The Aquatic Leaf Beetle *Macroplea mutica* (Coleoptera: Chrysomelidae): Population Structure in Europe and the Signature of Zoochorous Dispersal

Dissertation with the aim of achieving a doctoral degree
at the Faculty of Mathematics, Informatics and Natural Sciences
Department of Biology
of Universität Hamburg

submitted by
Jan-Jakob Laux

Hamburg, 2014

Genehmigt vom Fachbereich Biologie
der Fakultät für Mathematik, Informatik und Naturwissenschaften
an der Universität Hamburg
auf Antrag von Frau Professor Dr. S. DOBLER
Weiterer Gutachter der Dissertation:
Priv.-Doz. Dr. G. KÖLSCH
Tag der Disputation: 16. Februar 2015

Professor Dr. C. Lohr
Vorsitzender des
Fach-Promotionsausschusses Biologie

Abstract

Most aquatic invertebrates are incapable of actively dispersing between isolated wetlands. Mechanisms of passive transportation are therefore of great ecological significance for the biotic connectivity of freshwater habitats.

The fully aquatic leaf beetle *Macroplea mutica* (FABRICIUS, 1792) shows a wide Palearctic distribution despite being flightless and hardly able to walk when out of water. Range expansion and dispersal between isolated wetlands therefore have to involve mechanisms of passive transport.

The present study was dedicated to the analysis of the population genetic structure of *M. mutica* in Northern Europe, with respect to postglacial colonization, signatures of passive dispersal and special focus on the possibility of passive transport by waterbirds (zoochory).

Six highly polymorphic microsatellite markers were developed for *M. mutica* and used for population genetic analysis of 21 European *M. mutica* populations. As expected due to the low mobility of the species, genetic differentiation was strong across populations. Cluster analyses showed a clear hierarchical population structure with a western cluster (containing populations from Great Britain, the Netherlands and north-western Germany) and an eastern cluster (encompassing eleven Baltic Sea populations and two Danish inland populations). A zone of exceedingly strong genetic differentiation between Baltic Sea populations and closely neighbouring inland sites suggests a contact zone between two postglacial colonization waves. Geographic structure in analysed sequences of the mitochondrial cytochrome oxidase subunit 1 gene (COI) was, however, very low and did therefore not corroborate the hypothetical existence of separate glacial refugia.

Feeding trials with mallards (*Anas platyrhynchos* L.) showed that eggs of *M. mutica*, ingested with parts of its host-plant, are capable of viably passing through the digestive system of ducks, surviving retention times of at least five to eight hours and thereby demonstrating a clear potential for internal transport by waterbirds.

To test for genetic evidence for waterbird-mediated dispersal, population genetic structure of *M. mutica* was mapped against movements and local abundances of a potential vector species. More than 260,000 geo-referenced sightings of individually marked mute swans (*Cygnus olor* GMELIN) were analysed with focus on visitation of sampled *M. mutica* populations and predominant migration routes. A subsequent comparison with pairwise genetic distances across *M. mutica* populations showed that inferred movements and local abundances of mute

swans are consistent with a significant impact of waterbird-mediated dispersal on population genetic structuring in *M. mutica*.

Swan movements among sampled inland sites were found to be better predictors for genetic structure in *M. mutica* than geographic distance, and local swan abundances showed significant negative correlations with pairwise genetic differentiation across *M. mutica* populations in the Baltic Sea. The population genetic data furthermore showed the genetically isolating effect of geographic distance significantly decreasing with increasing swan abundances and the breakdown of isolation by distance between sampling sites with high swan abundances.

The results suggest that *M. mutica* is a rare example of zoochorous dispersal in aquatic insects, represent first evidence for waterbird-mediated dispersal of an aquatic beetle and corroborate the ecological significance of this mode of transport for a broad spectrum of aquatic invertebrate taxa.

Zusammenfassung

Aquatische Wirbellose sind meist nicht in der Lage sich aktiv zwischen isolierten Habitaten zu bewegen. Mechanismen passiven Transports sind daher von großer ökologischer Bedeutung für die biotische Konnektivität limnischer Lebensräume.
Der vollständig aquatisch lebende Blattkäfer *Macroplea mutica* (Fabricius, 1792) zeigt eine weiträumige paläarktische Verbreitung, obwohl die Art flugunfähig ist und außerhalb des Wassers kaum laufen kann. Passive Transportmechanismen dürften für diese Spezies daher von großer Bedeutung für Ausbreitung und Austausch zwischen isolierten Feuchtgebieten sein.

Die vorliegende Studie widmet sich der Analyse der populationsgenetischen Struktur von *M. mutica* in Nordeuropa, in Hinblick auf nacheiszeitliche Besiedlung, Spuren passiver Ausbreitungsmechanismen und mit besonderem Augenmerk auf die Möglichkeit von passivem Transport durch Wasservögel (Zoochorie).

Die populationsgenetische Struktur von 21 europäischen *M. mutica* Populationen wurde auf Basis von sechs neu entwickelten, hoch polymorphen Mikrosatelliten-Markern untersucht. Wie aufgrund der geringen Mobilität der Art zu erwarten, zeigte sich hierbei starke populationsgenetische Differenzierung. Clusteranalysen offenbarten zudem eine klare hierarchische Populationsstruktur, mit einem westlichen Cluster (bestehend aus Populationen aus Großbritannien, den Niederlanden und Nordwestdeutschland) sowie einem östlichen Cluster (bestehend aus elf Ostseepopulationen und zwei dänischen Binnen-Populationen). Eine Zone überproportional starker genetischer Differenzierung zwischen Ostseepopulationen und eng benachbarten Binnenlandpopulationen lässt dabei auf eine Kontaktzone zwischen zwei nacheiszeitlichen Besiedlungswellen schließen. Eine Untersuchung von mitochondriellen DNS-Sequenzen für einen Abschnitt des Gens der Cytochrom-Oxidase Untereinheit I (COI) zeigte allerdings kaum geographische Struktur und lieferte somit keine Belege für die hypothetische Existenz getrennter glazialer Refugien.

In Fütterungsversuchen mit Stockenten (*Anas platyrhynchos* L.) konnte gezeigt werden, dass mitsamt Teilen der Wirtspflanze verfütterte Eier von *M. mutica* intakt den Verdauungstrakt von Enten passieren können. Hierbei überleben sie nachweislich Retentionszeiten von mindestens fünf bis acht Stunden und zeigen somit ein klares Potenzial für den internen Transport durch Wasservögel.

Um auf genetische Spuren der Verbreitung durch Wasservogel zu testen, wurde die populationsgenetische Struktur von *M. mutica* mit Bewegungen und lokalen Abundanzen einer mutmaßlichen Vektor-Spezies verglichen. Mehr als 260.000 georeferenzierte Sichtungen individuell markierter Höckerschwäne (*Cygnus olor* GMELIN) wurden mit Fokus auf das Aufsuchen von *M. mutica*- Populationen und vorherrschende lokale und saisonale Migrationsrouten ausgewertet. Ergebnisse des Abgleichs erfasster Bewegungen und lokaler Abundanzen von Höckerschwänen mit paarweisen genetischen Distanzen zwischen *M. mutica*-Populationen sprechen für einen deutlichen Einfluss der Verbreitung durch Wasservögel auf die populationsgenetische Struktur von *M. mutica*: Erfasste Schwan-Bewegungen erwiesen sich als bessere Prädiktoren für die genetische Struktur von *M. mutica*-Binnenlandpopulationen als geographische Distanzen. Lokale Schwan-Abundanzen zeigten außerdem signifikante negative Korrelationen mit paarweisen genetischen Distanzen zwischen *M. mutica*-Populationen in der Ostsee. Darüber hinaus nahm der messbare, genetisch isolierende Effekt geographischer Distanz signifikant mit zunehmender Schwan-Abundanz ab und fehlte gänzlich zwischen Käferpopulationen an Standorten mit hohem lokalem Schwan-Aufkommen.

Die vorliegenden Ergebnisse sprechen für ein seltenes Beispiel von Ausbreitung durch Zoochorie bei einem aquatischen Insekt, repräsentieren erste Hinweise auf Ausbreitung eines aquatischen Käfers durch Wasservögel und bekräftigen die ökologische Bedeutung dieses Ausbreitungsmechanismus für ein breites Spektrum von aquatischen Wirbellosen.

List of Figures

List of Tables

1- Introduction

1.1 - Dispersal

Ecological conditions that allow for a species to thrive and persist in a certain habitat are only ever temporarily constant, ever-changing in a dynamic environment. Suitable habitats are furthermore not only temporarily but usually also spatially discontinuous. The movement in the landscape is therefore of highest importance to the long-term survival of any population. Range extension and colonization of new habitats can be necessary to escape deteriorating conditions in a current habitat or distribution range and to minimize pressure by predation (McKinnon *et al.*, 2010) or risk of disease (Altizer *et al.*, 2011). Gene flow between (sub-) populations can furthermore be crucial to prevent genetic impoverishment that would otherwise threaten (meta-) population survival (Hamilton & May, 1977; Fayard *et al.*, 2009). Biological dispersal, defined as directional movement away from a source to either establish or reproduce is therefore among the most essential processes influencing ecology and evolution (Dieckmann *et al.*, 1999). Dispersal occurs at all spatial levels, from small-scale locomotion of micro fauna (e.g. Thomas & Lana (2011)) over diel migration of zooplankton in the water column (e.g. Haney (1988)) to annual trans-continental migration of birds (e.g. Egevang *et al.* (2010)), and greatly influences the fate of individuals, populations and whole ecosystems (Colbert, 2001).

Dispersal capacities are, however, clearly limited in most species and global or local bio-diversity often does not reflect the entirety of suitable habitats for a given species (Kokko & López-Sepulcre, 2006) and rather represents the suitable habitats within a range that were reached. Landscape features (amongst other ecological factors) can act as barriers to dispersal and therefore greatly influence local bio-diversity and the exchange between con-specific populations and even trigger speciation events (Wiley, 1988).

Besides the *active* dispersal of animals by means of locomotion (autochory) diverse biotic and abiotic vectors can facilitate the *passive* dispersal of organisms. And although the recognition of the potential ecological significance of passive dispersal goes at least as far back as Darwin (1859) there is still much to know in order to understand how these mechanisms shape global biodiversity and the distribution and population structure of individual species.

Recent developments have led to an increasing importance of understanding mechanisms of dispersal. In times of rapid anthropogenic changes to ecosystems, the importance of dispersal

increases for species that are threatened by deterioration and fragmentation of suitable habitats (Amezaga *et al.*, 2002; Pearson, 2006). Furthermore, anthropogenic changes in the dispersal capacity of certain species (in the course of climate change-induced range changes or due to anthropogenic long-distance dispersal) are increasingly recognized to be of major ecologic and economic consequence, threatening local indigenous biodiversity, and ecosystem services world-wide (Pimentel *et al.*, 2005; Strayer, 2010; Vilà *et al.*, 2011).

1.1.1 - Passive Dispersal of Freshwater Organisms

Organisms in freshwater habitats are especially challenged by the need to disperse. Freshwater habitats are often relatively short-lived (De Meester *et al.*, 2002) and characterized by especially frequently and quickly fluctuating abiotic and biotic conditions, necessitating frequent (re-) colonization of suitable habitats to ensure meta-population survival (Fronhofer *et al.*, 2012). Darwin (C. Darwin in Darwin, 1909) coined the expression "islands in a sea of land" to describe the isolation of high altitude mountain habitats. In a very similar fashion isolated wetlands and ponds can be said to be "islands in a sea of land". In both cases the heterogeneous distribution of a certain habitat type in the landscape creates strong barriers to unlimited dispersal. Continental wetlands are usually divided by large areas of unsuitable habitat that keep their inhabitants from dispersing freely between catchments. Therefore dispersal is as difficult as vitally important for freshwater organisms. This might be increasingly the case due to anthropogenic wetland deterioration, which further increases fragmentation and loss of freshwater habitats. A better understanding of vectors and mechanisms facilitating the dispersal of aquatic organisms may therefore be of key importance for conservation efforts aiming to preserve diversity and functioning of wetland ecosystems (Santamaría & Klaassen, 2002).

Despite their short-lived and spatially isolated nature, freshwater habitats can show high biodiversity and are colonized by many species with low potential for active dispersal, often within short periods of time (Cáceres & Soluk, 2002). The extent of global distribution of many freshwater species might have been overestimated and falsely been considered to be cosmopolitan before molecular genetic methods improved insight into their taxonomy (Bohonak & Jenkins, 2003). Nevertheless, many freshwater species show wide distributions (Santamaría, 2002), especially considering the fact that they lack the capacity to actively disperse between catchments and isolated water bodies. This makes freshwater ecosystems especially interesting for the study of passive dispersal mechanisms.

Vectors Facilitating the Passive Dispersal of Aquatic Organisms

The great majority of aquatic organisms require mechanisms of passive transport to disperse between water bodies. This passive dispersal is facilitated by a number of different vectors, involving transport with flowing water (hydrochory), transport with wind (anemochory) and transport by different species of animal (zoochory).
While transport with wind is relatively random (lacking directionality), in most cases only suitable for comparatively small propagules (Van Leeuwen, 2012) and therefore might be neither a widespread nor a frequent phenomenon (Bohonak & Jenkins, 2003), transport with flowing water represents an important passive dispersal mechanism for most aquatic organisms, largely because propagules are transported within a medium that does not adversely affect survival (Van Leeuwen, 2012). Hydrochorous dispersal between different catchments is, however, largely limited to rather extreme flooding events and within hydrologically connected systems it is often unidirectional (downstream in lotic environments or along predominant currents in larger lentic water bodies) (Van Leeuwen, 2012; Srivastava & Kratina, 2013). Zoochorous dispersal is therefore of immense importance for organisms in aquatic habitats. And while numerous different animal taxa can potentially act as dispersal vectors for aquatic organisms (e.g. insects (Beladjal & Mertens, 2009), fish (Horn, 1997; Pollux, 2011) and mammals (Waterkeyn *et al.*, 2010; Vanschoenwinkel *et al.*, 2011; Van Leeuwen *et al.*, 2013)) waterbirds are generally considered to be of especially ubiquitous importance (Figuerola & Green, 2002; Green *et al.*, 2002; Bohonak & Jenkins, 2003; Nathan, 2006).

