significantly admixed AFLP genotypes or early generation hybrids. Concordant with mating signal analyses, these mismatched specimens in the Medway estuary represent a population of *A. aestuarina* completely fixed for *A. makarovi* mtDNA. Virtually no indication of introgression in any other population was identified, even at sympatric sites (Norfolk and Essex), therefore giving support for the lack of recent hybridisation between these species.

Table 5.4. AFLPOP (Duchesne & Bernatchez 2002) assignment success (%) for 1000 simulated genotypes for six genotypic classes, pure *A. makarovi,* pure *A. aestuarina*, F_1, F_2, first generation backcross (BC) to *A. makarovi* (BC mak) and first generation BC to *A. aestuarina* (BC aest). Simulations were based on pure reference genotypes for *Aphrodes aestuarina* (80 amplified fragment length polymorphism genotypes identified as pure in Bayesian clustering analyses) and *A. makarovi* (90 pure genotypes).

		Allocation (%)					
Population	N	*A. makarovi*	*A. aestuarina*	F_1	F_2	BC mak	BC aest
A. makarovi	1000	99.09				1.84	
A. aestuarina	1000		98.98				2.1
F_1	1000			60.08	32.84	2.38	5.32
F_2	1000			27.59	53.37	6.26	3.85
BC mak	1000	0.91		4.41	7.79	89.52	
BC aest	1000		1.02	7.92	6.00		88.73

5.4.2.2. Population structure of Aphrodes aestuarina *and the mismatched Medway estuary population*

To identify further population structure among mismatched specimens and *A. aestuarina* populations, analyses were done by omitting all *A. makarovi*. Results of the PCA analysis comparing all *A. aestuarina,* including the Medway estuary population are shown in Fig. 5.12. All mismatched Medway specimens cluster together with some overlap with Essex and Sussex *A. aestuarina* populations along the centre of the main cluster, which is separated along PC1 explaining 32.09% of the overall variation within the data set. Virtually no clustering was identified along PC2 and PC3 (explaining 19.07% and 13.84% of the variation), therefore, only PC1 plotted against PC2 is shown (Fig. 5.12).

STRUCTURE Bayesian clustering results comparing *A. aestuarina* and the mismatched Medway population reveal two clusters using both correlated and uncorrelated allele frequencies. Results for correlated allele frequencies when $K = 2$ are shown in Fig. 5.13 (Appendix II (2) for the maximal $\Pr(X|K)$ (Table 5.9), and the ΔK results (Fig. 5.20)). Nineteen non-mismatched *A. aestuarina* specimens cluster with a high assignment probability to cluster two with the remainder of specimens being admixed (Fig. 5.13). All but two mismatched Medway individuals have a high likelihood of admixture between the two clusters. Results for uncorrelated allele frequencies when $K = 3$ (Appendix II (2) for the maximal $\Pr(X|K)$ (Table 5.10), and the ΔK results (Fig. 5.22)) are shown in Fig 5.14, however, only two distinct clusters are observed. The third cluster is likely to represent a 'ghost' cluster with very few individuals showing any likelihood of belonging to this third cluster. Mismatched Medway specimens group as a single cluster at the > 0.9 probability threshold with 15 individuals classed as admixed. Non-mismatched *A. aestuarina* specimens are assigned to a single cluster at the > 0.9 probability threshold with four specimens showing likelihood of being admixed (Fig. 5.14).

STRUCTURE analyses were performed to identify further population structure within the data set. The data set was spilt based on the prior knowledge of the mismatch between mtDNA and nDNA in the Medway estuary population. Using both uncorrelated and correlated allele frequencies no further structure within the mismatched Medway population was identified (Appendix II (4)). Results comparing all non-mismatched *A.*

aestuarina populations for correlated allele frequencies are shown in Fig. 5.15. The most likely number of clusters identified was $K = 3$ (Appendix II (3)). Most Norfolk specimens showed a high probability of belonging to cluster 1 and the majority of Essex and Sussex specimens belong to the second cluster. Three individuals from Sussex and a single individual from Essex and Norfolk show a higher likelihood of belonging to the third cluster (Fig. 5.15). No structure was identified using uncorrelated allele frequencies comparing all *A. aestuarina* populations with all individuals assigned with a high probability to a single cluster (> 0.9).

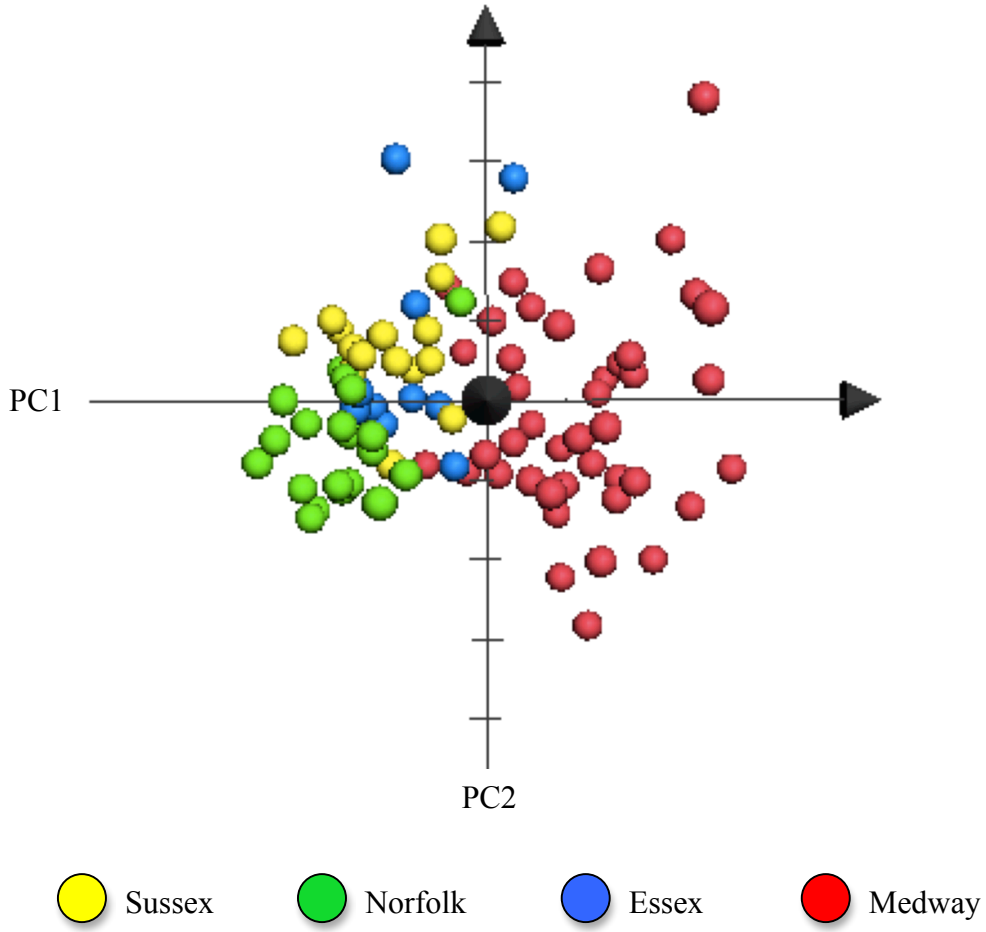


Figure 5.12. Principle coordinates analysis results based on Euclidean distances between amplified fragment length polymorphism (AFLP) multilocus phenotypes for 48 *Aphrodes aestuarina* specimens from Norfolk, Essex and Sussex and 42 mismatched Medway specimens shown to possess *A. makarovi* mitochondrial DNA (5.4.1). 3D plot illustrating the first two principle coordinates (PC1 = 32.09%, PC2 = 19.07% variation explained. Based on 309 polymorphic AFLP loci.

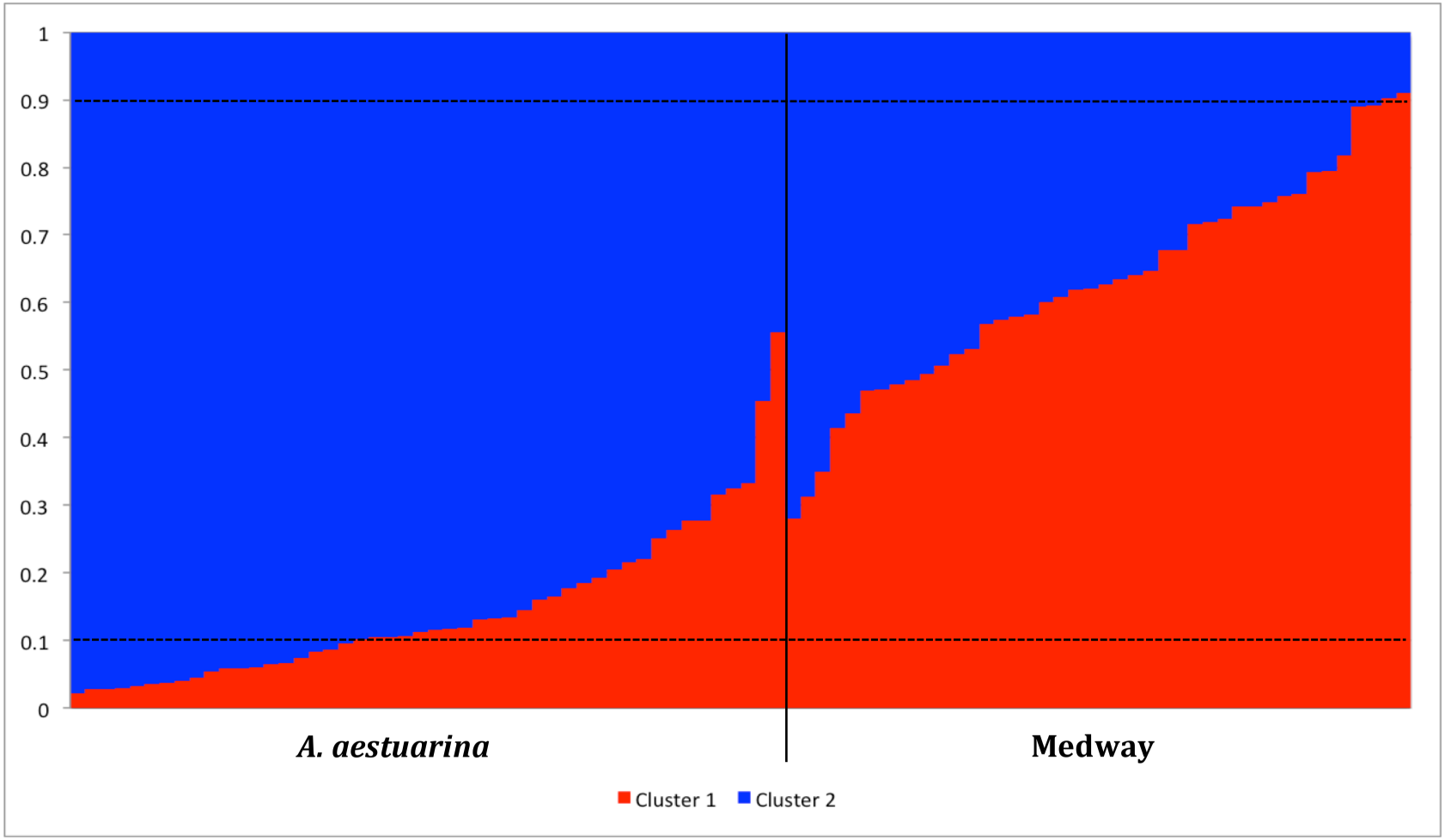


Figure 5.13. Results of STRUCTURE (Pritchard *et al.* 2000) clustering analysis using correlated allele frequencies, showing the assignment of 48 *Aphrodes aestuarina* specimens from Norfolk, Essex and Sussex and 42 mismatched individuals from the Medway estuary (shown to possess *A. makarovi* mitochondrial DNA, 5.4.1) (each thin vertical bar) to either cluster one (red) or cluster two (blue) (*K*=2). Results are based on 309 polymorphic amplified fragment length polymorphism loci. Horizontal dashed lines represent 0.9 probability thresholds for each cluster. Individuals were sorted according to their assignment values.