1.1.2 - Dispersal of Aquatic Organisms by Waterbirds

The main reason for the outstanding suitability of waterbirds as dispersal vectors are their frequent, fast and often long-distance movements between ecologically similar aquatic habitats (Figuerola & Green, 2002; Green *et al.*, 2002; Bohonak & Jenkins, 2003; Nathan, 2006). Waterbirds therefore surpass most other vectors in terms of directionality of transport (Van Leeuwen, 2012; Van Leeuwen *et al.*, 2012b).
Charles Darwin was among the first to suggest that migrating water birds might facilitate the dispersal of aquatic invertebrates and plants (Darwin, 1859). By now, more than 150 years later, waterbird-mediated dispersal is widely recognized to be of great ecological significance, facilitating colonization events and gene flow in aquatic organisms, often over long distances

and across geographical barriers (Figuerola & Green, 2002; Green *et al.*, 2002; Bohonak & Jenkins, 2003; Nathan, 2006). Due to the great importance of waterbird-mediated dispersal of aquatic organisms for biotic wetland connectivity (Amezaga *et al.*, 2002), waterfowl conservation might be crucial to the preservation of wetland biodiversity (Amezaga *et al.*, 2002; Bohonak & Jenkins, 2003).

The earliest studies on waterbird-mediated dispersal focused on the possibility of external transport of propagules (epizoochory or ectozoochory) attached to the plumage, beaks or feet of waterbirds (Darwin, 1859; de Guerne, 1887, 1888) (for a review of ectozoochory, see Sorensen (1986)). After Brown (1933) showed that bryozoan statoblasts fed to mallards (*Anas platyrhynchos* L.) were defecated in viable condition, it was slowly recognized that certain life stages of many aquatic plants and animals can survive being ingested and later defecated by waterbirds. This form of passive internal dispersal (endozoochory) is increasingly understood as an important and potent mechanism for the long-distance dispersal of many aquatic plants and animals (Figuerola & Green, 2002; Green & Figuerola, 2005; Brochet *et al.*, 2009, 2010a, 2010b, 2010d; Van Leeuwen *et al.*, 2012b; Green *et al.*, 2013; Van Leeuwen *et al.*, 2013), and held to be quantitatively even more important than external transport by waterbirds (Brochet *et al.*, 2010b; Sánchez *et al.*, 2012).

Figuerola and Green (2002) name three important requirements for internal dispersal by birds. Provided that an organism (1) is regularly ingested by birds, (2) is then capable of surviving gut passage while (3) remaining in the digestive tract long enough to be transported over considerable distances, this results in a very effective means of passive dispersal. Concordantly, the potential for internal dispersal is often assessed experimentally by feeding birds with a known quantity of propagules and subsequently examining faeces for retrieval of viable organisms (Charalambidou *et al.*, 2003; Brochet *et al.*, 2010c; Van Leeuwen *et al.*, 2012a, 2012c; Wada *et al.*, 2012). By monitoring the proportion of surviving propagules and the timing of retrieval, survival rates of maximum dispersal distances can be estimated. This approach also allows assessing the dispersal potential for propagules that do not occur in great densities in the wild and are therefore unlikely to be discovered in faecal samples in the field.

Dispersal of Aquatic Insects by Waterbirds

Internal transport by waterbirds has been shown to work for a large number of aquatic invertebrate taxa (e.g. Crustacea, Bryozoa, Gastropoda) (Figuerola & Green, 2002; Frisch *et al.*, 2007a; Brochet *et al.*, 2010a; Van Leeuwen *et al.*, 2012b). Zoochorous transport has, however, very rarely been connected to the dispersal of aquatic insects.

To the best of the author's knowledge, there are only two published accounts of internal dispersal of aquatic insects. Living chironomid larvae have been found in the faeces of waders (Green & Sanchez, 2006) and tipulid larvae have been reported from the faeces of coot (Frisch et al., 2007). Corixid eggs have been observed in waterbird faeces (Figuerola *et al.*, 2003). There is, however, no evidence that they are excreted while still viable. Further published evidence of internal transport of insects seems to exclusively concern terrestrial insects. Larvae of seed-inhabiting wasps and weevils are known to be internally dispersed by frugivore birds and mammals (Hernández & Falcó, 2008; Hernández, 2011).

Aquatic insects might have been overlooked in this context, as many otherwise fully aquatic insects still possess the ability to fly and are therefore mostly capable of active dispersal between water bodies (Bilton *et al.*, 2001). Furthermore, most published evidence for bird-mediated dispersal of aquatic invertebrates focuses on internal dispersal (Van Leeuwen *et al.*, 2012b) and most authors assumed that only invertebrates with physically and chemically resistant resting stages (resting eggs, cysts or ephippia) are capable of surviving the involved gut passage (e.g. Bilton *et al.*, 2001).

Recent studies have, however, revealed a potential importance of bird-mediated internal dispersal for organisms lacking obvious adaptations to surviving gut passage. Since digestive processes in waterbirds show high plasticity (Charalambidou *et al.*, 2005) and a tendency for reduced digestive efficiency and retention times in favour of maximized net energy intake (Van Leeuwen *et al.*, 2012a, 2012b) bird-mediated internal transport might be of strong significance for different life stages of a broad spectrum of aquatic invertebrates that have hitherto not been associated with this mode of dispersal (Green & Sánchez, 2006; Frisch *et al.*, 2007b; Wada *et al.*, 2012). In this context, the internal dispersal of aquatic insects is therefore a topic that deserves further attention.

1.2 - Population Genetic Analyses in Dispersal Ecology

One of the key aspects of dispersal is its role in facilitating gene flow between populations. Gene flow determines the relative effects of selection and drift on populations, homogenizes allelic frequencies and impedes and the fixation of alleles by local selection and genetic drift (and therefore genetic divergence and ultimately speciation) (Barton & Hewitt, 1985). In absence of sufficient gene flow populations are furthermore negatively affected in their evolutionary potential to resist the fixation of deleterious mutations (Wright, 1977; Frankham & Ralls, 1998; Higgins & Lynch, 2001). Due to the great ecological and evolutionary

importance of genetic exchange between populations, the understanding of gene flow is of major interest to numerous fields of research (e.g. population genetics, population ecology, conservation biology and epidemiology). And since patterns of genetic differentiation among populations often largely reflect genetic exchange through migration, the study of an organism's population genetic structure is also of obvious importance for tackling questions of dispersal ecology.

In order to understand how dispersal shapes population structure of a given species, it is often valuable to learn in how far certain landscape characteristics match population genetic data. To these ends landscape genetic approaches often utilize geographic information systems (GIS). GIS-based methods allow landscape variables to be overlaid onto genetic data and provide various geostatistical tools for interpolation (e.g. QUANTUM GIS (Quantum GIS Development Team, 2012)). A number of advanced GIS programs and a vast number of compatible geo-referenced geographic and ecological data sets are available free of charge and provide valuable tools for a wide variety of scientific applications. For reviews of available free and open-source GIS software see Steiniger & Hay, (2009) and Steiniger & Hunter (2012), for sources of free GIS-based environmental data see (Kozak *et al.*, 2008).

Geo-referenced data for population genetic structure of an organism can be mapped against landscape features and other environmental variables, in order to identify geographic features acting as barriers to gene flow and identify key factors for local adaptation (see e.g. Manel *et al.*, 2003; Spear *et al.*, 2005; Finn *et al.*, 2006; Pérez-Espona *et al.*, 2008) or to reconstruct (post-glacial) colonization histories (Taberlet *et al.*, 1998; Schmitt *et al.*, 2002, 2005; Adams *et al.*, 2006; Schmitt, 2007; Westberg & Kadereit, 2009; Theissinger *et al.*, 2013). To detect genetic traces of vector mediated dispersal events, data on genetic diversity (Wada *et al.*, 2012; Triest & Sierens, 2013) or genetic differentiation (Mader *et al.*, 1998; King *et al.*, 2002; Figuerola *et al.*, 2005; Van Leeuwen *et al.*, 2013) of studied organisms can be tested for correlation with the abundance or movements of potential vectors.

1.2.1 - Detecting Genetic Evidence of Vector-Mediated Dispersal

Trying to detect genetic traces of vector-mediated dispersal events can be a complex task and it is crucial to understand that observed genetic differentiation or differences in diversity between populations is not necessarily equivalent to the amount of dispersal of individuals (Bohonak & Jenkins, 2003). How much a given dispersal event adds to detected genetic differentiation can be strongly influenced by a number of potentially confounded factors.

Locally adapted individuals might sometimes outcompete immigrants, reducing immigrant reproductive success and therefore gene flow after dispersal (Orsini *et al.*, 2013a, 2013b) especially in combination with strong priority effects (De Meester *et al.*, 2002). As a result, population genetic structure can exceedingly reflect colonization events rather than contemporary gene flow, when it is strongly influenced by resilient founder effects (Orsini *et al.*, 2013b; Spurgin *et al.*, 2014).

It is furthermore of great importance to consider that patterns of spatial autocorrelation are common features of studied population genetic structure. Dispersal is ultimately strongly limited in most organisms and since even organisms with strong potential for dispersal over longer distances usually show distributions that exceed the individual's capacity for dispersal, patterns of spatial genetic structure often reflect a decrease of dispersal probability (and therefore gene flow and genetic relatedness) with increasing geographic distance. This common phenomenon is called *isolation by distance* (IBD). And since IBD is largely a function of limitations of an organisms dispersal range in the landscape, the influence of vector-mediated (long-distance) dispersal can be detected as changes to the patterns of IBD if the presence of dispersal vectors increases possible dispersal distances.

Studies reporting genetic evidence of vector-mediated dispersal therefore often base evidence on observations of changing patterns of spatial genetic structure rather than "raw" genetic differentiation The presence of zoochorous dispersal vectors could thereby be shown to reduce the genetically isolating effect of geographic distance in a transported organism (Mader *et al.*, 1998; King *et al.*, 2002) or cause patterns of IBD to breakdown altogether (Van Leeuwen *et al.*, 2013). Inferred routes of vector movement have also been shown to provide a better fit with population genetic distances in dispersed organisms than sheer geographic distance (Figuerola *et al.*, 2005).

Furthermore, spatial genetic structure can show strongly hierarchical patterns (i.e. two or more clusters of differentiated populations). While freshwater organisms can show such patterns because of hierarchic habitat structure (Bohonak & Jenkins, 2003), hierarchic population structure is often the result of postglacial colonization from multiple refugia (Meirmans, 2012). Hierarchical population structure and isolation by distance both represent forms of spatial autocorrelation in genetic data that can be informative for questions of gene flow and colonization history. It is, however, important to take the spatial dependence of investigated population genetic data into account to avoid potential bias in statistical tests (e.g. tests for association with mapped environmental variables) (Meirmans, 2012).

1.2.2 - Population Genetic Markers

Modern population genetic methods provide a variety of techniques to study patterns of genetic differentiation or diversity across (sub-) populations and scientists can choose from a number of available population genetic markers, such as amplified fragment length polymorphisms (AFLPs), microsatellites and mitochondrial DNA (mtDNA). Appropriate genetic markers should allow for revealing processes shaping population genetic structure within the time period of interest. While microsatellites are especially fast evolving markers that are well suited for revealing contemporary patterns of gene flow between conspecific populations (Selkoe & Toonen, 2006), other markers (such as mitochondrial DNA) can be more appropriate to infer genetic divergence on longer time scales or higher taxonomic levels (Hebert *et al.*, 2003a, 2003b; Papadopoulou *et al.*, 2010).

1.2.2.1 - Microsatellites

Microsatellites are among the most popular and versatile genetic markers in ecology and population genetics (Selkoe & Toonen, 2006). Also known as simple sequence repeats (SSR) or short tandem repeats (STR), microsatellites are stretches of nuclear DNA characterized by tandem repeats of 1–6 nucleotides which occur at high frequency in most taxa. Microsatellite loci tend to be highly polymorphic due to frequent mutation by slippage and proofreading errors during DNA replication. These mutations change the number of repeats and thereby the length of the microsatellite region. Since the DNA regions flanking the microsatellite are usually conserved within species (or across closely related species) they can provide binding sites for fitting oligonucleotides (primers), allowing for the amplification of a microsatellite locus with polymerase chain reaction (PCR). Differences in microsatellite length can subsequently be distinguished by high-resolution electrophoresis of amplification products, providing comparatively easy genotyping and the study of allelic differentiation. Due to their relatively high mutation rates (approximately 10^{-4} per locus per generation on average (Whittaker *et al.*, 2003)) and resulting high allelic diversity, microsatellites permit inference of contemporary levels of gene flow and resolution of comparatively low genetic differentiation across populations.

A drawback of microsatellite loci is, however, that the development of suitable PCR primers requires *de novo* isolation of microsatellite regions for organisms that are studied for the first time and therefore requires considerable amounts of sequence data. Traditionally this involves construction of a genomic library that is enriched for microsatellites (see Zane *et al.* (2002)

for a review of isolation techniques). Recently increasing affordability of new-generation sequencing methods might, however, increasingly render enrichment steps obsolete for most taxa (Silva *et al.*, 2013). Regardless of used isolation techniques, obtaining a working set of microsatellite primers requires testing of developed primers and screening of amplified loci. Once a set of polymorphic microsatellite loci has been established it has to be confirmed that genotyped loci are fundamentally selectively neutral, follow Mendelian inheritance and that amplification allows for unproblematic detection of alleles in order to be used as a tool for detecting demographic patterns. A number of free software tools can aid in this part of the development process (e.g. GENEPOP (Raymond & Rousset, 1995) and MICRO-CHECKER (Van Oosterhout *et al.*, 2004)).