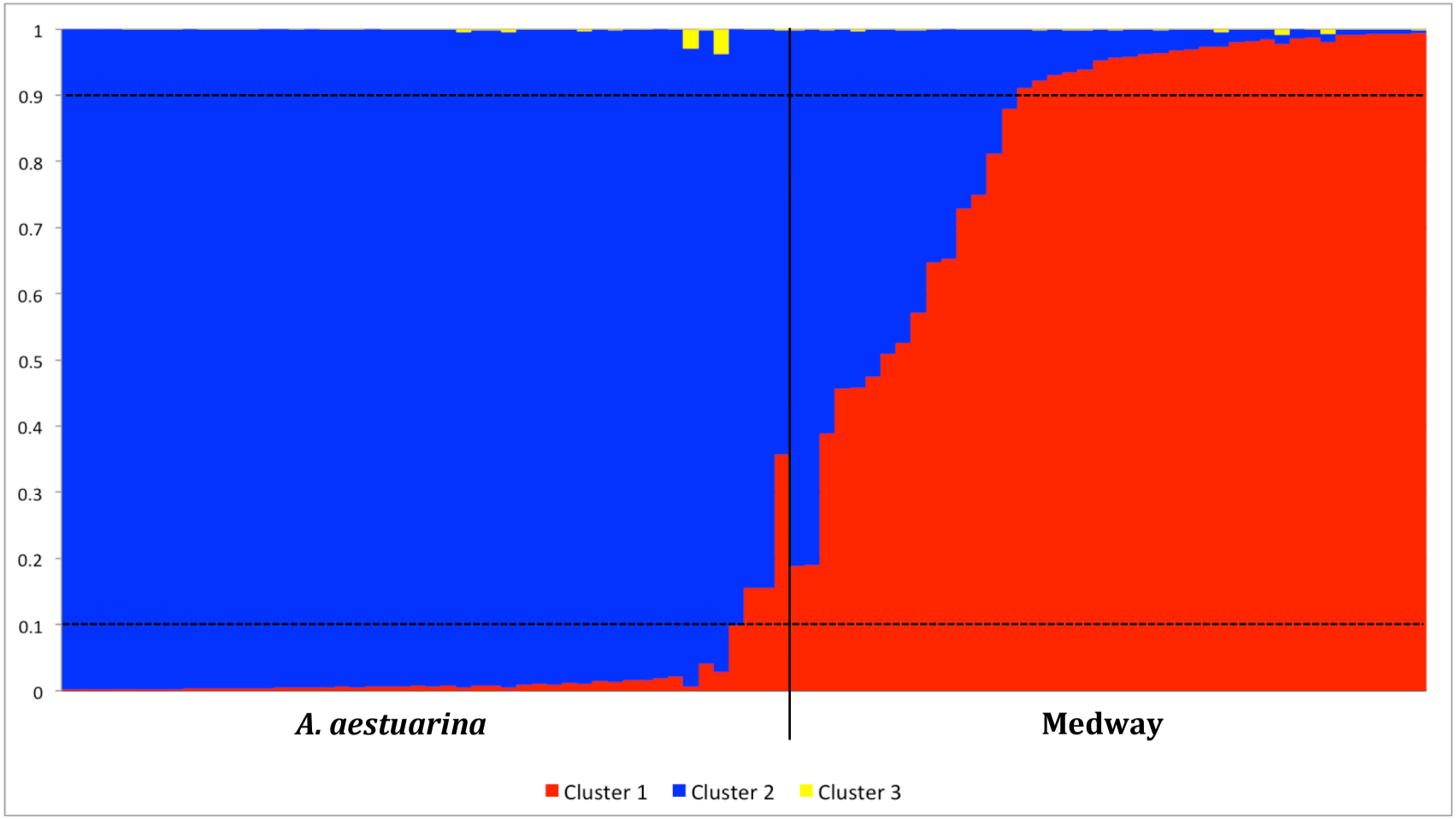


Figure 5.14. Results of STRUCTURE (Pritchard *et al.* 2000) clustering analysis using uncorrelated allele frequencies, showing the assignment of 48 *Aphrodes aestuarina* specimens from Norfolk, Essex and Sussex and 42 mismatched individuals from the Medway estuary (shown to possess *A. makarovi* mitochondrial DNA, 5.4.1) (each thin vertical bar) to either cluster one (red), cluster two (blue) or cluster 3 (yellow) (*K*=3). Results are based on 309 polymorphic amplified fragment length polymorphism loci. Horizontal dashed lines represent 0.9 probability thresholds for each cluster. Individuals were sorted according to their assignment values.

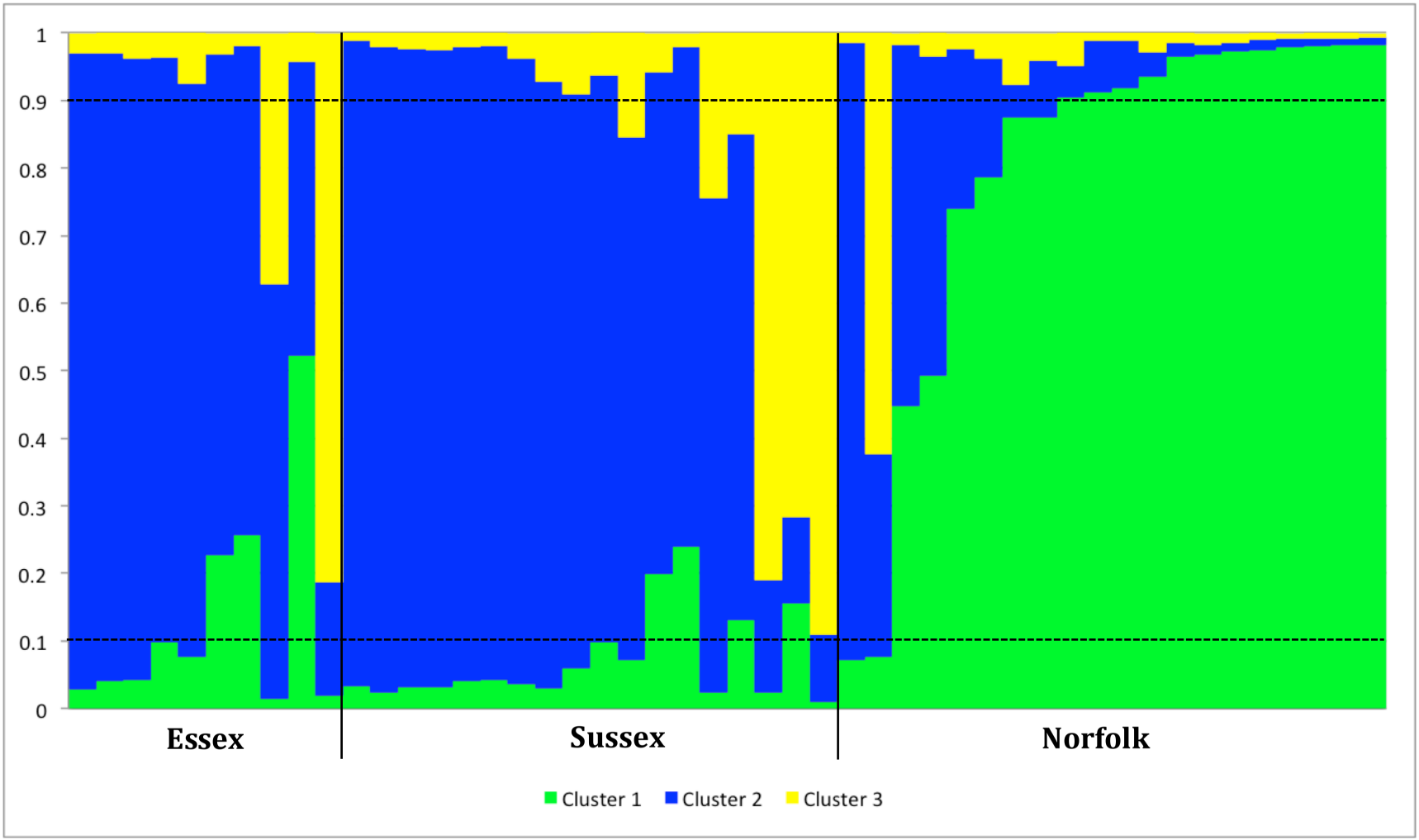


Figure 5.15. Results of STRUCTURE (Pritchard *et al.* 2000) clustering analysis using correlated allele frequencies, showing the assignment of 48 *Aphrodes aestuarina* specimens from Norfolk, Essex and Sussex (each thin vertical bar) to either cluster one (green), cluster two (blue) or cluster 3 (yellow) (*K*=3). Results are based on 208 polymorphic amplified fragment length polymorphism loci. Horizontal dashed lines represent 0.9 probability thresholds for each cluster. Individuals were sorted according to their assignment values.

The numbers of private AFLP bands were identified for each population using GENALEX. In the Sussex *A. aestuarina* population nine private bands were identified, in Norfolk six private bands were found. None were found in the Essex population and in the Medway estuary mismatched population 51 private bands were identified. The numbers of private bands in these populations were significantly reduced when populations of *A. makarovi* were included in the computation, suggesting that these AFLP markers are shared between both species. When typical *A. makarovi* populations were included, no private bands were identified in Norfolk, one was identified in Sussex and only four in the mismatched Medway population. The high number of private AFLP markers in the matched population that are shared with *A. makarovi* (47 bands) but not *A. aestuarina* may be a signature of introgression between these two species.

Mean genetic differentiation among populations, $\theta^{(\Pi)}$, calculated in HICKORY, was 0.07 (significant at the 95% credible interval, 0.0503 – 0.0853), suggesting moderate levels of differentiation among populations. Pairwise population $\theta^{(\Pi)}$ estimates (Table 5.5) show a higher level of genetic differentiation between Norfolk and the mismatched Medway population ($\theta^{(\Pi)} = 0.13$). The lowest differentiation was between Essex and Sussex populations ($\theta^{(\Pi)} = 0.02$). The deviance information criterion (DIC) statistic reported by HICKORY can be used as a model choice criterion (Holsinger *et al.* 2002). Values reported for the three models, full model, $F_{IS} = 0$ model and $\theta^{(\Pi)} = 0$ model were 4295, 5218, 4309 respectively. This clearly demonstrates preference for the full model compared to the $\theta^{(\Pi)} = 0$, supporting the existence of a significant level of differentiation between populations (much lower than $\theta^{(\Pi)} = 0$). A difference of 13.99 between the DIC for the full model and $F_{IS} = 0$ model which was not just due to the difference between the model dimensions (difference between pD values is only 3.9), which indicates some degree of departure from Hardy-Weinberg expectations.

Table 5.5. Pairwise population $\theta^{(\pi)}$ values using the f-free model (HICKORY, Holsinger *et al.* 2002) to estimate genetic differentiation among three populations of *Aphrodes aestuarina* (~A) and the mismatched Medway population (MH) shown to possess *A. makarovi* mitochondrial DNA (5.4.1). See Table 5.1 for full location abbreviations. All values are significantly different from zero based on the 95% credible interval.

	EA	**SA**	**NA**	**MH**
EA	/			
SA	0.0207	/		
NA	0.0996	0.0855	/	
MH	0.0574	0.0510	0.1337	/

Results for the AMOVA analyses based on Euclidean distance are shown in Table 5.6. A large proportion of genetic variation is attributed to differences between individuals within populations (91.13%) and only 8.87% of variation is associated with differences among populations (p-value < 0.0001). When populations are grouped according to non-mismatched *A. aestuarina* populations versus the mismatched Medway population, 4.89% of the variation is accounted for among groups, although this result was non-significant (p-value = 0.26). When the structure based on the groups identified in Bayesian clustering analysis (Norfolk, Essex/Sussex and the Medway mismatched population, Fig. 5.14 and 5.15) was tested, 7.73% of the variation was accounted for among groups, although this result was also non-significant (p-value = 0.16).

Table 5.6. Hierarchical analysis of molecular variance (AMOVA) results based on Euclidian distances between amplified fragment length polymorphism multilocus phenotypes, for 48 non-mismatched *Aphrodes aestuarina* individuals (sampled from Essex (EA), Norfolk (NA) and Sussex (SA)) and 42 mismatched specimens from the Medway (MH). Populations were analysed without structuring (four populations), comparing mismatched against non-mismatched populations (two populations) and based on Bayesian clustering results (three populations).