The mode of mutation associated with microsatellite length polymorphisms is arguably stepwise (i.e. usually adding or removing one repeat unit per mutation event). Therefore, allele identity-based measures of genetic differentiation (e.g. derivatives of Wright's F_{ST} (1951)) have been held to reflect mutation at microsatellite loci less accurately than allele size-based estimators that assume a stepwise mutation (derivatives of the F_{ST} analogue R_{ST} (Slatkin, 1995)). Recent studies showed, however, that patterns of microsatellite mutation are often likely to be less simple (Ellegren, 2000a, 2000b, 2004). For analysing microsatellite data, allele identity-based and allele size-based estimators of genetic differentiation differ considerably in their performance, depending on the relative significance of stepwise mutation processes to the studied genetic differentiation. It can therefore be advisable to test if a studied allele distribution fits a stepwise mutation model, in order to choose appropriate measures of genetic differentiation when analysing microsatellite data (Balloux & Lugon-Moulin, 2002; Hardy *et al.*, 2003).

1.2.2.2 - Mitochondrial DNA

The study of variation in the mitochondrial DNA (mtDNA) is a well-established and popular approach for reconstructing historical patterns of population demography, admixture, biogeography and speciation. Sequences of mtDNA can relatively easily be amplified in polymerase chain reactions (PCRs) for many taxa and due to (at least) very low rates of recombination can be assumed to largely represent the history of the whole molecule. Because of high mutation rates and an effective population size four times smaller than that of nuclear markers it allows for reconstruction of relatively recent events without extensive sequencing efforts (Hurst & Jiggins, 2005). Since mutation rates of mtDNA are furthermore assumed to be rather constant, mtDNA data is also popular for dating divergence times between taxa. The

sequence of the mitochondrial cytochrome oxidase subunit 1 (COI) gene is an especially popular marker, extensively used for dating of divergence times between taxa (Papadopoulou *et al.*, 2010) and identification of species by DNA-barcoding (Hebert *et al.*, 2003a, 2003b).

1.2.3 - Population Genetic Software Tools

Recent improvements in computing technology have considerably increased possibilities to use intensive statistical approaches such as maximum likelihood, Bayesian probability theory and Monte Carlo Markov chain simulation to (e.g.) detect patterns of gene flow and identify different levels of population genetic structure within population genetic data. Bayesian clustering algorithms (as implemented in software tools like BAPS (Corander *et al.*, 2003), GENELAND (Guillot *et al.*, 2005b) and STRUCTURE (Pritchard *et al.*, 2000) assign (optionally geo-referenced) genotyped samples to genetic clusters and thus infer the structure of population genetics data. Bayesian and maximum likelihood estimates based on coalescents for unequal migration rates and subpopulation sizes (as implemented in MIGRATE-N (Beerli & Felsenstein, 1999, 2001) have improved the inference of gene flow between groups of genotyped samples.

Numerous free software solutions furthermore aid in all steps of population genetic analyses. These include tools for microsatellite primer development (e.g. MSATCOMMANDER (Faircloth, 2008)), identification and correction of genotyping errors (e.g. MICRO-CHECKER (Van Oosterhout *et al.*, 2004)), assessment of statistical resolution power for tests of genetic differentiation (POWSIM (Ryman & Palm, 2006)), DNA sequence alignment (e.g. BIOEDIT (Hall, 1999)), inference, validation and visualization of phylogenetic trees and networks (e.g. TREEFIT (Kalinowski, 2009) and SPLITSTREE4 (Huson & Bryant, 2006)) and a number of versatile software packages that provide a variety of options for calculating different measures of genetic differentiation from allelic distribution data and analysing population genetic structure (e.g. GENEPOP (Raymond & Rousset, 1995; Rousset, 2008), GenAlEX (Peakall & Smouse, 2006, 2012) and SPAGEDI (Hardy & Vekemans, 2002)).

1.3 - Fully Aquatic Reed Beetles - The Genus *Macroplea*

Figure 1. ***Macroplea appendiculata*** **(PANZER, 1794).**
Photo: Samuel Waldron

The reed beetles (Donaciinae KIRBY, 1837) are a subfamily of the leaf beetles (Chrysomelidae LATREILLE, 1802) consisting of approximately 165 species that are predominantly found in the northern hemisphere (Kölsch & Pedersen, 2008). Larvae of the Donaciinae invariably develop in mud underwater. Endosymbiotic bacteria provide a secretion used by the larvae for building a cocoon for pupation (Kölsch *et al.*, 2009; Kölsch & Pedersen, 2010). Following this key adaptation to aquatic habitats, members of the Donaciinae underwent adaptive radiation resulting in different degrees of adaptation to aquatic lifestyles. Adult reed beetles live and feed (mostly oligo- or monophagous) on grasses in wet marshes, semi-aquatic on floating leaf plants or completely under water on submerged macrophytes.

The majority of reed beetles are terrestrial as imagines, with exception of the tribe Haemoniini (CHEN, 1941). The tribe consists of two genera; while adults in the new-world genus *Neohaemonia* (SZÉKESSY, 1941) live amphibious, the adaptation to an aquatic life style is more strongly realized in the Palearctic genus *Macroplea* (SAMOUELLE, 1819). Uniquely among the many leaf beetle species, members of this genus are fully aquatic. All life stages, including the adult beetles, are found on submerged host plants in freshwater and brackish

habitats, including the Baltic Sea. And given their high salinity tolerance, this makes members of the genus *Macroplea* rare examples of truly marine insects (Kölsch *et al.*, 2010).

The genus *Macroplea* currently comprises six known species with varying extent of known distributions (see Table 1) and an apparent hot spot of diversity in eastern Asia, with all recognized species occurring in China (Lou *et al.*, 2011). The recently discovered species *M. huaxiensis* (LOU & LIANG, 2011) and *M. ranina* (Lou & Yu, 2011) are exclusively known from China and *M. japana* (JACOBY, 1885) is known to occur in China, Japan and East Siberia. Further three *Macroplea* species can, however, also be found in Western Europe.

M. pubipennis (REUTER, 1875) has long been considered to be endemic to Finland (until its recent discovery in China (Askevold, 1990; Kölsch *et al.*, 2006)). The extreme discontinuity in the known distribution of *M. pubipennis* might reflect that *Macroplea* species are probably often overlooked due to their rather elusive aquatic life style and specimens therefore rarely occur in collections (Medvedev, 2006). The two sister species *M. appendiculata* (Panzer, 1794) and *M. mutica* (FABRICIUS, 1792) show a wide distribution throughout great parts of the Palaearctic. *M. appendiculata* (likely derived from *M. mutica* approximately 2.5 Ma ago (Kölsch *et al.*, 2006)) shows a trans-palearctic distribution not only overlapping with that of *M. mutica*, the two species even occur syntopically. The ecological differentiation between these two species is still poorly understood although slight differences in adult host plant use (oviposition) have been documented (Kölsch & Kubiak, 2011). Early assumptions about an ecological differentiation of *M. appendiculata* and *M. mutica* based on differences in salinity preference or tolerance (Freude *et al.*, 1966; Mohr, 1985) could not be corroborated based on laboratory experiments and recorded salinities in realized habitats of both species (Kölsch *et al.*, 2010; Kölsch & Krause, 2011). It cannot be excluded, however, that differences in habitat salinity played a role in speciation within the genus *Macroplea* (Kölsch *et al.*, 2010).

Table 1. Recognized species in the genus *Macroplea* (SAMOUELLE, 1819; Chrysomelidae: Donaciinae)

Information on known host plants and distribution of six recognized *Macroplea* species, according to Lou *et al.* (2011) with additional information on host plant use according to Zhang *et al.* (2010)(*), Saari (2007)(**) and Cox (2007)(***). Listed host plant taxa are used by larvae, adults or both.

Species	Host plants	Distribution
M. appendiculata (PANZER, 1794)	*Ranunculus* (L.) - Ranunculaceae *Carex* (L.) - Cyperaceae *Potamogeton* (L.) - Potamogetonaceae *Myriophyllum* (L.) - Haloragaceae *Sparganium* (L.) - Sparganiaceae	China, Siberia, Middle Asia, Europe, Northern Africa
M. huaxiensis (LOU & LIANG, 2011)	*Vallisneria natans* (Lour.) - Hydrocharitaceae *Ottelia acuminate* (Gagnep.) - Hydrocharitaceae	China
M. japana (JACOBY, 1885)	*Potamogeton* (L.) - Potamogetonaceae(*) *Myriophyllum* (L.) - Haloragaceae (*) *Hydrilla verticillata* (L.f.) Royle - Hydrocharitaceae *Vallisneria spiralis* (L.) - Hydrocharitaceae *Ottelia acuminata* (Gagnep.) Dandy - Hydrocharitaceae *Nymphoides peltatum* (S.G.Gmel.)Kuntze - Menyanthaceae *Alopecurus aequalis* (Sobol.) – Poaceae	China, Japan, East Siberia.
M. mutica (FABRICIUS, 1792)	*Brasenia* (Schreb.) - Cabombaceae *Potamogeton* (L.) - Potamogetonaceae *Zannichellia palustris* (L.) (***) - Potamogetonaceae *Ruppia* (L.), *Zostera* (L.) - Zosteraceae *Sparganium* (L.) – Sparganiaceae	China, Japan, Siberia, Mongolia, Middle Asia, Europe
M. pubipennis (REUTER, 1875)	*Potamogeton* (L.) - Potamogetonaceae(**) *Myriophyllum* (L.) - (Haloragaceae) (**)	China, Finland
M. ranina (LOU & YU, 2011)	*Hippuris vulgaris* (L.) - Hippuridaceae	China

1.3.1 - Macroplea mutica (FABRICIUS 1792)

Figure 2. *Macroplea mutica* (FABRICIUS, 1792).
Photo: Christiane Bramer

The present study focuses on the species *Macroplea mutica (FABRICIUS 1792)*, a slender beetle with an average length of 4,7 mm (male) to 5,5 mm (female) and average width of 1.7 mm (males) to 2.3 mm (female) (Türkgülü *et al.*, 2011). *M. mutica* shows the widest known distribution of all *Macroplea* species (Kölsch & Kubiak, 2011), having been found in Belgium, Belarus, Denmark, Estonia, Finland, France, Great Britain, Germany, Hungary, Italy, Latvia, The Netherlands, Norway, Russia, Poland, Romania, Sweden, Algeria, Kyrgyzstan, Mongolia, Uzbekistan (Silfverberg, 2010), Turkey (Türkgülü *et al.*, 2011) and China (Mende *et al.*, 2010; Lou *et al.*, 2011).

1.3.1.1 - Life Cycle

As adult, *M. mutica* lives exclusively on submerged macrophytes like the pondweed *Potamogeton pectinatus* L. Larvae might in some cases also use submerged rhizomes of plants rising above the water surface (see Table 1 for known host plant taxa). On *P. pectinatus* the female beetle lays rows of three to 20 eggs protected between the stem and the leaf sheath

of the host plant (Laux & Kölsch, 2014), where, covered with a translucent rubber-like substance, they are 'glued' in place (see Figure 40a). In temperate Europe, oviposition occurs between April and September (G. Kölsch, personal communication). Larvae develop attached to the base or rhizome of the host plant in cocoons built of secreted material produced by endo-symbiotic bacteria (Kölsch *et al.*, 2009; Kölsch & Pedersen, 2010). The fully developed beetle overwinters in this cocoon to hatch in the following spring.

1.3.1.2 - Respiration and Mobility

The capacity for locomotion out of water seems to be generally low in the genus *Macroplea*. This largely results from physiological adaptations to respiratory needs under water in course of a relatively recently evolved aquatic life style. The species *M. mutica* shows an especially striking lack of potential for active dispersal over land, even when compared to other *Macroplea* species.

While *Macroplea* larvae breathe by penetrating the aerenchyme of the host plant with two hollow abdominal hooks (see Figure 40c) that are connected to the tracheal system (Kölsch & Kubiak, 2011), *Macroplea* imagines breathe by means of a physical gill (plastron). This thin, incompressible layer of air covers the ventral and lateral sides of thorax and abdomen as well the antennae and most parts of the head (Thorpe & Crisp, 1949). Physiological adaptations to this mode of respiration seem to have been developed at cost of mobility out of water. The genus *Macroplea* is descended from a relatively recent terrestrial ancestor and seems to show low metabolic rates and reduction of metabolically highly active tissues (i.e. flight and leg muscles) due to selective pressure to minimize oxygen consumption under water (Kölsch & Krause, 2011). *M. appendiculata* and *M. mutica* have shown very low metabolic rates and oxygen consumption, relative to other chrysomelid beetles and aquatic insect taxa (Kölsch & Krause, 2011). Although the extent of immobility seems to vary somewhat between different species (see below), ability to disperse over land is probably severely limited in the whole genus.