Source of variation	d.f.	Sum of squares	Variance	% total	Φ statistic	p-value
Populations/Region n=4						
Among populations	3	244.86	2.66	8.87		
Within populations	86	2349.05	27.31	91.13		
Total	89	2593.91	29.97		Φ_{ST} 0.089	< 0.0001
***A. aestuarina* vs MH n=2**						
Among groups	1	141.73	1.49	4.89	Φ_{CT} 0.049	0.26
Among populations within groups	2	103.14	1.57	5.18	Φ_{SC} 0.054	< 0.0001
Within populations	86	2349.05	27.31	89.93	Φ_{ST} 0.101	< 0.0001
Total	89	2593.91	30.37			
EA/SA vs NA vs MH n=3						
Among groups	2	211.34	2.33	7.73	Φ_{CT} 0.077	0.16
Among populations within groups	1	33.53	0.48	1.60	Φ_{SC} 0.017	< 0.05
Within populations	86	2349.05	27.31	90.67	Φ_{ST} 0.093	< 0.0001
Total	89	2593.91	30.13			

The MANTEL test using pairwise F_{ST} estimated from HICKORY and coastal geographic distances revealed no significant correlation between genetic distances and geographic distances (r = 0.2664, *p*-value = 0.3288) between the mismatched Medway estuary and non-mismatched *A. aestuarina* populations, such that no significant pattern of isolation-by-distance was found (Fig. 5.16).

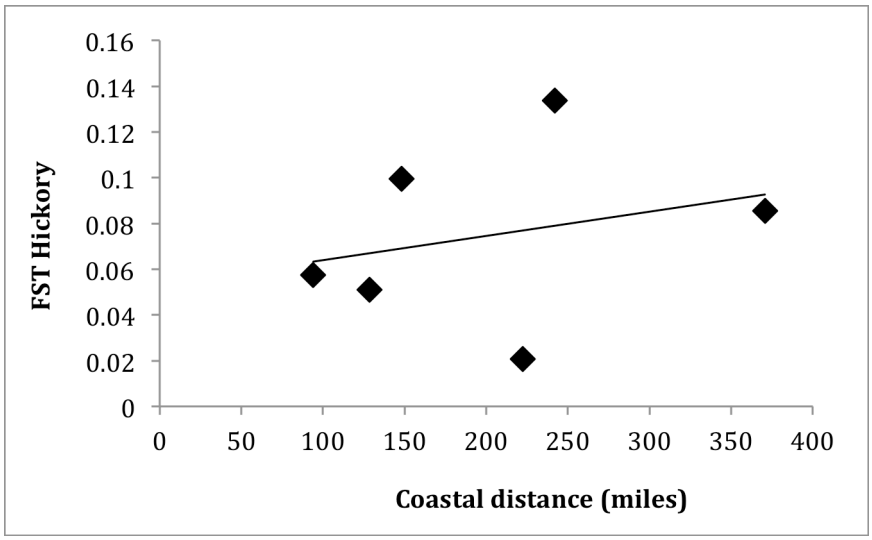


Figure 5.16. Scatter plot illustrating the relationship between pairwise coastal geographic distances (miles) versus the pairwise $\theta^{(n)}$ estimates obtained from HICKORY (Holsinger *et al.* 2002) for three non-mismatched *A. aestuarina* populations (Norfolk, Essex and Sussex) and the mismatched Medway population shown to possess *A. maka*rovi mitochondrial DNA (5.4.1).

5.5. Discussion

The aim of this study was to determine whether the mismatch identified between vibrational mating signal and mtDNA in the Medway estuary population (Chapter 3) was the result of introgressive hybridisation between *A. makarovi* and *A. aestuarina.* To understand the dynamics of hybridisation between saltmarsh-adapted *Aphrodes* three specific aims were addressed, 1) whether analyses of nuclear AFLP marker variation are concordant with mtDNA or mating signal data for the Medway population, 2) whether there is evidence of recent hybridisation or nuclear introgression in addition to mtDNA, and 3) how widespread across the UK this phenomenon is.

The AFLP results show that the mismatched Medway population is closely related to non-mismatched *A. aestuarina*, concordant with mating signal data, and therefore with the current taxonomic designations for *Aphrodes* species (Tishechkin 1998; Virant-Doberlet *et al.* 2005; Chapter 3). Unambiguous distinction between *A. aestuarina* (including the mismatched Medway population) and *A. makarovi,* from both sympatric and allopatric populations was shown in AFLP marker analyses. However, the Medway mismatched population is fixed for *A. makarovi* COI gene haplotypes. The majority of mismatched Medway specimens possessed the most common *A. makarovi* mtDNA haplotype mH1 found at all *A. makarovi* sampling localities, including the sympatric Medway *A. makarovi* population. Three additional haplotypes (two private haplotypes based on the current sampling effort) were recovered from this population, differing by either one or two nucleotides from *A. makarovi* haplotype mH1.

The incongruent patterns recovered for mtDNA and nuclear AFLP markers reflect a reticulate evolutionary pathway, which is likely to have arisen from introgression of *A. makarovi* mtDNA into a single *A. aestuarina* population in the Medway estuary. Retention of ancestral polymorphism or convergent evolution within the mtDNA, are alternative hypotheses that could account for the presence of *A. makarovi* mtDNA in the Medway estuary population. However these hypotheses are unlikely. Retention of ancestral polymorphism, where haplotypes are shared between taxa and are randomly retained in certain populations due to incomplete lineage sorting, has been proposed to explain mtDNA patterns in recently diverged taxa (Wilding *et al.* 2000). Both *A. makarovi* and *A. aestuarina* show levels of intraspecific mtDNA sequence divergence of

less than 1% among UK populations (Chapter 3). In contrast, interspecific mtDNA divergence between *A. makarovi* and *A. aestuarina* has been estimated at approximately 6.9% (K2P, Chapter 3). In phylogenetic analyses (Chapter 3, Fig. 3.9) *A. makarovi* shared a most recent common ancestor with *A. bicincta* (bootstrap support value of 98% in neighbour-joining analyses) rather than with *A. aestuarina,* and all four *Aphrodes* species formed reciprocally monophyletic clades (with high support) indicating that the mtDNA of these species is likely to have undergone complete lineage sorting. Furthermore, it is highly improbable that the locally retained paraphyletic haplotype would match the haplotype found in sympatric *A. makarovi* from the Medway estuary. The *A. makarovi* mtDNA haplotype identified in 95.6% of the Medway estuary mismatched *A. aestuarina* specimens, exactly matches that found in sympatric *A. makarovi* from the Medway estuary, making symplesiomorphic retention or convergent evolution unlikely.

Additionally, due to the lower effective population size of mtDNA relative to nDNA (as mitochondrial DNA is haploid and maternally inherited), mtDNA should resolve species phylogenetic patterns faster than nDNA in the absence of introgression (Palumbi *et al.* 2001; Funk & Omland 2003). Therefore, incomplete lineage sorting is more of a concern for nDNA (Funk & Omland 2003), and thus if mtDNA paraphyly was a product of incomplete lineage sorting then this is likely to be apparent for nDNA patterns also. However, the unambiguous distinction of *A. makarovi* from *A. aestuarina* based on genome-wide nuclear AFLP markers (and species specific vibrational mating signals) irrespective of the geographic region sampled provides further support against this hypothesis.

In relation to the second and third aims and the likelihood of nuclear introgression in addition to mitochondrial introgression across sampled population, no intermediate AFLP genotypes were identified showing that very recent or on-going hybridisation between *A. makarovi* and *A. aestuarina* is highly unlikely. Asymmetrical introgression is evident as all *A. makarovi* AFLP genotypes lacked any signs of admixture, and there is no apparent introgression of *A. aestuarina* AFLP loci (or mtDNA) into any *A. makarovi* population. A low level of nuclear introgression from *A. makarovi* was identified in 21.4% of the mismatched Medway estuary specimens and 8.3% of *A. aestuarina* (collected from Norfolk, Essex and Sussex locations), with the remainder clustering with high probability with *A. aestuarina*. However, in some Bayesian

clustering analyses (when using uncorrelated allele frequency option in STRUCTURE or the Uniform priors for the mixing proportion in NEWHYBRIDS) virtually all specimens were assigned as either pure *A. aestuarina* or *A. makarovi* and so the exact degree of nuclear introgression cannot be confidently determined from the present data. Since both Bayesian clustering methods were largely congruent, the overall level of nuclear introgression is clearly low across all sampling localities. A larger proportion of individuals would be expected to show significant admixture from both parental gene pools if recent introgression was prevalent. Additionally, other studies have identified higher proportions of polymorphic loci in introgressed populations (Albert *et al.* 2006), which would be expected for admixed vs. pure gene pools. However, this was not seen in either the Medway estuary mismatched population or any other *A. aestuarina* population with possible introgressed individuals, providing additional support for low levels of nuclear introgression. Further analyses with codominant nuclear markers, such as microsatellites, would provide insights into the degree of departure from HWE in possible introgressed hybrid individuals (Sušnik *et al.* 2007).

The presence of *A. makarovi* mtDNA in the mismatched Medway *A. aestuarina* population is therefore likely to be a case of 'mtDNA capture', based on the concordance of nuclear AFLP markers and mating signal data in grouping the mismatched Medway population with *A. aestuarina* populations, and lack of significant nuclear introgression. The reticulate evolutionary pathway generated by interspecific mtDNA exchanges has been documented for a number of animal groups, including fish (Dowling & DeMarais 1993; Bernatchez *et al.* 1995; Wilson & Bernatchez 1998; Sušnik *et al.* 2007), mice (Ferris *et al.* 1983), sparrows (Weckstein *et al.* 2001), *Nasonia* wasps (Raychoudhury *et al.* 2009), *Laupala* crickets (Shaw 2002) and *Lycaeides* butterflies (Gompert *et al.* 2006, 2008).

Uni-directional hybridisation (male *A. aestuarina* x female *A. makarovi*) followed by repeated backcrossing with only *A. aestuarina* could have produced the genetic architecture seen in the Medway estuary *A. aestuarina* population. Weckstein *et al.* (2001) provide a diagram explaining the principle of mtDNA capture. The asymmetrical hybridisation pattern suggests hybrids are fertile, but possibly capable of only breeding with one parent species. Further laboratory crosses between these species will determine whether hybrids are capable of backcrossing with either parent.

It is unlikely that incomplete speciation and secondary contact could explain the occurrence of hybridisation between *A. makarovi* and *A. aestuarina*, which has been suggested for certain species that survived climatic oscillations in Southern European glacial refugia (Taberlet *et al.* 1998; Hewitt 2001). Populations inhabiting such refugia are thought to have diverged through drift, and during postglacial periods Northern migration of populations resulted in formation of hybrid zones due to secondary contact between partially reproductively isolated populations from different refugia (Taberlet *et al.* 1998). Based on the estimate of c. 2% divergence per million years of invertebrate mtDNA (Hewitt 2001), speciation of *A. makarovi* and *A. makarovi* would have initiated c. 3.5 million years ago, giving ample time for complete speciation to occur.

Although the hybridisation event cannot be dated based on current data, support for a lack of very recent or on-going introgression is evident, due to a number of AFLP loci that are private to the global data set (four AFLP bands and two presently private mtDNA haplotypes) that were identified in the mismatched Medway population. Hence this population has been in isolation long enough for mutation and drift to generate a number of unique alleles, possibly for the last few hundred years or more. Conversely, 8.5% of AFLP bands in the Medway mismatched population were shared with *A. makarovi* but not *A. aestuarina* (although the individual frequencies of these AFLP bands in this population is unknown), so that later generation backcross hybrids (3^{rd} or 4^{th} generation or possibly later) are possible. This means that hybridisation could have taken place in the last decade or two, as a minimum possible estimate. High pairwise F_{ST} values between mismatched Medway *A. aestuarina* and non-mismatched *A. aestuarina* from other locations were recovered, could explain the number of shared *A. makarovi* AFLP bands in this mismatched population (although this is yet to be tested). The lack of a statistically significant correlation between pairwise F_{ST} estimates and geographic distance clearly indicates no pattern of isolation-by-distance. In addition, the AMOVA performed on localities assembled into groups according to both the knowledge of a mismatch between mtDNA and nDNA (or lack of) and results from Bayesian clustering analyses performed with STRUCTURE (5.4.2.2) did not show statistically significant apportionment of the genetic variance among regional groups (i.e., no apparent extrinsic barriers). Much of the genetic structuring observed could be explained through differentiation within localities.

A theoretical study by Chan & Levin (2005) revealed mitochondrial introgression from a rare species into that of a more common one occurs more readily than nuclear introgression and such effects were strongest when the number of immigrants (rare species) was low. They found that normally choosy females are more likely to accept heterospecific males when conspecific males are rare. Frequency-dependant prezygotic mating barriers therefore can explain why mtDNA introgression appears more common than nuclear introgression (Chan & Levin 2005).

There is currently no knowledge of disparate mating periods between *A. makarovi* and *A. aestuarina*. However, based on field observations, females show a prolonged period of reproductive activity compared with males of each species (based on the personal observation of females in the field at the end of the mating season when very few males are present). Hybridisation may occurr through a rare interspecific mating event between a female *A. makarovi* and male *A. aestuarina* towards the end of a mating season, where unmated females show reduced choosiness toward non-conspecific *A. aestuarina* males. Additionally, in the Medway estuary and surrounding Thames estuary, portions of the saltmarsh habitats have been destroyed through human disturbance (in the form of urban and industrialised areas, sea defences and increased pollution) and therefore, recent bottlenecks have probably occurred in this region. Such recent, localised events may not be related to the historical demography of the mismatched Medway population, or the fixation of *A. makarovi* mtDNA.

There are two likely explanations for the lack of significant nuclear introgression and the geographically local mtDNA introgression found in *A. aestuarina* in the Medway estuary population. First, selection sufficiently strong enough against *A. makarovi* x *A. aestuarina* hybrids and backcrosses to restrict nuclear introgression but which would not limit the interspecific exchange of neutral mtDNA alleles from *A. makarovi* to *A. aestuarina*. Genetic drift could result in fixation of neutral mtDNA variants in the genetic background of the other species. Under this scenario, strong reinforcing selection would be required to conserve species boundaries, likely aided by asymmetrical barriers to gene flow (Shaw 2002) and unidirectional introgression (Chan & Levin 2005). The lack of intermediate vibrational mating signal variants (Chapter 3) and nuclear introgression at numerous sympatric sites thus means that frequent or current hybridisation is unlikely. The lack of recent hybridisation is facilitated by assortative mating and mate choice (through vibrational communication and other

possible unexplored mechanisms). Such prezygotic reproductive isolating mechanisms permit introgression of maternally inherited genes to a greater extent than postzygotic barriers to gene flow (Chan & Levin 2005).

Second, a selective sweep of *A. makarovi* mtDNA could possibly fix *A. makarovi* mtDNA in the Medway *A. aestuarina* population. If the possession of the mitochondria of *A. makarovi* or associated nuclear genes conveyed a selective advantage in the genetic background of *A. aestuarina*, this haplotype could have spread rapidly (Ballard & Whitlock 2004). Selection on mtDNA may be dependant on environmental properties (e.g. fitness effects associated with metabolic rates at different temperatures, Ballard & Whitlock 2004). As mtDNA shows little or no recombination, selection at a single nucleotide causes a selective sweep of the whole molecule. Additionally, selection at nuclear loci associated with mtDNA affects rates of mtDNA evolution (Ballard & Whitlock 2004). Transmission of maternally inherited symbionts, such as *Wolbachia,* that cause cytoplasmic incompatibility in a population affect patterns of mtDNA diversity (Ballard & Whitlock 2004), as the mtDNA haplotype associated with the initial infection of a population sweeps to high frequency as the symbiont spreads (Turelli & Hoffmann 1995; Jiggins 2003; Narita *et al.* 2006; Gompert *et al.* 2008; Raychoudhury *et al.* 2009).

Patterns exhibited by selection causing a selective sweep to explain the fixation of mtDNA in this population are similar to those exhibited by population bottlenecks and it is therefore difficult to distinguish among these alternatives. For example, tests for departures from neutral evolutionary expectations (Tajima's D and Fu's F_S) are likely to yield both significant and negative test statistic values for processes such as population bottlenecks and selective sweeps (Tajima 1989; Ballard & Whitlock 2004), as recovered for *A. makarovi* mtDNA lineage (negative but non-significant values were recovered for *A. aestuarina*) in this study. Further evidence is required to determine whether selection or drift resulted in the fixation of *A. makarovi* mtDNA in the mismatched Medway estuary population. Additional crossing experiments between *A. makarovi* and *A. aestuarina* will provide further insights into the reproductive isolating mechanisms and hybrid fitness, and the possible selective or replicative advantage of *A. makarovi* mtDNA should be explored.

Although identification of interspecific mtDNA introgression from *A. makarovi* into *A. aestuarina* has only been identified in a single estuary, it is possible that this phenomenon is more widespread among these species. At present there is no conclusive evidence for the occurrence of estuarine *A. makarovi* or *A. aestuarina* anywhere else except in the UK. Specimens thought to be *A. aestuarina* have been identified in other coastal regions of the North and Baltic Sea, including Germany and Poland (Kirby 1992; Biedermann & Niedringhaus 2004). However, this was before knowledge that both species are present in the same estuarine habitat, often on the same host plant (Chapter 3). Further geographic sampling of European saltmarshes should elucidate the range of estuarine *Aphrodes* distributions and provide further insights into the extent of mitochondrial introgression and hybridisation among these species.

5.6. Conclusion

The results demonstrate the potential for introgressive hybridisation to have substantial and possible long-term effects on the genetic configuration of species and can produce considerable discrepancies among speciation histories based on nuclear and mitochondrial markers. Phylogenetic analyses made with the use of a single gene, and without consideration of reticulate evolution should be taken cautiously. Adequate taxon sampling is also a necessity (Funk & Omland 2003) as also highlighted in this study, and samples should be taken from a number of populations to confirm results. Inference of phylogeographic species histories should ideally be made using multiple markers that are of uni- and bi-parental inheritance. The AFLP techniques has proven to be a valuable tool allowing for substantial numbers of genome wide markers to be amplified per-individual, providing insights into the degree of nuclear DNA divergence that can often be misleading or conflicting when based solely on single gene sequences.

5.7. References

Ajmone-Marsan, P., Valentini, A., Cassandro, M., Vecchiotti-Antaldi, G., Bertoni, G., & Kuiper, M. (1997). AFLP^{TM} markers for DNA fingerprinting in cattle. *Animal Genetics, 28,* 418-426.

Albert, V., Jónsson, B., & Bernatchez, L. (2006). Natural hybrids in Atlantic eels (*Anguilla anguilla, A. rostrata*): evidence for successful reproduction and fluctuating abundance in space and time. *Molecular Ecology*, *15*, 1903-1916.

Anderson, E. C., & Thompson, E. A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, *160*, 1217-1229.

Aris-brosou, S., & Excoffier, L. (1996). The impact of population expansion and mutation rate heterogeneity on DNA sequence polymorphism. *Molecular Biology and Evolution*, *13*, 494-504.

Arnold, M. L. (1992). Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics*, *23*, 237-261.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*, 729-744.

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

Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*, 113-148.

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

Bensch, S., & Åkesson, M. (2005). Ten years of AFLP in ecology and evolution: why so few animals? *Molecular Ecology*, *14,* 2899-2914.

Bernatchez, L., Glémet, H., Wilson, C. C., & Danzmann, R. G. (1995). Introgression and fixation of Arctic char (*Salvelinus alpinus*) mitochondrial genome in an allopatric population of brook trout (*Salvelinus fontinalis*). *Canadian Journal of Fisheries and Aquatic Sciences*, *52*, 179-185.

Biedermann, R., & Niedringhaus, R. (2004). *Die Zikaden Deutschlands*. WABV, Scheeßel, (pp. 214-215).

Bluemel, J. K., King, R. A., Virant-Doberlet, M., & Symondson, W. O. C. (2011). Primers for identification of type and other archived specimens of *Aphrodes* leafhoppers (Hemiptera, Cicadellidae). *Molecular Ecology Resources, 11*, 770-774.

Bonin, A., Bellemain, E., Eidesen, P. B., Pompanon, F., Brochmann, C., & Taberlet, P. (2004). How to track and assess genotyping errors in population genetics studies. *Molecular Ecology*, *13,* 3261-3273.

Bonin, A., Ehrich, D., & Manel, S. (2007). Statistical analysis of amplified fragment length polymorphism data: a toolbox for molecular ecologists and evolutionists. *Molecular Ecology*, *16*, 3737-3758.

Buerkle, C. A., Morris, R. J., Asmussen, M. A., & Rieseberg, L. H. (2000). The likelihood of homoploid hybrid speciation. *Heredity*, *84*, 441-451.

Bull, V., Beltrán, M., Jiggins, C. D., McMillan, W. O., Bermingham, E., & Mallet, J. (2006). Polyphyly and gene flow between non-sibling Heliconius species. *BMC Biology*, *4*, doi:10.1186/1741-7007-4-11.

Chan, K. M. A, & Levin, S. A. (2005). Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution*, *59*, 720-729.

Claridge, M. F. (1985). Acoustic signals in the Homoptera: behaviour, taxonomy, and evolution. *Annual Review of Entomology, 30*, 297-317.

Čokl, A., & Virant-Doberlet, M. (2003). Communication with substrate-borne signals in small plant-dwelling insects. *Annual Review of Entomology, 48*, 29-50.

Colbeck, G. J., Turgeon, J., Sirois, P., & Dodson, J. J. (2011). Historical introgression and the role of selective vs. neutral processes in structuring nuclear genetic

variation (AFLP) in a circumpolar marine fish, the capelin (*Mallotus villosus*). *Molecular Ecology*, *20*, 1976-1987.

Coyne, J. A., & Orr, H. A. (2004). *Speciation.* Sinauer Associates, Sunderland, USA.

den Hartog, P. M., den Boer-visser, A. M., & ten Cate, C. (2010). Unidirectional hybridization and introgression in an avian contact zone: evidence from genetic markers, morphology, and comparisons with laboratory-raised F_1 hybrids. *The Auk*, *127*, 605-616.

Dowling, T. E., & DeMarais, B. D. (1993). Evolutionary significance of introgressive hybridisation in cyprinid fishes. *Nature*, *362*, 444-446.

Duchesne, P., & Bernatchez, L. (2002). AFLPOP: a computer program for simulated and real population allocation, based on AFLP data. *Molecular Ecology Notes*, *2*, 380-383.

Egger, B., Koblmüller, S., Sturmbauer, C., & Sefc, K. M. (2007). Nuclear and mitochondrial data reveal different evolutionary processes in the Lake Tanganyika cichlid genus *Tropheus*. *BMC Evolutionary Biology*, *7*, doi:10.1186/1471-2148-7-137.

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

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

Falush, D., Stephens, M., & Pritchard, J. K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, *164*, 1567-1587.

Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular Ecology Notes*, *7*, 574-578.

Ferris, S. D., Sage, R. D., Huang, C. M., Nielsen, J. T., Ritte, U., & Wilson, A. C. (1983). Flow of mitochondrial DNA across a species boundary. *Proceedings of the National Academy of Sciences of the United States of America*, *80*, 2290-2294.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & 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.

Fu, Y-X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, *147*, 915-925.

Funk, D. J., & Omland, K. E. (2003). Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, *34*, 397-423.

Gomes, B., Sousa, C. A., Novo, M. T., Freitas, F. B., Alves, R., Côrte-Real, A. R., Salgueiro, P., Donnelly, M. J., Almeida, A. P. G., & Pinto, J. (2009). Asymmetric introgression between sympatric molestus and pipiens forms of *Culex pipiens* (Diptera: Culicidae) in the Comporta region, Portugal. *BMC Evolutionary Biology*, *9*, doi:10.1186/1471-2148-9-262.

Gompert, Z., Nice, C. C., Fordyce, J. A., Forister, M. L., & Shapiro, A. M. (2006). Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly. *Molecular Ecology*, *15*, 1759-1768.

Gompert, Z., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2008). Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. *Molecular Ecology*, *17*, 5231-5244.

Grant, P. R., Grant, B. R., Markert, J. A, Keller, L. F., & Petren, K. (2004). Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution*, *58*, 1588-1599.

Grant, P. R., & Grant, B. R. (2010). Conspecific versus heterospecific gene exchange between populations of Darwin's finches. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 1065-1076.

Harpending, R. C. (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology, 66*, 591-600.

Harrison, R. G. (1993). Hybrids and hybrid zones: historical perspective. In: *Hybrid Zones and the Evolutionary Process* (ed. Harrison RG), Oxford University Press, Oxford. (pp. 3–12).

Hewitt, G. M. (2001). Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Molecular Ecology*, *10*, 537-549.

Holsinger, K. E., Lewis, P. O., & Dey, D. K. (2002). A Bayesian approach to inferring population structure from dominant markers. *Molecular Ecology*, *11*, 1157-1164.

Holsinger, K. E., & Wallace, L. E. (2004). Bayesian approaches for the analysis of population genetic structure: an example from *Platanthera leucophae* (Orchidaceae). *Molecular Evolution, 13*, 887-894.

Jiggins, F. M. (2003). Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics*, *164*, 5-12.

Kimura, M. (1980). A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *15*, 111–120.

Kirby, P. (1992). *A Review of the Scarce and Threatened Hemiptera of Great Britain.* No. 2. The Joint Nature Conservation Committee. (pp.154).