M. japana is likely to be the only *Macroplea* species to have retained (rudimentary) ability for flight. This is suggested by specimens caught in light traps (Lou *et al.*, 2011) and a single observation of an adult beetle that flew for a very short distance after it was taken out of water (Zhang *et al.*, 2010; Zhang, personal communication). *M. japana* might be the only *Macroplea* species retaining some of the ancestral ability to fly since it is by far the smallest member of the genus. Small size lowers energy cost of flight and is favorable for oxygen consumption and uptake (Kölsch & Krause, 2011). Therefore the need for flight muscle

reduction might be reduced relative to other *Macroplea* species. Phylogenetic reconstruction of the genus *Macroplea* (considering the species *M. appendiculata*, *M. japana*, *M. mutica* and *M. pubipennis*) furthermore showed *M. japana* as the most basal taxa. Remains of the ancestral ability for flight were probably lost entirely in larger, more derived *Macroplea* species. Reports of ability for flight in other *Macroplea* species seem to have been solely based on the discovery of existing hindwings. Mohr (1985) describes *M. mutica* as able to fly without providing further information or evidence for this claim besides the observation that the species is fully winged. Generally, the hindwing venation in *Macroplea* has been described as "most reduced known in Donaciinae" (Mann & Crowson, 1983) which is consistent with a reduced ability to fly. To the author's knowledge and own experience *M. mutica* has never been observed to actively lift its elytra or even fly. Mende *et al.* (2010) report that flight in *M. mutica* could not be induced in beetles taken out of water regardless of diverse experimental conditions with different temperatures and lighting. Reduced mobility out of water in *M. mutica* is not limited to a lost ability for flight. While *M. huaxiensis* has been reported to be able to slowly walk out of water for at least two hours (Lou *et al.*, 2011) the rather thin and weak legs of *M. mutica* hardly allow walking out of water (Mende *et al.*, 2010) and specimens furthermore seem to quickly die of desiccation within minutes after being taken out of water (own observation). Due to a lack of adaptations for active swimming, locomotion in *M. mutica* is therefore restricted to walking (mostly slowly) over substrate and vegetation under water.

1.3.1.3 - Population Genetic Structure of M. mutica

Population genetic structure and postglacial colonization history of *M. mutica* in Europe have recently been subject to a study by Mende *et al.* (2010). An AFLP analysis revealed pronounced population differentiation, signs of inbreeding and a population genetic signature of passive dispersal, as hypothesized based on the low mobility of this species. A comparatively higher genetic admixture among the Baltic Sea populations compared to inland populations suggested different relative influences of hydrochorous and zoochorous passive dispersal in inland- and Baltic Sea habitats (Mende *et al.*, 2010). Furthermore, *M. mutica* populations from the eastern part of Northern Germany appeared genetically similar to the and samples from the Baltic Sea and south-eastern Europe, while samples from the western part of Northern Germany appeared genetically close to British populations. Mende *et al.* (2010) therefore proposed that postglacial colonization of Europe might have originated from

two separate glacial refugia in south-eastern Europe and the area of present-day southern England or Ireland, resulting in a suture zone in Northern Germany.

1.3.1.4 - Passive Dispersal of M. mutica: Water and Waterbirds as Potential Vectors

The strikingly low potential for active dispersal in *M. mutica* strongly contrasts with the wide Palearctic distribution of this species. Mechanisms of passive transport must therefore be of great significance to its dispersal. This makes *M. mutica* an interesting model for the study of the passive dispersal of aquatic insects. Two mechanisms of passive transport are likely to be of major importance for dispersal of *M. mutica*: Transport with flowing water (with floating parts of host plant) has been documented but is naturally largely limited to dispersal within water bodies or catchments. The transport by waterbirds is potentially very effective even between isolated wetlands but has, however, never been documented for aquatic beetles.

M. mutica can, however, indisputably be transported with flowing water. This is demonstrated by the observation that live beetles inside cocoons attached to parts of the host plant are occasionally found in beach drift lines, washed up after severe weather (Mende *et al.*, 2010). But given the clear limitations of water flow-mediated transport for dispersal between hydrologically isolated wetlands, zoochorous transport might explain the wide distribution of this species in absence of the ability to actively move across dry land.

Since the life cycle of *M. mutica* is tightly associated with host plants that are food to many waterbird species (Figuerola & Green, 2002; Allin & Husband, 2003) waterbirds can be expected to frequently pick up considerable quantities of different *M. mutica* life stages with foraged plant material and potentially transport them between suitable habitats after ingestion. Cocoons or eggs of *M. mutica* are life stages that appear suitable to survive internal transport by waterbirds, due to features potentially resulting in resistance against mechanical and chemical stress during gut passage. The cocoons are made of a rigid and durable material that is known to withstand strong chemicals (Böving, 1910) and *M. mutica* eggs might gain some protection from an envelope of translucent rubber-like substance and ovipositioning between the stem and the leaf sheath of the host plant. Should at least one life stage be capable to survive the passage through the digestive tract of waterbirds, this could make *M. mutica* a rare example of (internal) zoochorous dispersal of an aquatic insect.

1.3.1.5 - The Mute Swan Cygnus olor - a Potential Vector Species

Figure 3. Mute swan - *Cygnus olor* (GMELIN, 1783).
Photo: Lisa Laux

Several taxonomic groups of waterbirds have been connected to the (internal) dispersal of aquatic invertebrates, mainly the Anatidae (especially the dabbling ducks (Anatinae) and diving ducks (Aythyinae)) and Rallidae, with the majority of experimental trials and field collections focused on dabbling ducks (Van Leeuwen *et al.*, 2012b). A meta-analysis of the suitability of these groups for the internal dispersal of plant and animal propagules did not reveal differences in their quantitative dispersal capacity (Van Leeuwen *et al.*, 2012b). Suitability as dispersal vector is, however, likely to vary with differences in diet and digestive physiology (Figuerola *et al.*, 2003).

A number of different bird species is likely to frequently ingest life stages of *M. mutica* with foraged plant material due to their diet and thereby represent potential dispersal vectors for *M. mutica*. But since the efficiency of internal dispersal of aquatic organisms increases with bird body mass (Van Leeuwen *et al.*, 2012b) the large and heavy mute swan (*Cygnus olor* (GMELIN, 1783)) might be an especially suitable vector species. Mute swans qualify as a potential vector species for the dispersal of *M. mutica* for further reasons: They show an extent of Palearctic distribution that is similar to that of *M. mutica*; between 40° and 60° N from Western Europe to Northeast China (Atkinson *et al.*, 2006). And while host plants of

M. mutica (e.g. *Potamogeton* sp.*, Zannichellia palustris, Ruppia* sp.) are food to a number of different waterbird species (Figuerola & Green, 2002; Allin & Husband, 2003), grazing by mute swans has been shown to significantly impact biomass in these plants (Allin & Husband, 2003; Stafford *et al.*, 2012). Moreover, swans have been connected to (long distance-) dispersal and influencing population genetic structure in these species (e.g. Mader *et al.*, 1998). Furthermore, mute swan foraging shows impact on above-ground *and* below-ground parts of submerged aquatic vegetation (Stafford *et al.*, 2012) and swans are therefore bound to come into contact with (and ingest) all life-stages of *M. mutica* (eggs, larvae, cocoons and adult beetles), further enhancing chances of swan-mediated dispersal. Other species that might be potentially important as vector species due to their diet and distribution include, but are probably not limited to, the Eurasian coot (*Fulica atra,* (L.), Rallidae) and dabbling ducks like the Eurasian wigeon (*Anas penelope* (L.)) or the mallard (*Anas platyrhynchos* (L.)).

1.4 - Objectives

A first study of the population genetic structure in *M. mutica* (Mende *et al.*, 2010) showed a suture zone in Northern Germany that might result from (re-) colonization of Europe from two separate glacial refugia. But it was revealed that closer study of Northern Germany, Western Denmark (Jutland) and the Netherlands was needed to better characterize the putative suture zone and verify the detected close genetic relatedness between western German and British *M. mutica* populations (Mende *et al.*, 2010). Furthermore, *Macroplea mutica* shows a fascinating contrast of a vast distribution range and strikingly low potential for active dispersal. And while the transport of *M. mutica* by waterbirds has been proposed as a possible explanation (Mende *et al.*, 2010), there is no evidence for this mode of dispersal in aquatic beetles to date.

The present study accordingly has three main objectives. First, microsatellite markers are developed for the species *Macroplea mutica* and its population genetic structure is analysed based on microsatellites and mitochondrial DNA, with respect to signatures of passive dispersal mechanisms, colonization of Northern Europe and landscape features acting as potential barriers to gene flow. Second, a possible role of zoochorous transport for the dispersal of *Macroplea mutica* is investigated by analysing abundances and movements of a potential avian dispersal vector (the mute swan *Cygnus olor*) for potential correlations with population genetic differentiation across *M. mutica* populations. Third, *M. mutica* eggs and cocoons are tested for the potential for surviving digestion in waterbirds, following the hypothesis that certain life stages of *M. mutica* might facilitate waterbird-mediated dispersal by surviving internal transport.

2 - Material and Methods

2.1 - Origin of Samples

Animal samples originated from 24 European *M. mutica* populations, two Chinese *M. mutica* populations and two European *M. appendiculata* populations (out-group) (see Table 2). European samples were collected between 2001 and 2012; the Chinese populations were sampled in 1994 and 1995, Chinese specimens were pooled for analysis. Exact location of beetle field sampling is not given due to conservational concerns. Specimens were collected as larvae or adult beetles and stored in 100 % ethanol and/ or frozen at -20° C. Median sample size per site was 15. Species had been determined beforehand for most populations and individuals as they had been subject to AFLP analyses by Mende et al. (2010). Analyses of mtDNA fragments allowed species identification for the remaining populations.

Table 2. Origin of Samples

The abbreviation, sample size (n), location, habitat type and collectors of analyzed sampling sites.

Abbreviation	n	Origin	Habitat type	Collectors
Macroplea mutica				
WAL	15	Wales, United Kingdom	coastal lake	G. Kölsch, E. Meichssner
YOR	15	East Yorkshire, United Kingdom	coastal lake	J. Laux
ESU	15	East Sussex, United Kingdom	coastal lake	G. Kölsch, E. Meichssner
CAM	15	Cambridgeshire, United Kingdom	inland lake	G. Kölsch, E. Meichssner
HOL	8	North Holland, Netherlands	coastal lake	J. Laux, G. van Ee, A. Bouman
NFL	13	North Frisia, Germany	coastal lake	G. Kölsch, R. Suikat
NFS	15	North Frisia South, Germany	coastal lake	G. Kölsch
PLW	15	Plön, Germany	coastal lake	G. Kölsch, R. Suikat
JUN	15	Northern Jutland, Denmark	coastal lake	J. Laux
JUW	15	Western Jutland, Denmark	coastal lake	J. Laux
HEL	11	Little Belt, Denmark	marine	G. Kölsch
LEM	15	Fehmarn East, Germany	marine	G. Kölsch, R. Suikat
ORT	15	Fehmarn West, Germany	marine	J. Laux, G. Kölsch
OBJ	15	Sjaelland, Denmark	marine	G. Kölsch, R. Suikat
RUG	15	Rügen, Germany	marine	G. Kölsch, E. Meichssner
OST	15	Västerbotten, Sweden	marine	A. Nilsson
VAX	6	Stockholm North, Sweden	marine	H.-E. Wanntorp
UTO	5	Stockholm South, Sweden	marine	H.-E. Wanntorp
VOR	8	Vormsi, Estonia	marine	O. Biström
DRA	5	West. Finland, Finland	marine	O. Biström
KIR	5	Kirkkonummi, Finland	marine	O. Biström
BRA	10	Lazio, Italy	inland lake	G. Kölsch, E. Meichssner
SAR	10	Sardinia, Italy	coastal lake	G. Kölsch, E. Meichssner
BAL	4	Transdanubia, Hungary	inland lake	I. Musko
CHI	8	Daqing, PR of China	inland lake	G.A. Buckingham, C. Zhiqun, C.A. Bennett, D. Jianqing
SEE	8	Plön, Germany	inland lake	G. Kölsch
Macroplea appendiculata (Outgroup)				
MAS	1	Plön, Germany	inland lake	G. Kölsch
MAN	4	North Holland, Netherlands	inland lake	J. Laux, G. van Ee, A. Bouman

2.2 - Molecular Biology

2.2.1 - DNA Extraction

DNA was isolated from two to three legs of adult beetles or slices of larval tissue, using a micropestle to grind the sample in a microcentrifuge tube while cooling the tip of the tube in liquid nitrogen. Extraction was performed using a DNeasy blood and tissue kit (Qiagen, Hilden, Germany) and checked for quality and quantity by using a Nanodrop 2000 UV-Vis Spectrophotometer (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA).

2.2.2 - Development and Characterization of Microsatellite Markers for Macroplea mutica

2.2.2.1 - Construction of a Genomic DNA Library Enriched for Microsatellites

Genomic DNA for construction of the library was isolated from three legs of a beetle caught at the Northern German Baltic coast (for the extraction protocol see 2.2.1). Isolated genomic DNA was dried and shipped to the USA where a microsatellite library (library of genomic fragments enriched for short tandem repeats) was constructed as a commercial service by Steve Bogdanovicz at the Evolutionary Genetics Core Facility (*EGCF*) at Cornell University (Ithaca, NY, USA).

The following methodology for library construction is a summary based on an EGCF quote for microsatellite library development:

Genomic DNA was digested with a restriction enzyme (five-base cutter), generating blunt-ended fragments. Linkers were ligated to the digested DNA and the resulting fragments were enriched for microsatellites by hybridization to biotinylated repeat probes. Repeat probes represented two unique dimer motifs, five unique trimer motifs, and four unique tetramer motifs. After magnetic capture of enrichment probes, the enriched genomic fragments were amplified by PCR, ligated to Roche/454 Titanium Multiplex Identifier adapters and size fractioned in an agarose gel. A library of sequences was then generated with Roche/454 GS FLX Titanium reagents, protocols and hardware (Roche Applied Science, Basel, Switzerland). After trimming of adapter sequences and assembly 454 reads were then made available as fasta format files.