Koblmüller, S., Duftner, N., Sefc, K. M., Aibara, M., Stipacek, M., Blanc, M., Egger, B., & Sturmbauer, C. (2007). Reticulate phylogeny of gastropod-shell-breeding cichlids from Lake Tanganyika – the result of repeated introgressive hybridization. *BMC Evolutionary Biology*, *7*, doi:10.1186/1471-2148-7-7.

Kronforst, M. R., Young, L. G., Blume, L. M., & Gilbert, L. E. (2006). Multilocus analyses of admixture and introgression among hybridizing *Heliconius* butterflies. *Evolution*, *60*, 1254-1268.

Lefebvre, T., Châline, N., Limousin, D., Dupont, S., & Bagnères, A-G. (2008). From speciation to introgressive hybridization: the phylogeographic structure of an island subspecies of termite, *Reticulitermes lucifugus corsicus*. *BMC Evolutionary Biology*, *8*, doi:10.1186/1471-2148-8-38.

Linnen, C. R., & Farrell, B. D. (2007). Mitonuclear discordance is caused by rampant mitochondrial introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. *Evolution*, *61*, 1417-1438.

Litzkow, M. J., Livny, M., & Mutka. M. W. (1998). Condor – A hunter of idle workstations. *Proceedings of 8th IEEE International Conference on Distributed Computing Systems (ICDCS8), San Jose, California, USA,* (pp. 104–111).

Llopart, A., Lachaise, D., & Coyne, J. A. (2005). Multilocus analysis of introgression between two sympatric sister species of Drosophila: *Drosophila yakuba* and *D. santomea*. *Genetics*, *171*, 197-210.

Lynch, M., & Milligan, B. G. (1994). Analysis of population genetic structure with RADP markers. *Molecular Ecology*, *3,* 91-99.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*, 229-237.

Mallet, J. (2007). Hybrid speciation. *Nature*, *446*, 279-283.

Martinsen, G. D., Whitham, T. G., Turek, R. J., & Keim, P. (2001). Hybrid populations selectively filter gene introgression between species. *Evolution*, *55*, 1325-1335.

McKinnon, G. E., Smith, J. J., & Potts, B. M. (2010). Recurrent nuclear DNA introgression accompanies chloroplast DNA exchange between two eucalypt species. *Molecular Ecology*, *19*, 1367-1380.

Meudt, H. M., & Clarke, A. C. (2007). Almost forgotten or latest practice? AFLP applications, analyses and advances. *Trends in Plant Science*, *12*, 106-117.

Moritz, C., Dowling, T. E., & Brown, W. M. (1987). Evolution of animal mitochondrial DNA: relevance for population biology and systematics. *Annual Review of Ecology and Systematics*, *18*, 269-292.

Narita, S., Nomura, M., Kato, Y., & Fukatsu, T. (2006). Genetic structure of sibling butterfly species affected by *Wolbachia* infection sweep: evolutionary and biogeographical implications. *Molecular Ecology*, *15*, 1095- 1108.

Nickel, H. (2003). *The leafhoppers and planthoppers of Germany (Hemiptera, Auchenorrhyncha): patterns and strategies in a highly diverse group of phytophagous insects.* Pensoft Publishers, Sofia-Moscow. (pp. 103-105).

Peakall, R., & Smouse P. E. (2006). GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6, 288-295.

Powell, J. R. (1983). Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, *80*, 492-495.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.

Raychoudhury, R., Baldo, L., Oliveira, D. C. S. G., & Werren, J. H. (2009). Modes of acquisition of *Wolbachia*: horizontal transfer, hybrid introgression, and codivergence in the *Nasonia* species complex. *Evolution*, *63*, 165-183.

Renoult, J. P., Geniez, P., Bacquet, P., Benoit, L., & Crochet, P-A. (2009). Morphology and nuclear markers reveal extensive mitochondrial introgressions in the Iberian Wall Lizard species complex. *Molecular Ecology*, *18*, 4298-4315.

Rogers, A. R., & Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution, 9*, 552-569.

Ryan, F. P. (2006). Genomic creativity and natural selection: a modern synthesis. *Biological Journal of the Linnean Society*, *88*, 655-672.

Schneider, S., & Excoffier, L., (1999). Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics 152*, 1079-1089.

Schwenk, K., Brede, N., & Streit, B. (2008). Introduction. Extent, processes and evolutionary impact of interspecific hybridization in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*, 2805-2811.

Seehausen, O. (2004). Hybridization and adaptive radiation. *Trends in Ecology & Evolution*, *19*, 199-207.

Shaw, K. L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 16122-16127.

Sušnik, S., Weiss, S., Odak, T., Delling, B., Treer, T., & Snoj, A. (2007). Reticulate evolution: ancient introgression of the Adriatic brown trout mtDNA in softmouth trout *Salmo obtusirostris* (Teleostei: Salmonidae). *Biological Journal of the Linnean Society*, *90*, 139-152.

Taberlet, P., Fumagalli, L., Wust-Saucy, A. G., & Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology*, *7*, 453-464.

Tajima, F. **(**1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123*, 585-595.

Tishechkin, D. Y. (1998). Acoustic signals and morphological characters of leafhoppers belonging to *Aphrodes bicinctus* group (Homoptera, Cicadellidae) from European Russia. *Entomological Review, 78*, 370-377.

Turelli, M., & Hoffmann, A. A. (1995). Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. *Genetics*, *140*, 1319-1338.

Vähä, J-P., & Primmer, C. R. (2006). Efficiency of model-based Bayesian methods for detecting hybrid individuals under different hybridization scenarios and with different numbers of loci. *Molecular Ecology, 15*, 63-72.

Vallender, R., Robertson, R. J., Friesen, V. L., & Lovette, I. J. (2007). Complex hybridization dynamics between golden-winged and blue-winged warblers

(*Vermivora chrysoptera* and *Vermivora pinus*) revealed by AFLP, microsatellite, intron and mtDNA markers. *Molecular Ecology*, *16*, 2017- 2029.

Vekemans, X., Beauwens, T., Lemaire, M., & Roldán-Ruiz, I. (2002). Data from amplified fragment length polymorphism (AFLP) markers show indication of size homoplasy and of a relationship between degree of homoplasy and fragment size. *Molecular Ecology*, *11,* 139-151.

Virant-Doberlet, M., & Čokl, A. (2004). Vibrational communication in insects. *Neotropical Entomology, 33*, 121–134.

Virant-Doberlet, M., Prešern, J., Stewart, A. J. A., & Wilson, M. R. (2005). The genus *Aphrodes*: bioacoustic approach to solve taxonomic problems. *Congress of International BioAcoustic Council (IBAC).* (pp. 35).

Virant-Doberlet, M., Prešern, J., Bluemel, J., King, A. R., Wilson, M. R., Symondson, W. O. C., & Stewart, A. J. A. (2006). Vibrational signals and reproductive isolation in the genus *Aphrodes* (Hemiptera: Cicadellidae). *11th International Meeting on Invertebrate Sound and Vibration*, University of Toronto at Scarborough.

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Fritjers, A., Pot, J., Peleman, J., Kuiper, M., & Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research*, *23*, 4407-4414.

Weckstein, J. D., Zink, R. M., Blackwell-Rago, R. C., & Nelson, D. A. (2001). Anomalous variation in mitochondrial genomes of white-crowned (*Zonotrichia leucophrys*) golden-crowned (*Z. atricapilla*) sparrows: pseudogenes, hybridisation, or incomplete lineage sorting? *The Auk*, *118*, 231-236.

Weir, B. S., & Cockerham, C. C. (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution*, *38*, 1358-1370.

Whitlock, R., Hipperson, H., Mannarelli, M., Butlin, R. K., & Burke, T. (2008). An objective, rapid and reproducible method for scoring AFLP peak-height data that minimizes genotyping error. *Molecular Ecology Resources*, *8,* 725-735.

Wilding, C. S., Grahame, J., & Mill, P. J. (2000). Mitochondrial DNA COI haplotype variation in sibling species of rough periwinkles. *Heredity*, *85*, 62-74.

Wilson, C. C., & Bernatchez, L. (1998). The ghost of hybrids past: fixation of arctic charr (*Salvelinus alpinus*) mitochondrial DNA in an introgressed population of lake trout (*S. namaycush*). *Molecular Ecology*, *7*, 127-132.

Zhivotovsky, L. A. (1999). Estimating population structure in diploids with multilocus dominant DNA markers. *Molecular Ecology*, *8,* 907-913.

5.8. Appendix

Appendix I – Sampling locations and geographic coordinates

Location	Host plant/ habitat type	Country	Geographic co-ordinate
Kent (KM)			
Pegwell Bay	*Atriplex* - saltmarsh	England	N 51° 18.954' E 001° 21.590'
Essex (EM/EA)			
Canvey Bridge	*Atriplex* - saltmarsh	England	N 51° 32.551' E 000° 33.834'
Canvey Island	*Atriplex* - saltmarsh	England	N 51° 31.343' E 000° 37.025'
Mersea Island	*Atriplex* - saltmarsh	England	N 51° 47.728’ E 000° 55.322’
Horsey Island	*Atriplex* - saltmarsh	England	N 51° 51.552’ E 001° 14.649’
Gower (GM)			
Penclaudd	*Atriplex* - saltmarsh	Wales	N 51° 38.646’ W 004° 06.659’
Norfolk (NM/NA)			
Stiffkey Marsh (NM/NA)	*Atriplex* - saltmarsh	England	N 52° 57.621' E 000° 55.389'
Warham Marsh (NM)	*Atriplex* - saltmarsh	England	N 52° 57.367' E 000° 54.505'
Wells East Quay (NA)	*Atriplex* - saltmarsh	England	N 52° 57.417' E 000° 51.851'
Medway (MM/MH)			
R (site destroyed) (MM/MH)	*Atriplex* - saltmarsh	England	N 51° 23.460' E 000° 30.560
R1 (MH)	*Atriplex* - saltmarsh	England	N 51° 23.597' E 000° 29.637'
R2 (MH)	*Atriplex* - saltmarsh	England	N 51° 23.500' E 000° 29.670'
R3-Wouldham Marsh (MH)	*Atriplex* - saltmarsh	England	N 51° 22.367' E 000° 27.994'
R4-Rochester Castle (MH)	*Atriplex* - saltmarsh	England	N 51° 23.280' E 000° 29.952'
R5-Baty’s Marsh (MM)	*Atriplex* - saltmarsh	England	N 51° 22.588' E 000° 28.986'
R6-Riverside Walk (MM)	*Atriplex* - saltmarsh	England	N 51° 23.348’ E 000° 30.617’
R7-Gillingham Pier (MM)	*Atriplex* - saltmarsh	England	N 51° 23.849’ E 000° 33.327’
R8-Chatham Reach (MH)	*Atriplex* - saltmarsh	England	N 51° 24.015’ E 000° 30.945’
R9-Hoo St Werbergs (MM/MH)	*Atriplex* - saltmarsh	England	N 51° 24.677’ E 000° 33.727’
R10-Stoke/Grain (MM/MH)	*Atriplex* - saltmarsh	England	N 51° 26.959’ E 000° 39.356’
R11-Lower Twydall (MM/MH)	*Atriplex* - saltmarsh	England	N 51° 23.198' E 000° 35.610'
R12-Funton Creek (MM)	*Atriplex* - saltmarsh	England	N 51° 22.807' E 000° 41.825'
Sussex (SA)			
Shoreham	*Atriplex* - saltmarsh	England	N 50° 50.456' W 000° 17.387'
Rye Harbour	*Atriplex* - saltmarsh	England	N 50° 56.223' E 000° 45.822'

Gower (GIM)			
Penclaudd	*Urtica* - Grassland	Wales	N 51° 38.609’ W 004° 05.742’
Church Lane	*Urtica* - Grassland	Wales	N 51° 38.020’ W 004° 05.992’
Graveyard	*Urtica* - Grassland	Wales	N 51° 38.262’ W 004° 05.925’
Near Ilston	*Urtica* - Grassland	Wales	N 51° 35.482’ W 004° 05.429’
Norfolk (NIM)			
Warham Marsh coastal footpath	*Urtica* - Grassland	England	N 52° 57.367' E 000° 54.505'
Stiffkey Marsh coastal footpath	*Urtica* - Grassland	England	N 52° 57.409’ E 000° 55.384’
Stiffkey Marsh Road	*Urtica* - Grassland	England	N 52° 57.272’ E 000° 55.431’
Stiffkey-Wells Road	*Urtica* - Grassland	England	N 52° 56.946’ E 000° 54.404’
Lisvane (LIM)			
Lisvane	*Urtica* - Grassland	Wales	N 51° 32.160’ W 003° 10.173’
Castle Hill (CIM)			
Castle Hill	*Urtica* - Grassland	England	N 50° 50.473’ W 000° 04.400’

A. makarovi locations (estuarine ~M or inland ~IM)

A. aestuarina locations (~A)

Introgressed *A. aestuarina* specimens possessing *A. makarovi* mitochondrial DNA (MH)

Appendix II – Determining the number of genetic clusters (*K*) in analyses using STRUCTURE version 2.2 (Pritchard *et al.* 2000).

1. *A. makarovi* versus *A. aestuarina* and the Medway estuary population.

The Δ*K* method of Evanno *et al.* (2005) that calculates the rate of change in the log probability of data between successive *K*-values identified two genetic clusters (*K* = 2; Fig. 5.17) when comparing *A. makarovi*, *A. aestuarina* and the mismatched Medway population, using correlated allele frequencies. For *K* = 2 assignment results see Fig. 5.8, Section 5.4.2.1. The maximum log probability of the data, Pr(*X*|*K*), reported by STRUCTURE is shown at *K* = 5 (see value highlighted in Table 5.7). For *K* = 5 (Fig. 5.18, below), the results are similar to that shown in Fig. 5.8, with clustering of *A. aestuarina* and the Medway mismatched population (cluster 1), which are distinct from *A. makarovi*. Admixture across species is minimal as seen when *K* = 2. The main difference is the separation of estuarine and inland *A. makarovi* into distinct clusters with some admixture (clusters 3 and 5, respectively). Habitat related genetic structuring in this species is reported in Chapter 4. Individuals across populations and species were assigned with low probability to clusters 2 and 4.

Table 5.7. Log probability values given by STRUCTURE using correlated allele frequencies, averaged over 10 replicate runs for each value of *K* tested (*K* = 1-10).

K	Log probability	*K*	Log probability
1	-45434.15	6	-41780.63
2	-41398.97	7	-41317.36
3	-43766.47	8	-46329.77
4	-42542.23	8	-47557.97
5	**-40884.61**	10	-46637.14

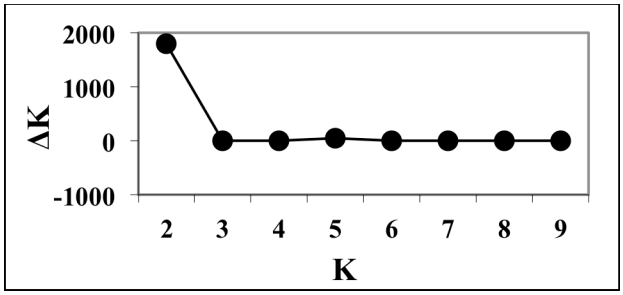


Fig. 5.17. Δ*K* values for *K* genetic clusters (*K* = 2-9).

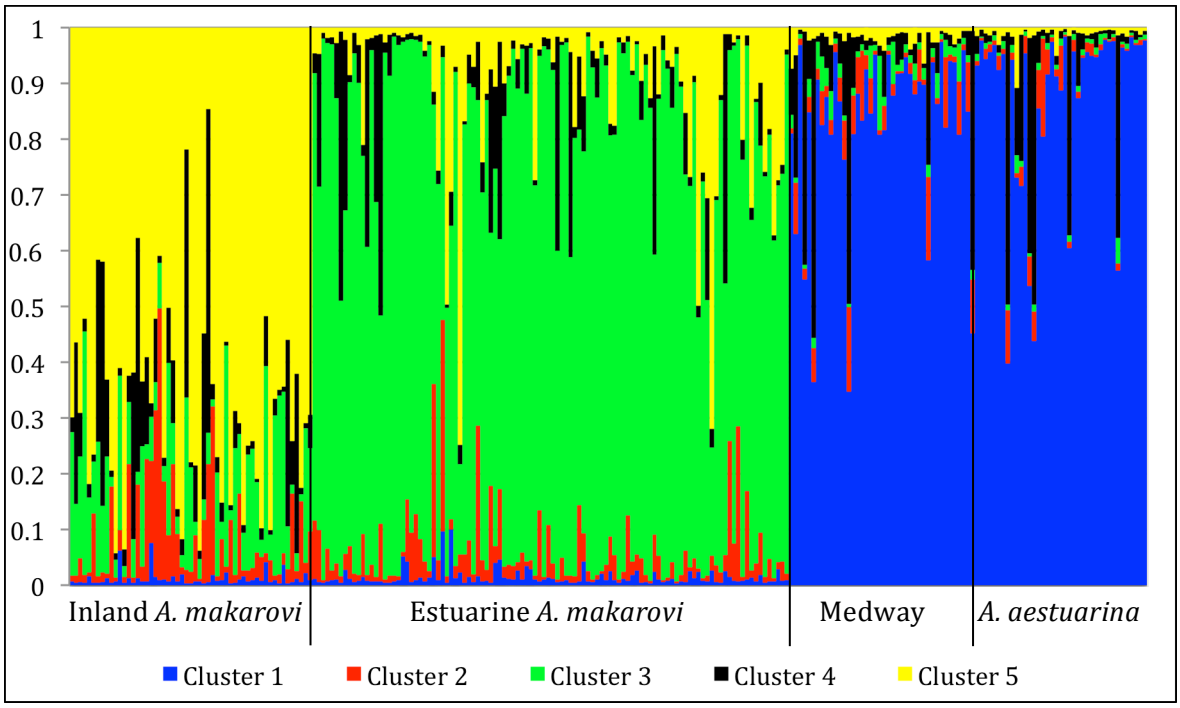


Fig. 5.18. Assignment probabilities using correlated allele frequencies for *A. makarovi* populations from inland and estuarine sites, all *A. aestuarina* populations and the mismatched specimens from the Medway estuary to five clusters, $K = 5$.

When using uncorrelated allele frequencies the ΔK method (Fig. 5.19) indicates the true number of genetic clusters to be two ($K = 2$). For $K = 2$ assignment results see Fig. 5.9, Section 5.4.2.1. The maximum log probability (see value highlighted in Table 5.8) is shown at $K = 3$. For $K = 3$ assignment results two genetic clusters were identified (with high assignment probability) with virtually no assignment of individuals to the third cluster. Results when $K = 3$ were very similar to those seen in Fig. 5.9 for $K = 2$ and are therefore not repeated.

Table 5.8. Log probability values given by STRUCTURE using uncorrelated allele frequencies, averaged over 10 replicate runs for each value of *K* tested ($K = 1$-10).

K	Log probability	*K*	Log probability
1	-45439.43	6	-41700.13
2	-41661.73	7	-41725.71
3	**-41651.49**	8	-41752.93
4	-41656.54	8	-41799.04
5	-41677.56	10	-41818.77

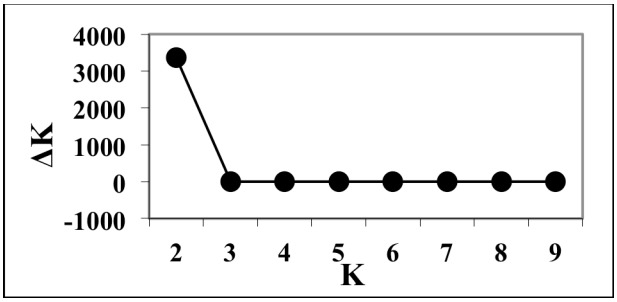


Fig. 5.19. ΔK values for *K* genetic clusters ($K = 2$-9).

2. *A. aestuarina* and the mismatched Medway estuary population.

The Δ*K* method (Fig. 5.20) indicates two genetic clusters (*K* = 2) when comparing *A. aestuarina* and the mismatched Medway population using correlated allele frequencies.. For *K* = 2 assignment results see Fig. 5.13, Section 5.4.2.2. The maximum log probability (see value highlighted in Table 5.9) is shown at *K* = 3 (Fig. 5.21 below). Results were similar to those seen in Fig. 5.13 with some clustering of *A. aestuarina* and the mismatched Medway population, except that a higher degree of admixture can be seen and three individuals show > 0.7 assignment to the third cluster (all originating from different populations).

Table 5.9. Log probability values given by STRUCTURE using correlated allele frequencies, averaged over 10 replicate runs for each value of *K* tested (*K* = 1-10).

K	Log probability	*K*	Log probability
1	-12261.04	6	-12201.41
2	-11870.89	7	-12177.44
3	**-11773.41**	8	-12648.09
4	-11944.42	8	-12594.02
5	-12314.53	10	-13456.03

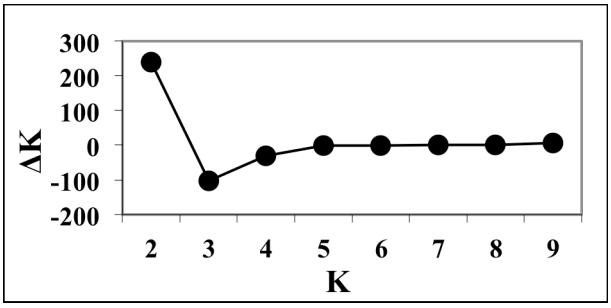


Fig. 5.20. Δ*K* values for *K* genetic clusters (*K* = 2-9).

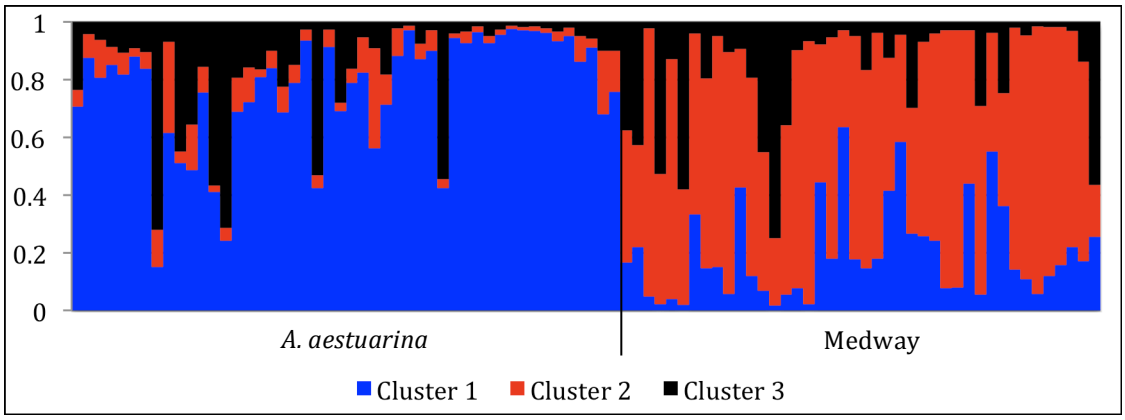


Fig. 5.21. Assignment probabilities using correlated allele frequencies for all *A. aestuarina* populations and the mismatched specimens from the Medway estuary to three clusters, *K* = 3.

When *A. aestuarina* and the mismatched Medway population were tested using uncorrelated allele frequencies both the maximum log probability (see value highlighted in Table 5.10) and the Δ*K* method (Fig. 5.22) indicate the true number of genetic clusters as two (*K* = 3). For *K* = 3 assignment results see Figure 5.14, Section 5.4.2.2.

Table 5.10. Log probability values given by STRUCTURE using uncorrelated allele frequencies, averaged over 10 replicate runs for each value of *K* tested (*K* = 1-10).

K	Log probability	*K*	Log probability
1	-12279.84	6	-12002.61
2	-12293.63	7	-11970.66
3	**-11945.57**	8	-11979.37
4	-11957.6	8	-12086.39
5	-12138.63	10	-11988.97

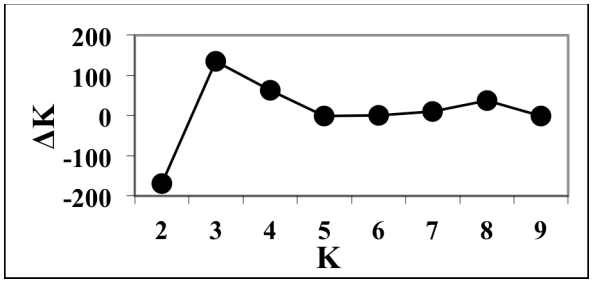


Fig. 5.22. Δ*K* values for *K* genetic clusters (*K* = 2-9).