2.2.2.2 - Microsatellite Primer Design

Microsatellite primers were developed using PRIMER3 (Rozen & Skaletsky, 2000) as implemented in MSATCOMMANDER (Faircloth, 2008). The software detects user-specified classes (di-, tri-, tetra-nucleotide, etc.) of microsatellite arrays in fasta-formatted sequence files. Minimal repeat numbers considered were six (di-nucleotids) and five (tri- and tetra-nucleotids). Given that repeats were located within the source file, MSATCOMMANDER suggested sequences for forward- and reverse primers within a user specified distance to the detected repeat (default value: 50 bp) and following desired primer properties. The default primer values were used. These included a product size range from 150-450 bp, a primer melting temperature T_m between 57°C and 62°C (and an optimal T_m of 60°C), a maximum difference in T_m between primers of 5°C, an optimal primer size of 19 bp and a GC content of 35-75%.

2.2.2.3 - Microsatellite Primer Testing and Locus Screening

Optimal annealing temperatures for primer pairs were determined on a gradient from 45°C to 65°C. To screen for polymorphic loci, polymerase chain reactions (PCRs) were performed using extracted *M. mutica* DNA from six different European populations. Products were separated on a 2.5 % agarose gel (Ultra-Pure Agarose, Life Technologies GmbH, Darmstadt, Germany). Forward primers for reliably amplifying loci that showed size differences between tested individuals (suggesting a polymorphic locus) were redesigned with a fluorescent label on the 5'-end (MWG Eurofins, Ebersberg, Germany).

After markers had been screened in simplex and annealing temperatures and expected PCR-product size were determined, primer pairs showing compatible allele sizes and similar annealing temperatures were amplified in multiplex sets of three primers each. Annealing temperatures and the relative amounts of primers for each marker were adjusted until all loci were reliably amplified.

Loci showing problematic "triallelic" patterns (chromatogram features that interfered with unambiguous allele calling) after separation by fluorescent capillary electrophoresis were omitted when a redesign of primers did not improve the signal.

2.2.2.4 - PCR Conditions

Table 3. Composition of Primer-Mixture for Multiplex PCR

Set	Primer	Dye-label	Proportion	Sequence
1	16359 f	DY-682	0.2	TTTGGCGGGATTGCACTTG
1	16359 r	-	0.2	AGAGGTTCTATCAAACTGTACCAC
1	4107f	DY-682	0.1	TGTTGTCTGACGTAACTCTGC
1	4107r	-	0.1	GAGTCTAACAAGACCATCTGTCG
1	12208f	CY5	0.2	GTGAGACGTGAAACGGCAG
1	12208r	-	0.2	AGGGTTCGTAGTCGGTATGC
2	321f	DY-682	0.083	CCTTCGTAGGAACTTTAGGCG
2	321r	-	0.083	GACGAGGCGTGCTGTTTAG
2	3012f	DY-682	0.245	ATCCAGCTAACCAGATGGC
2	3012r	-	0.245	GTTGGGTTCAGCGCGTATC
2	1624f	CY5	0.165	TAAGGGTCGAATGGGCAGG
2	1624r	-	0.165	GTGCAGCATCTGGTTCACG

Set: multiplex set; f: forward primer; r: reverse primer; Proportion: proportion of primer mixture.

PCR's and genotyping analyses were performed as two multiplex sets with three primer pairs each (see Table 3). Differently colored forward primers were used to distinguish between products with overlapping allele sizes. PCR's were performed in volumes of 12.5 µl with a *Eppendorf Mastercycler Gradient* (Eppendorf AG, Hamburg), containing the following components: 1 to 8 µl of template DNA and PCR water (equaling ≥ 40 ng DNA), 1 µl of 5 µM primer-mix with Cy5- or DY-682 dye (Eurofins MWG Operon, Ebersberg, Germany) labeled forward primers, 1.25 µl PCR Rxn buffer (10x), 0.1 µl Taq polymerase (Taq DNA polymerase 5 u/µl), 0.5 µl 50 mM $MgCl_2$ (Life Technologies GmbH, Darmstadt, Germany) and 1.25 µl dNTP's (2.5mM each) (Roth, Karlsruhe, Germany). Cycling conditions were a hot start at 93°C for 3 minutes, followed by 35 cycles of denaturation at 93°C for 1 minute, annealing at 60°C (set 1) / 51°C (set 2) for 1 minute and elongation at 72°C for 30 seconds. After a final elongation step for 5 minutes at 72°C, samples were stored at 4°C until electrophoresis.

2.2.2.5 - Genotyping / Scoring

PCR products were analyzed by fluorescent capillary electrophoresis (Beckman CEQ8800 sequencer). Alleles were sized by comparison to a DNA Size Standard Kit - 600 (Beckman-Coulter, Galway, Ireland). To detect and size DNA fragments, chromatogram files were individually inspected and alleles were identified and scored manually.

In case of ambiguity in allele calling or non-distinctive peaks in the chromatogram, single samples were amplified in simplex, using only one primer pair instead of the primer mix.

2.2.3 - Amplification and Sequencing of Mitochondrial DNA

A 600 bp long section of the mitochondrial genome, containing the partial *cytochrome oxidase I* gene was amplified for 186 individuals of *Macroplea mutica* from 21 sampling sites and 5 individuals of *Macroplea appendiculata* from two sampling sites (as outgroup) applying standard PCR techniques using an Eppendorf Mastercycler gradient cycler (Eppendorf AG, Hamburg) and primers S2183 ("Jerry"; Simon *et al.*, 1994) and A3022 (Dobler et al. unpubl.). Thermal cycle amplifications were performed in 25 µL reactions, containing 2 µL dNTPs (2.5 mM each) (Roth, Karlsruhe, Germany), 1 µL 50 mM $MgCl_2$, 0.2 µL *Taq* polymerase, 0.5 µL (10 µM) each of primers S2183 and A3022 (see Table 4), 18.5 µL PCR water and template DNA (equaling ≥ 80 ng DNA). Cycling conditions were a hot start at 94°C for 5 minutes, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 54°C for 1 minute and elongation at 72°C for 1 minute. After a final elongation step for 5 minutes at 72°C, samples were stored at 4°C. PCR products were directly sequenced in both directions using the PCR primers. Sequencing was done by a commercial service (GATC Biotech, Konstanz).

Table 4. Primer Pair for Amplification of a CO I Fragment

Primer	Source	Sequence
S2183 ("Jerry")	Simon *et al.* (1994)	CAACATTTATTTTGATTTTTTGG
A3022	Dobler *et al*. (unpubl.)	GGGRTTTAAATCCAAYGCACTAATCTG

2.3 - Population Genetic Analyses

2.3.1 - Characterization of Developed Microsatellite Markers

2.3.1.1 - Tests for Null Alleles, Stuttering, Hardy-Weinberg Equilibrium and Linkage Disequilibrium

Prior to all further analyses, microsatellite data was checked for deviations from Hardy-Weinberg equilibrium, for linkage disequilibrium (with GENEPOP 4.2 (Raymond & Rousset, 1995; Rousset, 2008), for the presence of null alleles and scoring errors due to stuttering and for large allele dropout (MICRO-CHECKER 2.2.3 (Van Oosterhout *et al.*, 2004)).

Deviations from Hardy-Weinberg equilibrium were tested as a probability test for all populations and loci, deviations expressed as two estimates of F_{IS}, Weir & Cockerham's estimate (Weir & Cockerham, 1984) and Robertson & Hill's estimate (Robertson & Hill, 1984) representing deficit (positive values) or excess (negative values) of heterozygotes. The Markov chain method with 5,000 batches of 10,000 iterations each was used to estimate if deviations from the expected ratio were significant (Guo & Thompson, 1992). P-values were adjusted for multiple testing with a sequential Bonferroni correction (Holm, 1979).

As a test of the composite pairwise linkage disequilibrium, the data was tested for any possible non-random association of alleles at any pair of loci. P-values were again adjusted for multiple testing with a sequential Bonferroni correction.

2.3.1.2 - Validation of Markers - Test for Congruence with AFLP Data

Results of the microsatellite analysis were tested for congruence with results of an AFLP analysis by Mende *et al.* (2010). Pairwise F_{ST} data for 16 populations that had been subject to both studies (Wales ("WAL"), Cambridgeshire ("CAM"), Sussex ("ESU"), North Frisia ("NFL"), Plön ("PLW"), Fehmarn ("ORT" and "LEM"), the Stockholm archipelago ("UTO" and "VAX"), the Little Belt ("HEL"), Sjaelland ("OBJ"), Rügen ("RUG"), Estonia ("VOR"), Västerbotten ("OST") and Finland ("KIR" and "DRA")) - were compared with a Mantel test with 10,000 permutations and a Spearman's rank correlation test.

2.3.2 - Population Genetic Differentiation and Isolation by Distance

2.3.2.1 - Assessing Statistical Power of the Used Set of Microsatellites

The statistical power of the used set of microsatellites to detect genetic differentiation was tested with POWSIM (Ryman & Palm, 2006). For the present set of samples, number of loci and allele frequencies, POWSIM assesses the probability of rejecting the null hypothesis of genetic homogeneity when it is true (α-error) and accepting it when it is not true (β-error).

Sampling is simulated from a specified number of populations with different predefined degrees of hypothetic genetic differentiation (measured as F_{ST}). Resolution power is tested for different expected values for F_{ST} based on the effective population size (N_e) and the number of generations of drift (t) and Nei's (1987) definition of $F_{ST} = 1-(1 - 1/2\ N_e)^t$. Power was estimated as the proportion of significant outcomes of Fisher's exact tests and χ^2 tests when repeating the simulations 10,000 times for each of 10 different levels of F_{ST} (from 0 to 0.0129). The α-error was calculated as the probability to (falsely) find significant

differentiation with F_{ST} = 0 and t = 0. With this setting samples are drawn directly from the base population. The simulated drift process and the used effective generation sizes do not necessarily reflect assumptions about the real demographics or evolutionary history of the studied populations but are merely used to simulate different F_{ST}-levels in a biologically reasonable fashion.

2.3.2.2 - Population Genetic Differentiation

Testing for significant genetic differentiation between pairs of populations was performed by calculating genotypic differentiation with the software GENEPOP 4.2 (Raymond & Rousset, 1995; Rousset, 2008). The test was performed as an exact G-test (Goudet et al. 1996), the null hypothesis being: "genotypes are drawn from the same distribution for all populations".

Population genetic differentiation was further calculated as fixation indices based on allele identity (F_{ST}) and allele sizes (R_{ST}) with GENEPOP. The program calculates unbiased estimators for F_{ST} and Rho_{ST} (ρ_{ST}, an unbiased estimate of Slatkin's R_{ST} (Slatkin, 1995; Rousset, 1996)), based on a weighted analysis of variance (Cockerham, 1973; Weir & Cockerham, 1984; Michalakis & Excoffier, 1996). Computations of multilocus estimates were performed following Weir and Cockerham (1984).

Estimates of fixation indices based on allele identity (F_{ST}) and allele sizes (R_{ST}) were calculated for all 210 pairs of 21 sampling sites / (sub-) populations.

Furthermore, a matrix of Nei's chord distance D_A (Nei *et al.*, 1983) for all 210 pairs of (sub-) populations was calculated with TREEFIT (Kalinowski, 2009).

2.3.2.3 - Testing the Relative Performance of F_{ST} and R_{ST} Estimates

To test whether F_{ST} or R_{ST} statistics better reflected differentiation at the used microsatellite loci, the performance of F_{ST} and R_{ST} estimates was compared for the present data.

Since the relative performance of allele size-based statistics (R_{ST}) versus allele identity-based statistics (F_{ST}) depends on the relative contributions of mutations following a stepwise mutation model (SMM) versus the effects of drift and migration to population differentiation, mean square errors of pairwise F_{ST} and R_{ST} were compared for the present microsatellite data by randomly permuting allele sizes for all 6 loci, 256 samples and 210 population pairs along allelic states (while maintaining allele-identity) with 10,000 permutations using SPAGEDI 1.4 (Hardy & Vekemans, 2002). Subsequently, global and pairwise R_{ST} values before and after randomization were compared. According to Hardy *et al.*(2003), significant contribution of SMM-like mutation to genetic differentiation should cause computed R_{ST} values after

randomization (called pR_{ST} by the authors) to be significantly smaller than R_{ST} values observed before the randomization (while F_{ST} estimates remain identical). A significant test result would imply a significant contribution of SMM-like mutations to the observed genetic differentiation and sufficiently large mutation rates relative to the effects of migration and drift. This would therefore suggest that R_{ST} might outperform F_{ST}-estimates for the present data (Balloux & Lugon-Moulin, 2002; Hardy *et al.*, 2003).

2.3.2.4 - Subdivision of Macroplea mutica Sampling Sites

In addition to analysis of the whole microsatellite dataset for all 21 (Northern) European *M. mutica* sampling sites, the data was subdivided according to either habitat type or membership to main genetic clusters.

Two groups represented different types of habitat. The *inland group* consists of sampling sites in inland- and coastal lakes, without direct hydrographic connection to the sea. The *Baltic Sea group* includes all sampling sites situated within the Baltic Sea. Independently of habitat type the data set was alternatively divided to represent the two main population genetic clusters (western- and eastern cluster) detected for *M. mutica* as the uppermost level of population structure using the ΔK method by Evanno *et al.* (2005) (see 2.3.3.1 and 3.1.2.6).

The division by habitat type (inland- and Baltic Sea group) was included in analyses since different possible dispersal mechanisms in inland water bodies and the Baltic Sea could potentially lead to differences in gene flow and genetic structure. Separate analysis of the two main genetic clusters was performed to avoid potential bias whenever spatial structure of genetic differentiation had to be considered. The assignment of all 21 sampling sites in the data set to the respective groups is shown in Table 5. Habitat type groups and main genetic clusters are very similar in composition and only differ in the assignment of the two sampling sites in Jutland (“JUW” and “JUN”). Samples from Jutland represent the only inland populations not assigned to the western cluster (and the only sites within the eastern cluster not representing samples from the Baltic Sea).

Table 5. Habitat Type and Main Genetic Cluster Membership of Sampled *M. mutica* Populations

Population (Abbreviation)	Habitat type	Genetic cluster
Wales (WAL)	Inland	West
Yorkshire (YOR)	Inland	West
Cambridgeshire (CAM)	Inland	West
Sussex (ESU)	Inland	West
North Holland (HOL)	Inland	West
North Frisia (NFL)	Inland	West
North Frisia South (NFS)	Inland	West
Plön Lakes (PLW)	Inland	West
Northern Jutland (JUN)	Inland	East
Western Jutland (JUW)	Inland	East
Little Belt (HEL)	Baltic Sea	East
Fehmarn East (LEM)	Baltic Sea	East
Fehmarn West (ORT)	Baltic Sea	East
Sjaeland (OBJ)	Baltic Sea	East
Rügen (RUG)	Baltic Sea	East
Norther Sweden (OST)	Baltic Sea	East
Stockholm North (VAX)	Baltic Sea	East
Stockholm South (UTO)	Baltic Sea	East
Estonia (VOR)	Baltic Sea	East
Finland West (DRA)	Baltic Sea	East
Finland East (KIR)	Baltic Sea	East

2.3.2.5 - Testing for Isolation by Distance

The population genetic data was tested for *isolation by distance*, the (positive) correlation of geographic and genetic distance among populations, with ISOLDE as included in GENEPOP. GENEPOP performs Mantel tests (Mantel, 1967) based on data in two semi-matrices. Permutations of lines or columns of the (semi-) matrix provide the distribution of a statistic under the null hypothesis of independence between two variables (e.g. genetic and geographic distance). Instead of Mantel approximations of "Z" the program uses a rank correlation coefficient. The program provides results as two one-sided tests (for either positive or negative correlation). Semi-matrices of pairwise R_{ST} values and linear geographic distances were tested for significant correlation. At first the default value for the minimal distance considered was used (0.0001). Accordingly, all pairs of (sub-) populations (down to spatial distances of < 3 km) were considered. The tests were repeated considering only data for (sub-) population pairs showing geographic distances of at least 50 km, to test whether non-linearity of IBD at small spatial scales might lead to a biased estimate of IBD. Geographic distances between sampling sites were inferred using QGIS 2.0.1 (Quantum GIS Development Team, 2013) (see 2.4.4). The Mantel tests were run with 10,000 permutations. The whole dataset of

210 population pairs and four data subsets (for population pairs from the western- and eastern cluster, Baltic Sea populations and inland populations (see 2.3.2.4)) were tested. P-values were adjusted for multiple testing with a sequential Bonferroni correction (Holm, 1979).

2.3.2.6 - Calculating Residual R_{ST}-Values

To avoid bias towards smaller p-values caused by potentially overlapping patterns of spatial auto-correlation when testing groups showing spatial structure of genetic distances, residual R_{ST}-values were calculated for linear regressions of pairwise R_{ST} on geographic sampling distance. Provided that Mantel tests suggested a significant (linear) correlation between pairwise R_{ST} values and geographic distances, pairwise R_{ST} was calculated (separately for each tested group) to rise with geographical distance as $R_{ST} = x*d+y$ (*d* being the geographical distance between populations in km). An "expected" R_{ST} value for every measured geographic distance between populations was calculated and the deviation from the "expected" R_{ST} value (the "residual" R_{ST} value) was inferred for every population pair by subtracting the "expected" R_{ST} value from the measured R_{ST} value.

2.3.2.7 - Construction of a Neighbor-Joining Tree

A neighbor-joining tree (Saitou & Nei, 1987) was constructed using TREEFIT (Kalinowski, 2009) based on Nei's modified chord distance D_A (Nei *et al.*, 1983). Statistical support for interior branches was calculated in 10,000 bootstrap runs and the fit of observed genetic distances and genetic distances in the tree was calculated as R^2. TREEVIEW (Page, 1996) and FIGTREE (http://tree.bio.ed.ac.uk/software/figtree/) were used for tree layout and display.

2.3.2.8 - Principal Coordinate Analysis

A principal coordinate analysis (PCoA) was performed with GenAlEX 6.5 (Peakall & Smouse, 2006, 2012) via a matrix of standardized Nei's D_A distances (Nei *et al.*, 1983) generated with TREEFIT (Kalinowski, 2009).

2.3.3 - Inferring the Number of Genetically Distinct Clusters

Different methods for Bayesian inference of population genetic structure as implemented in different software were used to characterize population genetic structuring among Northern European *M. mutica* (sub-) populations. Available software products provided a choice of

distinct models for the inference of population structure. Unless mentioned, the default options of each program were used.

Bayesian inference of the number of genetically distinct clusters within the data (on different levels of population genetic structuring) and allocation of studied populations to the inferred clusters was conducted using BAPS 5.2 (Corander *et al.*, 2003), GENELAND (Guillot *et al.*, 2005b) and STRUCTURE 2.3.4 (Pritchard *et al.*, 2000). Key differences between the used programs and used settings encompass the consideration of geographic coordinate data for sampling sites, the individual- or group-wise assignment to genetic clusters, the (optional) consideration of admixture and the use of Markov Chain Monte Carlo (MCMC) methods.

2.3.3.1 - Inferring the Uppermost Level of Population Structure

The number of genetic clusters K or "real number of populations" was inferred using STRUCTURE 2.3.4 (Pritchard et al., 2000) and an ad hoc statistic devised by Evanno et al. (2005): Structure calculates estimates for the posterior probability of the data for a given K, Pr(X|K) (Pritchard et al., 2000). The maximal average of the log likelihood of the data for each step of the MCMC (with half of the variance subtracted to the mean) is given as 'log probability' Ln P(D) or (from now on referred to as) 'L(K)'.

Instead of the distribution of L(K) the "ad hoc quantity based on the second rate order change of the likelihood function with respect to K" (Evanno et al., 2005) (ΔK) was used to detect the uppermost hierarchical level of structure.

Parameters were set as recommended by Evanno *et al.* (2005). The admixture model and the option of correlated allele frequencies between populations were chosen. The degree of admixture alpha was inferred from the data, Lambda, the parameter of the distribution of allelic frequencies, was set to one, burn-in and MCMC to 10,000 each. Each run was carried out 20 times in order to quantify the amount of variation of the likelihood for each K. The range of possible Ks tested was from 1 to 25 (the number of sampled populations plus 4).

First, the mean likelihood L(K) over 20 runs for each K was calculated. Secondly, the mean difference between successive likelihood values of K, $L'(K) = L(K) - L(K - 1)$ was calculated. Thirdly, the (absolute value of the) difference between successive values of $L'(K)$, $|L''(K)| = |L'(K + 1) - L'(K)|$ was calculated (the second order rate of change of L(K) with respect to K). Finally, ΔK as the mean of the absolute values of $L''(K)$ averaged over 20 runs divided by the standard deviation of L(K), $\Delta K = m(|L''(K)|)/s[L(K)]$ was assessed, which expands to $\Delta K = m(|L(K + 1) - 2\,L(K) + L(K - 1)|)/s[L(K)]$.

2.3.3.2 - BAPS 5.2

The software BAPS performs Bayesian analyses of population genetic structure, treating either both the allele frequencies and the number of genetically divergent groups or, alternatively, only the allele frequencies as random variables. BAPS uses a Bayesian approach to estimate allele frequencies and analyses population genetic structure with a stochastic optimization algorithm (instead of the more common Markov Chain Monte Carlo (MCMC) methods).

To estimate the number of clusters K the samples are most likely sub-divided into according to the data, the estimation process was repeated several times with different upper limits for K. The analyses were run with the "mixture of groups of individuals" option (Corander *et al.*, 2006, 2008; Cheng *et al.*, 2011) and repeated ten times for each upper limit for k (k=1, k=2, k=3, k= 4, k=5, k= 10, k=20 and k=50). BAPS subsequently assigns groups of individuals to the inferred clusters.

To reduce possible bias caused by overestimation of genetic structure due to weak stochastic fluctuations in allele frequencies, this analysis was repeated considering geographic coordinate data for predefined groups. For this option the program implements a spatial prior for assessing the population genetic structure (Corander *et al.*, 2007; Cheng *et al.*, 2013).

Additionally, the analysis of mixture of groups of individuals was run with predefined numbers of divergent groups (a fixed number of clusters) of K=2 and K=3 to infer which populations would be grouped under the assumption of partition sizes of K=2 and K=3.

2.3.3.3 - GENELAND

Similarly to the program BAPS, GENELAND (Guillot *et al.*, 2005a, 2005b, 2008; Guillot, 2008; Guillot & Santos, 2010; Guedj & Guillot, 2011) as package for R (R Core Team, 2013) uses Bayesian inference of probability to determine the optimal number of virtual populations (or clusters) and subsequently assigns individuals or groups of individuals to these clusters. In contrast to BAPS, GENELAND uses Markov Chain Monte Carlo (MCMC) methods for Bayesian inference and explicitly bases inference of cluster membership on geo-referenced genetic data.

Calculations were performed with different upper limits for the number of virtual populations k = 2, k = 3 and k = 25, with an uncorrelated allele frequencies model, with 5 independent runs of 100,000 MCMC iterations each (thinning: 100).

For post processing, information was extracted from the MCMC simulation with a horizontal discretization of 200 pixels and a vertical discretization of 100 pixels.

2.3.4 - Bayesian Inference of Migration Rates with MIGRATE-N

Mutation-scaled effective population sizes and migration rates between genetic clusters were calculated with MIGRATE-N 3.6.4 (Beerli & Felsenstein, 1999, 2001). This software simultaneously estimates effective population sizes and migration rates between populations.
MIGRATE-N estimates the mutation-scaled effective population size Θ (Theta) (defined as $4N_e\mu$ for the present diploid system with nuclear microsatellite loci, where N_e is the effective population size and μ is the mutation rate per generation and loci), M (defined as m/μ, where m is the immigration rate), calculates the number of effective migrants per generation as N_em = Θ M/4 and therefore the amount and direction of gene flow between populations based on population genetic data.
For each analysis 30,000,000 genealogies were sampled combined over 3 replicates. 50,000 steps were recorded per chain. Further settings for search parameters included Bayesian inference as search strategy and one long chain with a sampling increment of 100. The continuous Brownian motion model was chosen instead of a discrete stepwise mutation model to allow for faster parameter estimation. Runs were first optimised until they were found to converge and deliver acceptable posterior distributions. Migration rates were subsequently estimated for all directions and between all clusters. The uniform priors for Θ and M were initially set at a minimum of 0, a maximum of 100 and a delta value of 10. According to the observed posterior distributions after test runs uniform priors were later set to a minimum of 0, a maximum of 5 and a delta value of 0.5 for Θ and to a minimum of 0, a maximum of 300 and a delta value of 30 for M.

2.3.5 - Analysis of Mitochondrial DNA

Sequence Alignment and Haplotype Network Construction
Forward and reverse sequences and resulting consensus sequences for 186 individuals of *Macroplea mutica* from 21 sampling sites and 5 individuals of *Macroplea appendiculata* from two sampling sites (as out-group) were aligned using CLUSTAL W (Thompson *et al.*, 1994) as implemented in BIOEDIT (Hall, 1999) and edited with BIOEDIT. The alignment was

exported in a "Phylip 4" format and a median joining network (Bandelt *et al.*, 1999) was calculated using SPLITSTREE4 4.12.8 (Huson & Bryant, 2006).

2.4 – GIS-Based Analyses

2.4.1 - Mapping and Visualizing of Swan Sightings and Movements

Mute swan sighting data was obtained from EURING for sightings from Germany, Denmark, Belgium, the Netherlands, Great Britain, Ireland, Sweden, Norway, Finland, Poland and the Baltic States (Lithuania, Latvia and Estonia).
Data was sorted by year and month of sighting (discarding records before 1974 and after 2008 to reduce bias due to lack of data for some regions) and coordinates were transformed to decimal degrees using Microsoft Excel.
Data was then mapped as point data for single sightings using the free software Quantum GIS 2.0.1 (QGIS) (Quantum GIS Development Team, 2013). Individual movements were derived from and visualized by converting the point data for re-sightings of individual swans to line data, connecting the points using the "Points to Paths" plug-in. The direction of movement (as bearing of the resulting line data) was inferred using the date of sightings to determine direction (see 2.4.2).
Data had initially been converted to decimal degrees and mapped using a geographic coordinate system (WGS84). To fit the needs of the respective queries it was then re-projected. For inference of length and bearing (of movements in Northern Europe) an equidistant projection (World Equidistant Cylindrical (Sphere)) was used.

2.4.2 - Inferring the Direction of Swan Movements

The bearing of swan movements was calculated and then added as an attribute to line data shape files representing swan movements by using the field calculator function in QGIS.
The inverse tangent of x and y differences is calculated and converted to degrees (180/pi). Either 180 or 360 is then added to the resulting figure to obtain a bearing of 0° to 360°.
The following formula was used in the "expression box" of the field calculator:
(atan((xat(-1)-xat(0))/(yat(-1)-yat(0)))) * 180/3.14159 + (180 *(((yat(-1)-yat(0)) < 0) + (((xat(-1)-xat(0)) < 0 AND (yat(-1) - yat(0)) >0)*2))).

This gives direction from the beginning point (coordinate for the earliest sighting of any given individual) to the end point (coordinate for the latest sighting of any given individual) of a (poly-) line.

Similarly, the distance between sightings of individual swans (as an expression of the distance covered between sightings) was inferred by adding "length" to the attribute table by using the field calculator function (the field calculator expression is: '$length / 1000' to obtain values in kilometers in a layer projected in meters).