3. All *A. aestuarina* populations.

Both the maximum log probability (see value highlighted in Table 5.11) and the Δ*K* (Fig. 5.23) indicate the true number of genetic clusters as two (*K* = 3) for analyses comparing all *A. aestuarina* populations using correlated allele frequencies. For *K* = 3 assignment results see Figure 5.15, Section 5.4.2.2.

Table 5.11. Log probability values given by STRUCTURE using correlated allele frequencies, averaged over 10 replicate runs for each value of *K* tested (*K* = 1-5).

K	Log probability
1	-5336.37
2	-5549.44
3	**-5252.4**
4	-5480.74
5	-5801.29

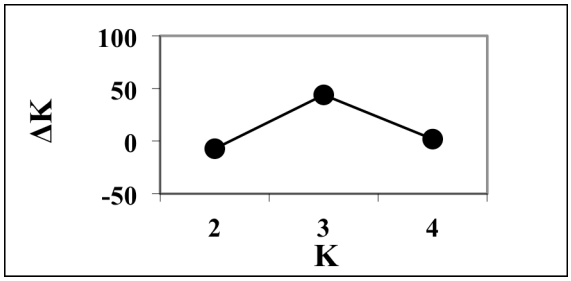


Fig. 5.23. Δ*K* values for *K* genetic clusters (*K* = 2-4).

For analyses comparing all *A. aestuarina* populations, using uncorrelated allele frequencies, no genetic structure was seen, with the highest log probability typically shown at *K* = 1 and very little variation in Δ*K* (results not shown). For values of *K* tested, all individuals were assigned to a single cluster with high probability (> 0.9) and are therefore not reported.

4. Medway estuary mismatched population.

The maximum log probability (see value highlighted in Table 5.12) is shown at $K = 1$ for the mismatched Medway estuary population and correlated allele frequencies. The ΔK method (Fig. 5.24) indicates two genetic clusters ($K = 2$), although very low values of ΔK are shown (<5). Results for $K = 2$ assignment probabilities show no genetic structure with all individuals assigned with similar probabilities to both cluster 1 and cluster 2 (Fig. 5.25).

Table 5.12. Log probability values given by STRUCTURE using correlated allele frequencies, averaged over 10 replicate runs for each value of *K* tested (*K* = 1-5).

K	Log probability
1	**-5777.4**
2	-5827.46
3	-6164.74
4	-6096.43
5	-6295.65

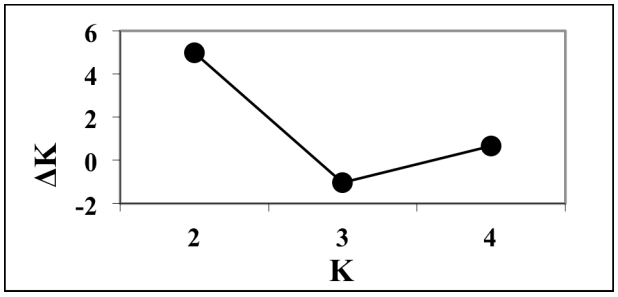


Fig. 5.24. ΔK values for *K* genetic clusters (*K* = 2-4).

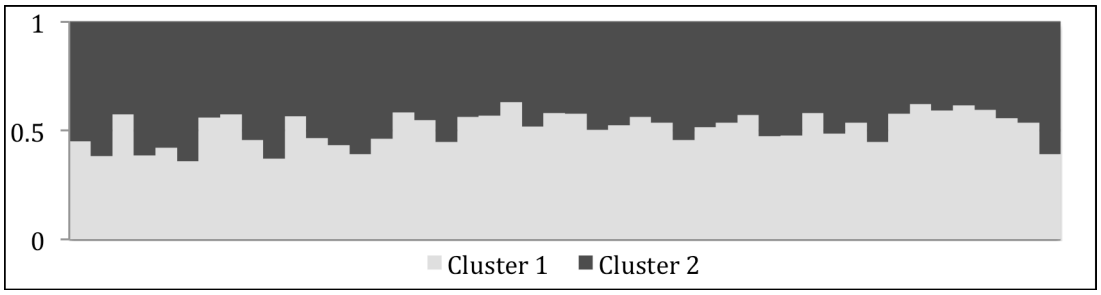


Fig. 5.25. Assignment probabilities using correlated allele frequencies for all of the mismatched specimens from the Medway estuary to two clusters, *K* = 2, showing no genetic structure.

For the mismatched Medway population using uncorrelated allele frequencies, no genetic structure was seen with the highest log probability shown at $K = 1$ and very little variation in ΔK values (results not shown). For values of *K* tested, all individuals were assigned to a single cluster with high probability (> 0.9) and are therefore not shown.

Chapter 6: General conclusions

The main focus of the present study was elucidation of the nature and extent of differentiation and processes involved in shaping diversity within and between species of the *Aphrodes* leafhopper genus, Curtis 1833, from the UK. This chapter presents the key outcomes of this investigation and recommends future research directions.

6.1. *Aphrodes* taxonomy

The morphological similarities among *Aphrodes* species and the multiple synonyms that have been described in the past made species identification problematic before employing the use of additional techniques. It is important to clearly define species and identify the extent of variation within and between taxa prior to addressing evolutionary questions, using a variety of techniques (Hendry *et al.* 2000; Sites & Marshall 2004). Thus, to resolve the status of taxa within the *Aphrodes* complex with confidence, an array of tools were used, including morphological, behavioural and molecular analyses.

A general concordance was found among molecular data and vibrational mating signals recorded for this study (Chapter 3, with the exception of the Medway estuary population) and those previously described for the genus (Tishechkin 1998; Virant-Doberlet *et al.* 2005). The evolution of species-specific mating signals suggests reproductive isolation is near complete, regardless of apparently overlapping morphology. Mating signals are used for species recognition and mate choice by *Aphrodes,* which is likely to facilitate reproductive isolation among the four species (Virant-Doberlet, unpublished data), and have been used to identify other invertebrate taxa (Claridge 1985; Virant-Doberlet & Čokl 2004).

Each species was discovered co-occurring with another *Aphrodes* species in at least one sampling locality giving evidence that they can be found and do remain distinct in sympatry (Chapter 3). In light of the behavioural and genetic differentiation identified among *Aphrodes* in sympatry, despite considerable morphological similarity, most species concepts would recognise that the four *Aphrodes* species clearly represent four

distinct taxa (Coyne & Orr 2004; Mallet 2008). However, Chapter 5 provides evidence that *A. makarovi* and *A. aestuarina* remain distinct despite some level of hybridisation and gene flow in the past.

The species status of *A. aestuarina* (Edwards 1908), which has been previously questioned due to the lack of mating signal data and overlapping morphology with *A. makarovi* (Tishechkin 1998; Biedermann & Niedringhaus 2004), was confirmed using a combination of analyses. Mitochondrial DNA sequences from *A. aestuarina* syntype museum specimens (Bluemel *et al.* 2011 – Chapter 2) were compared with freshly collected specimens from saltmarsh habitats at the syntype location (Norfolk) that had been unequivocally identified as members of a particular species (Chapter 3). As archive specimens kept in museum collections become increasingly important in genetic analyses (e.g. Harper *et al.* 2006; Stuart & Fritz 2008), errors in taxonomic identification can represent a major issue (Graham *et al.* 2004; Wandeler *et al.* 2007). A high number of incorrectly designated specimens were found (Chapter 2), and at least three of the specimens that represent the syntype series for the purported UK endemic species *A. aestuarina* belong to *A. makarovi* (seven of the 13 syntype specimens were tested in Chapter 2 – Bluemel *et al.* 2011). The combined use of historical, ecological, behavioural and molecular data to validate freshly collected specimens that were compared to museum syntype specimens using DNA analysis clearly represents a valuable approach when analysing morphologically similar taxa with a history of taxonomic uncertainty. Only genetic sequences obtained from such specimens can provide a reliable standard for species discrimination of archived material. I propose that validated specimens representing each *Aphrodes* species identified in Chapter 3 are duly deposited as reference material into museum collections for future reference.

The song pattern of *A. aestuarina* is composed of elements also present in *A. bicincta* males found in Aberdare (Fig. 3.4, Virant-Doberlet *et al.* 2005), and female preference experiments indicated no behavioural barrier based on signals between *A. bicincta* and *A. aestuarina* (Virant-Doberlet, unpublished data). The occasional identification of similar or overlapping song features in morphologically cryptic song species has also previously been documented (Henry *et al.* 1999), although only in allopatric species pairs. This pattern may be expected, as prezygotic isolation is likely to evolve more rapidly in sympatric compared to allopatric species pairs (Coyne & Orr 2004). To date, *A. bicincta* has not been found in sympatry with *A. aestuarina* and they are found in

very different habitats (inland and estuarine, respectively) (Chapter 3). The possibility of evolutionary convergence due to similar environmental constraints or selection pressures cannot be ruled out without further investigations of habitat preference, host fidelity and signal transmission properties in different host plant substrates. Alternatively, *A. bicincta* and *A. aestuarina* could exhibit similarities in song features due to the retention of an ancestral (plesiomorphic) state from a more distant common ancestor (Henry *et al.* 1999). They are not members of the same species and do not share the most recent common ancestor based on the mitochondrial DNA (mtDNA) phylogeny presented in Chapter 3 (Fig. 3.9). Random mutations may also explain the convergence seen in vibrational signal song elements in these species if signal traits can be altered considerably by single mutations (Henry *et al.* 1999). Laboratory based crossing and mate choice experiments will provide insights relating to the degree of reproductive isolation among *Aphrodes* species.

6.2. Phylogeography

Four statistically supported monophyletic mtDNA lineages were recovered, which correspond with the species specific vibrational mating signals reported at all localities except the Medway estuary (Chapter 3). The two thoroughly sampled lineages of *A. makarovi* and *A. aestuarina* exhibited shallow phylogeographic structure with a low incidence of locality-specific haplotypes across the UK. For *A. makarovi* a single common haplotype was identified across all sampling localities, which is likely to be ancestral for this species, and a number of less common geographically local haplotypes (Chapter 3), which in turn are likely to be recent mutations that have not spread throughout populations (Avise *et al.* 1987). This pattern was similar to that identified for *A. aestuarina* (Chapter 3). Species with this pattern of mtDNA haplotype distributions indicate phylogeographic continuity and life histories associated with intermediate gene flow with weak long-term barriers to gene flow (Avise *et al.* 1987).

Only a few mutational steps separate mtDNA haplotypes across the geographical area of the UK examined for both *A. makarovi* and *A. aestuarina*. Such star-like haplotype networks (Chapter 5) are often indicative of recent population expansions from a small effective population size (Avise 2000), possibly as the result of population cycles in response to climatic conditions. This was also supported in the mismatch distribution

analysis for *A. makarovi* (but not *A. aestuarina*), although the inferred sudden expansion in population size could not be dated based on the current data (Chapter 5). This geographic expansion may have followed habitat modification associated with recent environmental changes. The low nucleotide diversity and moderate haplotype diversity indices (Chapter 3) for these two *Aphrodes* species are supportive of a bottleneck in recent times (Avise 2000). However, low mtDNA diversity patterns are also expected when natural selection is acting on the mitochondrial genome itself, causing selective sweeps of a single or a few mtDNA haplotypes (Ballard & Whitlock 2004). Tests for departures from neutral evolutionary expectations recovered negative and significant test statistic values for *A. makarovi* and negative but non-significant for *A. aestuarina* mtDNA lineages (Chapter 5), which is an expected pattern for both population bottlenecks and selective sweeps.

Further geographic sampling across the species' ranges and estimation of divergence times that may link with historical climatic changes (Hewitt 2001) could be undertaken to gain a better insights into the phylogeography and mode of speciation in this genus of insects. Additionally, further geographic sampling of *A. bicincta* and *A. diminuta* are required to make any inferences relating to the history of these species as only a single population of each was included in analyses performed in Chapter 3. The identification of *A. diminuta* is the first known record of this species in the UK. *Aphrodes aestuarina* has been documented as endemic to the UK, although specimens thought to be *A. aestuarina* have been reported in coastal saltmarsh habitats in the North and Baltic Sea, in Germany and Poland. Exploration of *Aphrodes* distributions across Europe should be examined by adopting a multidisciplinary identification approach, as used in Chapter 3.

Phylogeographic histories based solely on mtDNA can lead to incorrect conclusions due to other processes (introgression, selective sweeps and cytoplasmic infections) that can bias patterns of mtDNA variation (Funk & Omland 2003; Ballard & Whitlock 2004). Evidence of hybridisation and interspecific mtDNA introgression in the Medway estuary population (6.4 and Chapter 5) suggests that additional nuclear gene markers should also be employed to further explore the phylogeographic history of the *Aphrodes* genus.