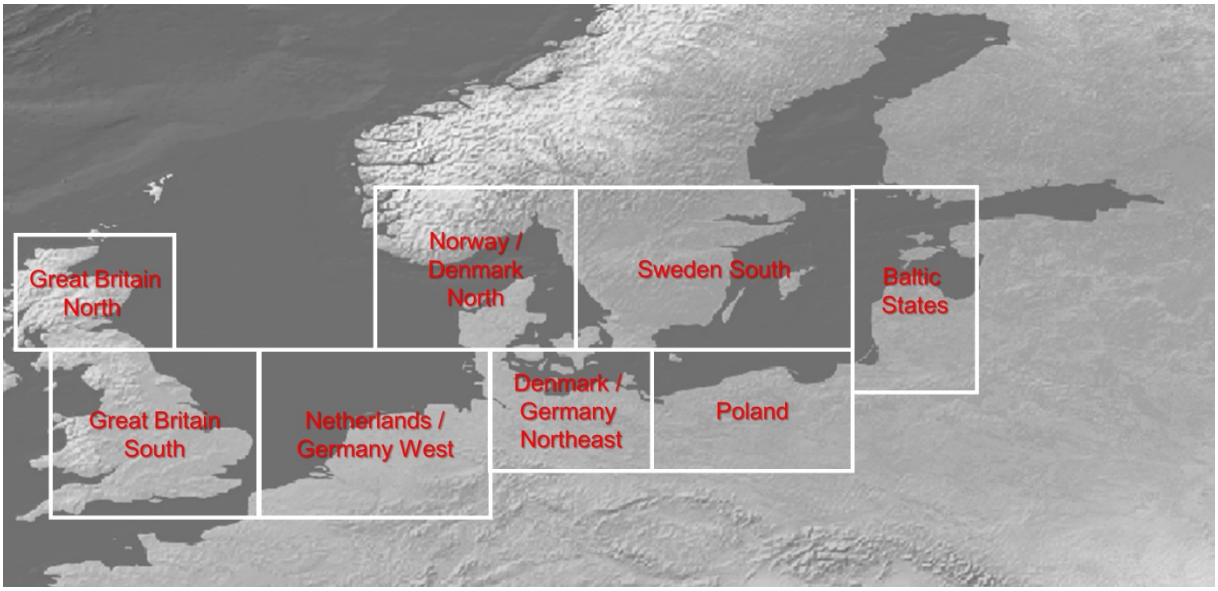


Figure 4. Query Regions for Predominant Routes of Mute Swan Movements

White rectangles represent areas considered for inference of regional mute swan movements.

To test for prevailing routes of swan movement, bearing (azimuth) of mapped movements longer than 10 km was plotted in rose diagrams for movements within six months, recorded in "summer" (April to September) and "winter (October to March) and for eight regions on the coast of Baltic and North Sea. Additionally, directional distribution of regional movements was plotted after weighting by the covered distance. The regions were chosen to encompass major parts of important sites for mute swans in Northern Europe (Names in parentheses refer to populations or parts of populations as defined by Atkinson *et al.* (2006): Southern Great Britain (England and Wales Group), Northern Great Britain (Scotland Group), Netherlands and North West Germany (Netherlands Group), Norway and Northern Denmark (Southern and eastern Norway, Norway and Northern Denmark) Denmark and North East Germany (East Germany, Schleswig Holstein), Southern Sweden (Southern Sweden), Poland (Poland) and Baltic States (Scandinavia-Baltic Group) (see Figure 4).

The directional mean was calculated and the bearing data was tested for non-random directional distribution using Rayleigh's Z equation. The null hypothesis H_0 for this procedure is a random distribution of movement directions, the alternative hypothesis is a non-random distribution.

The mean angle ($\bar{\alpha}$) is calculated as:

$$X = \frac{\sum_{i-1}^{n} \cos_{\alpha}}{n} \qquad Y = \frac{\sum_{i-1}^{n} \sin_{\alpha}}{n} \qquad r = \sqrt{X^2 + Y^2} \qquad \cos\bar{\alpha} = \frac{X}{r} \qquad \sin\bar{\alpha} = \frac{Y}{r}$$

$$\bar{\alpha} = \arctan\left(\frac{\sin\bar{\alpha}}{\cos\bar{\alpha}}\right)$$

And the critical value Rayleigh's Z is calculated as

$$Z = nr^2$$

where n is the number of observations, α_i is the i^{th} azimuth and r is the magnitude of the mean vector (and a measure of angular concentration). A table of critical Z values (Zar 1999) was used to determine if H_0 was accepted or rejected.

Testing whether the sample was oriented in a particular direction (as suggested by plots of movements or calculated mean direction) was done using the V test of circular uniformity:

V is calculated as $V = R\cos(\bar{\alpha} - \mu_0)$

μ_0 being the predicted angle and R = r n

The critical value u was calculated as $u = V\sqrt{\frac{2}{n}}$

The null hypothesis H_0 in this case states that bearings are randomly distributed with respect to the predicted direction. A table of critical V values was used to determine if H_0 was accepted or rejected. Predictions of predominant directions were based on movement plots.

Equations and tables for critical values were taken from Zar (1999).

2.4.3 - Local Swan Abundances and Swan Traffic between Sites

To infer and compare local abundances of mute swans in proximity of sampled *M. mutica* populations, spatial queries were performed. Using the software's "geoprocessing tools" option, a vector layer was created featuring buffer zones of 50 km radius around each mapped

M. mutica population. Using the "spatial query" plugin, it was inferred how many and which mapped swan sightings coincided spatially with the respective buffer zones and had hence been recorded in proximity (within a 50 km radius) of a sampled beetle population.

Similarly, the relative amount of swan traffic crossing an area comprising Jutland (Denmark) and parts of Northern Schleswig Holstein (Germany) was compared for the period between April and September and between March and October. The spatial query was performed for an area with a maximum extent from (Latitude and Longitude in decimal degrees) 57.621° to 53.924° and 7.895° to 10.765° (see Figure 28 and Figure 29), querying for line data (representing swan movements) spatially coinciding with the area.

To reduce bias from spatial and temporal differences in ringing and recovery activity, swan abundance was calculated as average number of sightings per year. Swan sighting data was recorded for each year from 1974 to 2008, discarding records above the 95 % percentile and below the 5 % percentile for each site to yield a corrected annual average of local swan sightings. This eliminates records for cases where, e.g., marked individuals were frequently recorded in smaller areas of open water near human settlements during harsh winters (and other cases of unproportionately high or low observer activity).

Since every single record for a swan sighting included an individual ring number it was possible to compare to what extent a pair of sites was connected by migration routes / movements of *individual* swans by recording how many individual swans had been sighted within a 50 km radius around *one* beetle population and - at another time - within 50 km of *another* beetle population. Two lists of ring numbers for sighted individuals were checked for entries appearing in both lists using Microsoft Excel. This was done for all 210 pairs of sampled beetle population sites resulting in a measure of swan traffic between sites. Swan traffic was recorded as the *total number* of individual swans moving between two given sites and as the *percentage* of swan individuals moving between these two sites relative to the sum of swan sighting records for both sites in that pair.

Counts of swan traffic and relative swan traffic between beetle sampling sites and average annual sighting numbers were tested for spatial structure (correlation with geographic distance between sampling sites) (similar to tests of spatial structure of genetic distances; see 2.3.2.5) with a Mantel test with 10,000 iterations.

2.4.4 - Geographic Distances between Macroplea mutica Sampling Sites

To infer geographic distances between locations of sampled beetle populations a matrix of pairwise geographic distances was generated using the analysis tools implemented in QGIS 2.0.1 (Quantum GIS Development Team, 2013).

2.4.5 - Mapping Beetle Sampling Locations against Landscape Features

Geographic coordinate data for 21 sampled *M. mutica* populations were mapped against data for European river catchments and Biogeographical regions (containing the official delineations used in Habitats Directive (Council Directive 92/43/EEC, 1992) and for the EMERALD Network set up under the Convention on the Conservation of European Wildlife and Natural Habitats (Bern Convention)). Data were provided by the European Environment Agency:
(http://www.eea.europa.eu/data-and-maps/data/european-river-catchments-1);
(http://www.eea.europa.eu/data-and-maps/data/biogeographical-regions-europe-1).

2.4.6 - Testing for Correlation of Swan Movements and Abundances with Population Genetic Structure in Macroplea mutica

2.4.6.1 - Pairwise Genetic Differentiation in M. mutica and Swan Sighting Data

Pairwise genetic distances (R_{ST}) and residual R_{ST} (see 2.3.2.6) for pairs of beetle (sub-) populations were tested for statistically significant correlation with swan abundances (the average number of swans sighted per year or half-year averaged for pairs of populations) and swan traffic (the total and the relative number of swan individuals that had been recorded at both sites for a given pair of beetle sampling sites (see 2.4.3)).
Semi-matrices of pairwise values for genetic distance between sites and similarity of sites based on detected swan individuals were tested for significant correlation performing Mantel tests (Mantel, 1967) with "ISOLDE" as included in GENEPOP (Raymond & Rousset, 1995; Rousset, 2008). R_{ST} and residual R_{ST} values were tested for statistically significant correlation with swan abundances and swan traffic with Spearman's rank correlation.

Tests included either the whole dataset or different data subsets (see 2.3.2.4). Probability values were always adjusted for multiple testing with a sequential Bonferroni correction (Holm, 1979).
To test whether observed relationships stayed true (i.e. statistically significant), when excluding smaller geographical distances, selected tests were repeated while only considering sampling site pairs of a certain minimum distance. The minimum geographic distance was gradually raised between tests to infer the largest minimum distance for which the results still implied a significant correlation.

2.4.6.2 - Breakdown of Isolation by Distance and Swan Abundances

To test the hypothesis that *M. mutica* (sub-) populations that are frequently visited by potential vectors (mute swans) might show significantly less signature of geographic distance on genetic differentiation (therefore suggesting swan mediated gene flow), isolation by distance (IBD) was compared between beetle sampling sites with high and low counts of swan sightings. Tests were performed individually for each sampling location with Spearman's rank correlation test and a subsequent Fisher's test of a 2x2 contingency table and additionally using matrices of pairwise genetic and geographic distances (representing groups of sampling sites) for Mantel tests.
First, each of the 21 beetle sampling sites was tested individually for statistically significant (positive) correlation between geographic distance to all other sites and the respective genetic differentiation towards these sites (IBD). The test was performed for each site using Spearman's rank correlation coefficients with p-values adjusted by sequential Bonferroni correction (Holm, 1979).
It was subsequently tested whether significant IBD for one site towards all other sites was significantly associated with an average number of local swan sightings *of less than* ten sightings per year (or, accordingly: If local swan abundances *of more than* ten sighted individuals per year correlate with a breakdown of the signature of IBD) with an exact two-sided Fisher's test of a 2x2 contingency table (using the method of summing small p-values). A significantly reduced influence of geographic distance on genetic differentiation among beetle sampling sites with high recorded numbers of swans would imply higher (possibly bird-mediated) gene flow between these sites.
The whole dataset of 210 population pairs and four subsets (of population pairs from the western- and eastern cluster, Baltic Sea populations and inland populations (see 2.3.2.4) were tested. P-values were adjusted for multiple testing with a sequential Bonferroni correction.

Additionally, a group of all population pairs with an average number of local swan sightings *of less than* ten sightings per year and a group of all population pairs with an average number of local swan sightings *of more than* ten sightings per year were tested for IBD with Mantel tests, testing the whole dataset and datasets for the western- and the eastern cluster (see 2.3.2.4 for composition of data subsets). P-values were adjusted for multiple testing with a sequential Bonferroni correction.
To assure that potentially non-linear IBD at small spatial scales did not significantly bias tests for IBD, all Mantel tests were repeated while only considering pairs of sites with geographic distances of 50 km or more.

2.5 - Feeding Trials and Simulation of Gut Passage

2.5.1 - Origin and Keeping Conditions of Plant Material and Beetles

Specimens of *M. mutica* and fennel-pondweed were collected at the Baltic coast of Germany. Fennel-pondweed (*Potamogeton pectinatus*) was kept in a 25 L aquarium with 120 adult *M. mutica* specimens for a week prior to the experiments to ensure a high density of clutches within the plant material. Beetles showed mating and oviposition behaviour when kept in Baltic Sea water at 16°C water temperature and 16 hours light per day.
Cocoons containing beetles were collected at Lake Selent (Schleswig Holstein) in October and tested still attached to parts of *P. pectinatus* rhizome. 30 cocoons were not treated and kept in a well-oxygenated aquarium at 12°C as a control group.

2.5.2 - Simulation of Gut Passage

Simulation of waterbird digestion was based on a protocol devised by Furman *et al*. (2006). Plant material (*Potamogeton pectinatus*) containing eggs and cocoons of *M. mutica,* respectively, was treated with artificial gizzard fluids and artificial intestinal fluids for a total time of 6 hours. Samples were constantly shaken vigorously in an incubation shaking cabinet to apply physical force in addition to the chemical treatment.

2.5.2.1 - Gizzard Phase

The artificial gizzard fluid consisted of 1M NaCl (58.44 g/L, Carl-Roth, Karlsruhe, Germany) and 10 g/L pepsin (Merck-Millipore, Billerica, USA) acidified with HCl (Carl-Roth,

Karlsruhe, Germany) to pH 2.6. Thirty millilitres (ml) were combined with plant material (*P. pectinatus*) containing eggs or cocoons of *M. mutica* in 50 ml plastic-tubes. The samples were shaken for three hours at 42°C and 250 rpm.