6.3. Ecological adaptation and ecotype formation in *Aphrodes makarovi*

Ecological speciation occurs when divergent natural selection on traits between populations in different environments leads to the evolution of reproductive isolation (Schluter 2001). The likelihood of divergent ecologically-driven natural selection promoting genetic and morphological divergence among *A. makarovi* populations inhabiting inland and estuarine habitats on two primary host plants (*Urtica* sp. and *Atriplex portulacoides* respectively) was tested in Chapter 4. Despite the polyphagous feeding habits of this species, significant morphological and nuclear genetic differentiation identified between populations inhabiting inland and estuarine habitats, even in close geographical range, suggests that adaptation may be driven along such steep environmental gradients. Adaptation to a more extreme saltmarsh environment (and the habitat-associated array of host plant species) is likely to be an important factor facilitating divergence in this species.

Significant morphological variation (banding pattern and pigmentation intensity) and nuclear genetic population structure associated with habitat type was identified, relative to that explained by geographic locality (Chapter 4). Inland populations were typically made up of specimens with varied head banding patterns but typically with a higher degree of banding with darker pigmentation than their estuarine counterparts. Estuarine adapted *A. makarovi* showed less varied pigmentation (lighter and less banded), very similar to *A. aestuarina* occurring in the same habitat. Phylogenetic analyses using amplified fragment length polymorphism (AFLP) markers resulted in near monophyletic habitat-associated lineages within *A. makarovi*, revealing the importance of habitat type in structuring the genetic diversity of this species. Mitochondrial DNA sequence data however, revealed no structure relating to habitat or geographic locality. The lack of fixed, divergent AFLP loci or significant mtDNA structure suggests that *A. makarovi* populations have diverged very recently and are in the early stages of sympatric ecotype formation. As speciation is not an inevitable consequence of population differentiation (Nosil *et al.* 2009), further molecular evidence of reduced gene flow is needed to strengthen support for the incidence of sympatric ecological speciation (Schluter 2001; Via 2001; Drès & Mallet 2002).

Ecological speciation might occur in either allopatry or sympatry (Schluter 2001). Thus, the initial divergence of inland and estuarine *A. makarovi* populations in allopatry

cannot be ruled out. However, the observed pattern does not suggest that populations have experienced a long period of vicariance. Further work exploring the historical genetic structuring of *A. makarovi* over a larger geographic area, the degree of host/habitat fidelity, fitness costs associated with each habitat/host plant type, intrinsic genetic incompatibilities and hybrid fitness are needed. Likewise, additional behavioural and genetic analyses along fine-scale transects across locally adjacent/sympatric populations of alternate habitats are required to examine this hypothesis further.

Lastly, further analysis exploring the possible reasons for the maintenance of colour polymorphism (Gray & McKinnon 2007), and apparent convergence in lack of banding pattern and lighter colouration in estuarine adapted *A. makarovi* and *A. aestuarina*, is required. Possible processes involved in facilitating colour/pattern differentiation include random genetic drift, sexual selection or physiological responses to varying predation levels and the need for morphological background crypsis in different habitats (Nosil & Crespi 2006; Rosenblum 2006; Gray & McKinnon 2007).

This study introduces a suitable model system of presently sympatric ecotypes that can be used for further exploration of the relative importance of geography, predation, ecological specialisation driving divergence and speciation of phytophagous insects.

6.4. Hybridisation and mitochondrial introgression

The likelihood of hybridisation and introgression between *A. makarovi* and *A. aestuarina* was tested (Chapter 5) in order to elucidate the possible cause of the mismatch identified between mating signal and mtDNA data in the Medway estuary population (Chapter 3). Unambiguous distinction between *A. makarovi* and *A. aestuarina* was recovered in nuclear AFLP Bayesian clustering analyses. The AFLP results show that the mismatched Medway population is closely related to non-mismatched *A. aestuarina* (Chapter 5), concordant with mating signal data, and therefore with the current taxonomic designations for *Aphrodes* species (Tishechkin 1998; Virant-Doberlet 2005; Chapter 3). Complete fixation of mtDNA from *A. makarovi* was observed in the mismatched Medway estuary *A. aestuarina* population, which is 6.9% (K2P distance) divergent from that of *A. aestuarina* mtDNA found at

other localities. Of the 42 mismatched specimens identified, 95.6% were found to possess the most common *A. makarovi* haplotype (mH1), also present among sympatric Medway *A. makarovi*. Together, these results mean that interspecific mtDNA exchange is likely to explain the reticulate evolutionary pathway for this mismatched population, rather than retention of an ancient ancestral polymorphism or convergence.

Low levels of uni-directional nuclear introgression were observed, hence the hybridisation event is likely to be of historical origin, followed by repeated backcrossing of hybrids with *A. aestuarina.* A number of private AFLP loci (and two private single base pair divergent mtDNA haplotypes based on current sampling effort) were recovered from the mismatched population, providing additional support for a lack of on-going or recent hybridisation. Increased geographic sampling representative of the distributional range of the species is needed to determine how widespread hybridisation among *A. makarovi* and *A. aestuarina* is, or whether *A. aestuarina* is found anywhere else other than in the UK.

Hybridisation is becoming increasingly well acknowledged in natural populations (Mallet 2005) and many studies exploring hybridisation have reported significant mitochondrial introgression with little or no nuclear introgression (Chan & Levin 2005). Thus mtDNA introgression may be more common (Ballard & Whitlock 2004), possibly because foreign mtDNA alleles are relatively neutral and not linked to alleles with associated fitness costs in novel environments and/or genetic backgrounds in comparison to foreign nuclear alleles (Martinsen *et al.* 2001; Funk & Omland 2003).

More research is required to determine among several possible reasons for the fixation of *A. makarovi* mtDNA in this population. Probably a selective sweep of *A. makarovi* mtDNA occurred within *A. aestuarina* populations in the Medway estuary following introgressive hybridisation, due to: a) an unknown direct fitness advantage to possessing *A. makarovi* mtDNA and/or linked nuclear genes in the genetic background of *A. aestuarina*; b) to indirect selection on the mtDNA genome due to cytoplasmic infections commonly identified in many invertebrate taxa (such as *Wolbachia,* Jiggins 2003; Gompert 2008); or c) by chance (i.e. drift). Experiments aimed at identifying factors responsible for the replacement of *A. aestuarina* mtDNA by *A. makarovi* mtDNA in the Medway should look for possible selective or replicative advantage of *A. makarovi*

mtDNA, as well as at the reproductive behaviour and success of the two species of leafhoppers through the use of crossing experiments.

The research results demonstrate that introgressive hybridisation has had substantial and possible long-term effects on the genetic configuration of species, which can produce considerable discrepancies among speciation histories based on nuclear and mitochondrial markers. Phylogenetic analyses made with the use of a single marker, and without consideration of reticulate evolution should be taken cautiously. Adequate taxon sampling is also necessary (Funk & Omland 2003) as also highlighted in this study, and samples should be taken from a number of populations to confirm results. Inference of phylogeographic species histories should ideally be made using multiple markers that are of uni- and bi-parental inheritance. AFLP has proven valuable in allowing a substantial number of genome-wide markers to be amplified per-individual, providing insights into the degree of nuclear DNA divergence that can often be misleading or conflicting when based solely on single gene sequences.

Finally, mitochondrial DNA has an important role in animal biology and due to the interesting qualities of the molecule (outlined in Chapter 1) and the variety of processes that can affect mtDNA patterns, proves this marker to have additional uses beyond its use in phylogenetic and population genetic studies. Clearly mtDNA provides interesting avenues of research and further attempts should be made to explore the ecology and evolution of mtDNA and to understand the nature of selection acting on mtDNA itself (Ballard & Whitlock 2004).

6.1. References

Avise, J. C., Arnold, R. M., Ball, E., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A., & Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review Ecology and Systematics*, *18*, 489-522.

Avise, J. C. (2000). *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*, 729-744.

Biedermann, R., & Niedringhaus, R. (2004). *Die Zikaden Deutschlands*. WABV, Scheeßel, (pp. 214-215).

Bluemel, J. K., King, R. A., Virant-Doberlet, M., & Symondson, W. O. C. (2011). Primers for identification of type and other archived specimens of *Aphrodes* leafhoppers (Hemiptera, Cicadellidae). *Molecular Ecology Resources, 11*, 770-774

Chan, K. M. A, & Levin, S. A. (2005). Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution*, *59*, 720-729.

Claridge, M. F. (1985). Acoustic signals in the Homoptera: behaviour, taxonomy, and evolution. *Annual Review of Entomology, 30*, 297-317.

Coyne, J. A., & Orr, H. A. (2004). *Speciation.* Sinauer Associates, Sunderland, USA.

Drès, M., & Mallet, J. (2002). Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *357*, 471-492.

Funk, D. J., & Omland, K. E. (2003). Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, *34*, 397-423.

Gompert, Z., Forister, M. L., Fordyce, J. A., & Nice, C. C. (2008). Widespread mito-nuclear discordance with evidence for introgressive hybridization and selective sweeps in *Lycaeides*. *Molecular Ecology*, *17*, 5231-5244.

Graham, C. H., Ferrier, S., & Huettman, F., Moritz, C, & Peterson, A. T. (2004). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution, 19,* 497–503.

Gray, S. M., & McKinnon, J. S. (2007). Linking color polymorphism maintenance and speciation. *Trends in ecology & evolution*, *22*, 71-9.

Harper, G. L., Maclean, N., & Goulson, D. (2006). Analysis of museum specimens suggests extreme genetic drift in the Adonis blue butterfly (*Polyommatus bellargus*). *Biological Journal of the Linnean Society*, *88*, 447-452.

Hendry, A. P., Vamosi, S. M., Latham, S. J., Heilbuth, J. C., & Day, T. (2000). Questioning species realities. *Conservation Genetics*, *1*, 67-76.

Henry, C. S., Brooks, S. J., Johnson, J. B., & Duelli, P. (1999). Revised concept of *Chrysoperla mediterranea* (Hölzel), a green lacewing associated with conifers: courtship songs across 2800 kilometres of Europe (Neuroptera: Chrysopidae). *Systematic Entomology*, *24*, 335-350.

Hewitt, G. M. (2001). Speciation, hybrid zones and phylogeography - or seeing genes in space and time. *Molecular Ecology*, *10*, 537-549.

Jiggins, F. M. (2003). Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics*, *164*, 5-12.

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution*, *20*, 229-237.

Mallet, J. (2008). Hybridisation, ecological races and the nature of species: empirical evidence for the ease of speciation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363,* 2971-2986.

Martinsen, G. D., Whitham, T. G., Turek, R. J., & Keim, P. (2001). Hybrid populations selectively filter gene introgression between species. *Evolution*, *55*, 1325-1335.

Nosil, P., & Crespi, B. J. (2006). Experimental evidence that predation promotes divergence in adaptive radiation. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 9090-9095.

Nosil, P., Harmon, L. J., & Seehausen, O. (2009). Ecological explanations for (incomplete) speciation. *Trends in Ecology & Evolution*, *24*, 145-156.

Rosenblum, E. B. (2006). Convergent evolution and divergent selection: lizards at the white sands ecotone. *The American Naturalist*, *167*, 1-15.

Schluter, D. (2001). Ecology and the origin of species. *Trends in Ecology and Evolution*, *16*, 372-380.

Sites, J. W., & Marshall, J. C. (2004). Operational criteria for delimiting species. *Annual Review of Ecology, Evolution, and Systematics*, *35*, 199-227.

Stuart, B. L., & Fritz, U. (2008). Historical DNA from museum type specimens clarifies diversity of Asian leaf turtles. *Biological Journal of the Linnean Society, 94*, 131–141.

Tishechkin, D. Y. (1998). Acoustic signals and morphological characters of leafhoppers belonging to *Aphrodes bicinctus* group (Homoptera, Cicadellidae) from European Russia. *Entomological Review, 78*, 370-377.

Via, S. (2001). Sympatric speciation in animals: the ugly duckling grows up. *Trends in Ecology & Evolution*, *16*, 381-390.

Virant-Doberlet, M., & Čokl, A. (2004). Vibrational communication in insects. *Neotropical Entomology, 33*, 121–134.

Virant-Doberlet, M., Prešern, J., Stewart, A. J. A., & Wilson, M. R. (2005). The genus *Aphrodes*: bioacoustic approach to solve taxonomic problems. Congress of International BioAcoustic Council (IBAC). (pp. 35).

Wandeler, P., Hoeck, P. E. A., & Keller, L.F. (2007). Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution, 22*, 634–642.