2.5.2.2 - Intestinal Phase

One ml of $NaHCO_3$-solution (saturated, Carl-Roth, Karlsruhe, Germany) was added to each tube (resulting in measured pH-values of 6.2-7.0). Bile salts (0.105 g, equivalent to 0.35% bile, Sigma-Aldrich, St. Louis, USA) and 0.0105 g pancreatin (equivalent to 0.035% pancreatin, Merck-Millipore, Billerica, USA) were then added to each tube and samples were again subjected to three hours of shaking at 250 rpm and 42°C.

The samples were afterwards rinsed with tab water and searched for cocoons or eggs. Retrieved eggs and cocoons were then kept in well-oxygenated aquaria and checked daily for hatching larvae or beetles.

2.5.3 - Feeding Trials

Feeding trials were conducted in June 2012 with two mallards (one female and one male). Mallards were chosen mainly since they are the species most studied as potential dispersers (Charalambidou & Santamaría, 2002). The birds were kept in an enclosure of 6 m^2 containing shelter and a water basin. They were fed food pellets for waterfowl (made from milled wheat, corn and soy), salad and pondweed and had access to grit (fine gravel) ad libitum. For the trials, the animals were presented with fennel pondweed containing *Macroplea* eggs. The experiment was conducted four times at one-week intervals.

Plant material was offered in an area covered with Perspex sheeting, to control for spilt food items. During feeding, the birds were watched from a distance to make sure that no food was carried away while avoiding undue stress to the birds.

Faeces were collected two, five and eight hours after feeding, discarding parts in contact with the soil. Samples within a 50 cm radius around the feeding site were excluded to avoid collecting uneaten plant material. Faeces were suspended in water directly after collecting and searched for eggs under a dissecting microscope (Leica, Wetzlar, Germany). Retrieved eggs were transferred to small plastic cylinders with nylon membrane-covered openings, kept in well-oxygenated aquaria with Baltic Sea water and checked daily for hatched larvae. Larvae displaying normal movements and reaction to physical stimuli were scored as “alive”.

Housing conditions and feeding trials were approved by the veterinary offices of the city of Kiel (Schleswig Holstein) and Hamburg and the animal protection representative of the city of Hamburg.

3 - Results

3.1 - Population Genetic Analyses

3.1.1 - Characterization of Developed Microsatellite Markers

3.1.1.1 - Basic Information on the Genomic Library and Amplification of Loci

The finished genomic library (enriched for microsatellites) contained sequences for 18,738 DNA fragments (contigs). 3182 contigs contained simple sequence repeats (SSRs). 933 contigs containing SSRs remained after elimination of redundant sequences (complementary sequences and multiple copies). 685 contigs containing SSR fulfilled the requirements for a minimal number of motif repeats. Including redesign of primers, 62 primer pairs for 58 loci were screened for polymorphisms. Six polymorphic loci could reliably be amplified in two multiplex sets (of three primer pairs each). 256 individuals from 21 Northern European sampling sites were genotyped for these six newly developed microsatellite loci. All loci proved to be highly polymorphic with an average of 17 detected alleles per loci (for size ranges and detected alleles per locus see Table 6). A full list of all called alleles for each tested individual (allele names equaling measured PCR product sizes in base pairs (Appendix A) as well as expected and observed heterozygosities (Appendix B) for each locus and population are presented in the appendices.

Table 6. Basic Information on Primers

Basic information on primers used to amplify six microsatellite loci in *Macroplea mutica* and summary statistics based on the amplification in 256 individuals from 21 populations in Northern Europe.

Locus	Dye-label	Sequence	Repeat motif	Size Range	Alleles	T_a (C°)
loc16359	DY-682	F: TTTGGCGGGATTGCACTTG R: AGAGGTTCTATCAAACTGTACCAC	ATAC	127-179	12	60
loc4107	DY-682	F: TGTTGTCTGACGTAACTCTGC R: GAGTCTAACAAGACCATCTGTCG	TC	177-433	13	60
loc12208	CY5	F: GTGAGACGTGAAACGGCAG R: AGGGTTCGTAGTCGGTATGC	GT	130-206	16	60
loc321	DY-682	F: CCTTCGTAGGAACTTTAGGCG R: GACGAGGCGTGCTGTTTAG	GTT	275-443	16	51
loc3012	DY-682	F: ATCCAGCTAACCAGATGGC R: GTTGGGTTCAGCGCGTATC	ATT	145-286	37	51
loc1624	CY5	F: TAAGGGTCGAATGGGCAGG R: GTGCAGCATCTGGTTCACG	CAA	377-416	10	51

F: forward primer sequence; R:reverse primer sequence;Size Range: detected allele sizes in bp; Alleles: number of detected alleles; T_a, primer annealing temperature.

3.1.1.2 - Tests for Null Alleles and Stuttering

An analysis with MICRO-CHECKER (Van Oosterhout *et al.*, 2004) showed no evidence for scoring errors due to stuttering or evidence for large allele dropouts. Relative homozygote excess for several size classes indicating possible presence of null alleles was detected for five out of six loci. The estimated average frequency of possible null-alleles for a single locus was always below 8% and below 6% across loci.

3.1.1.3 - Tests for Hardy-Weinberg Equilibrium and Linkage Disequilibrium

According to a test with GENEPOP 4.2 (Raymond & Rousset, 1995; Rousset, 2008) 18 of 21 populations showed single-locus deviations from Hardy-Weinberg equilibrium, involving 23 of 124 estimates. However, only 10 remained significant after a sequential Bonferroni correction and those showed homozygote excess. Out of 315 estimates of linkage disequilibrium 11 showed interlocus associations, none of which remained significant after a sequential Bonferroni correction. No locus pair showed significant interlocus associations across all populations.

3.1.2 - Population Structure and Genetic Differentiation

3.1.2.1 - Assessing Statistical Power of the Used Set of Microsatellites

Table 7. Estimates of Resolution Power

Estimate of the resolution power of the used set of microsatellite markers using POWSIM (Ryman & Palm, 2006).

Average F_{ST}	Expected F_{ST}	N_e	t	χ^2-test	Fisher's test
0	0	500	0	0.0845	0.0689
0.005	0.005	500	5	0.8076	0.7443
0.006	0.006	500	6	0.9010	0.8562
0.007	0.007	500	7	0.9536	0.9239
0.008	0.008	500	8	0.9825	0.9665
0.009	0.009	500	9	0.9916	0.9848
0.010	0.010	500	10	0.9975	0.9947
0.011	0.011	500	11	0.9993	0.9980
0.012	0.012	500	12	0.9998	0.9996
0.013	0.013	500	13	1	1

Resolution power is assessed by simulating different expected levels of F_{ST} according to the effective population size (N_e) and generations of divergence (t). The last two columns show the proportion of significant tests for Fisher's exact test and χ^2 tests, reflecting the power to detect a given level of differentiation (average F_{ST}) with the used sampling design and used markers. A setting of $F_{ST} = 0$ and $t = 0$ estimates the α-error in the absence of differentiation. Values for N_e and t used for simulated drift processes do not reflect assumptions about the real demographics or evolutionary history of the studied populations.

Estimating the resolution power of the used set of microsatellite markers using POWSIM (Ryman & Palm, 2006), the α-error for Fisher`s test was closer to 5% than the α-error for χ^2 and therefore closer to the desired value for a reliable test when using the 5% significance limit. According to the tests of resolution power (and Fisher's test), the six used microsatellite markers should allow for a detection of F_{ST} values as low as 0.008 with a probability of > 96%. The probability for detecting a differentiation equivalent to $F_{ST} = 0.0129$ is 100% (for the present set of six microsatellite loci, the detected allele frequencies, the present number of samples, and their distribution over 21 sampling locations / (sub-) populations (Table 7).

3.1.2.2 - Global and Pairwise Population Genetic Differentiation

The exact G test for pairs of populations as implemented in GENEPOP 4.2 (Raymond & Rousset, 1995; Rousset, 2008) showed significant genetic differentiation ($P \leq 0.05$) for all but two of the 210 tested pairs of populations in the data set when testing across all loci. The only exception was a population from Finland ("DRA") that did not show significant differentiation to two neighbouring populations from Finland ("KIR") and Estonia ("VOR").

The highest pairwise R_{ST} value (0.898) was recorded between Cambridgeshire, UK ("CAM") and West Finland ("DRA") while the lowest pairwise R_{ST} value (-0.095) was measured for Sjaelland, Denmark ("OBJ") and Stockholm South, Sweden ("UTO"). Incongruently, the lowest pairwise F_{ST} value (0.0192) was recorded for Estonia ("VOR") and Finland West ("DRA") and the highest F_{ST} value (0.470) was measured between populations in Wales, UK ("WAL") and Estonia ("VOR"). The ten percent of sampling site pairs showing the highest differentiation were 23.8 % identical (i.e. consisted of the same sampling site pairs) between F_{ST}- and R_{ST} estimates. Likewise, the ten percent of sampling site pairs showing the lowest differentiation were 28.6 % identical between F_{ST}- and R_{ST} estimates

Table 8. Population Genetic Differentiation.

F_{ST} values, R_{ST} values and exact G-test results per locus and over all loci. Asterisks for the exact G-test indicate significant genetic structuring among samples: ***- $p \leq 0.001$.

Locus	F_{ST}	R_{ST}	Exact G-test
loc16359	0.1435	0.1527	***
loc4107	0.2273	0.0167	***
loc12208	0.1427	0.2033	***
loc321	0.1522	0.3573	***
loc3012	0.1757	0.1759	***
loc1624	0.1389	0.0629	***
All:	0.1629	0.2517	***

Overall population differentiation assessed by the exact G-test was highly significant for all loci and across all loci (see Table 8). F_{ST} values varied comparatively little between loci and ranged from 0.1389 (loc1624) to 0.2273 (loc4107) with an averaged value of 0.1629. R_{ST} showed a wider range of values from 0.0167 (loc4107) to 0.3573 (loc321) and a much higher average of 0.2517. Full lists of pairwise distances (Nei's D_A) and pairwise R_{ST}- and F_{ST} values are presented in the appendices (Appendix C).

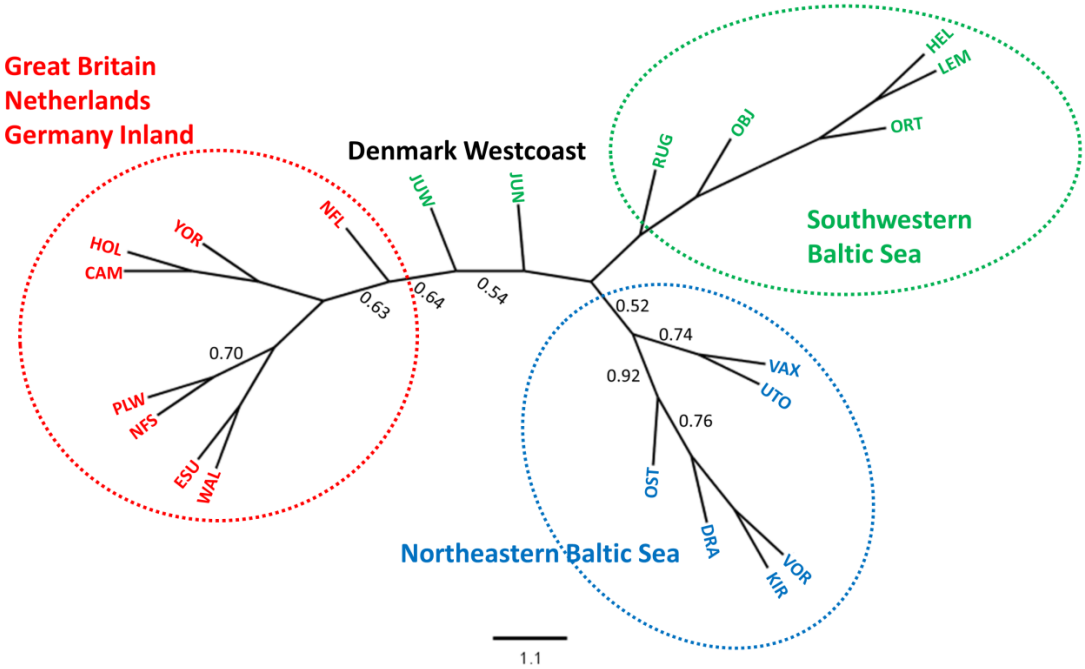


Figure 5. Neighbour Joining Tree

Neighbour joining tree (Saitou & Nei, 1987) based on Nei's chord distance, D_A, (Nei *et al.*, 1983), fit: R^2= 0.825. Bootstrap values show results of 10 000 bootstrap runs for results > 0.5.

A neighbour joining tree (Saitou & Nei, 1987) based on Nei's chord distance, D_A, (Nei *et al.*, 1983) (Figure 5) showed the British populations clustering paraphyletically at one end of the cladogram with the Dutch population and the German populations from Southern North Frisia ("NFS") and the Plön Lakes ("PLW") nested in between. The Baltic Sea populations clustered at the opposite end of the cladogram with the (North-) Eastern Baltic Sea populations ("VAX", "UTO", "OST", "DRA, "KIR" and "VOR") as a clearly separated sistergroup to the (South-) Western Baltic Sea populations ("RUG", "OBJ", "ORT", "LEM" and "HEL").

While the tree topography fits the data well (R^2=0.825) the bootstrap support is weak (<0.75) for most steps, with the exception of subdivision between the Western and the Eastern Baltic populations and within the Eastern Baltic populations.

While the populations from North Frisia ("NFL") and the two Danish inland populations from Jutland ("JUW" and "JUN") are geographically and genetically positioned inbetween the Western and the Eastern cluster, the two German inland populations from Southern North Frisia ("NFS") and the Plön Lakes ("PLW") showed up exceedingly well differentiated from geographically neighbouring Baltic Sea populations (e.g. from the Little Belt population ("HEL") and the samples from off the coast of Fehmarn ("LEM" and "ORT")).

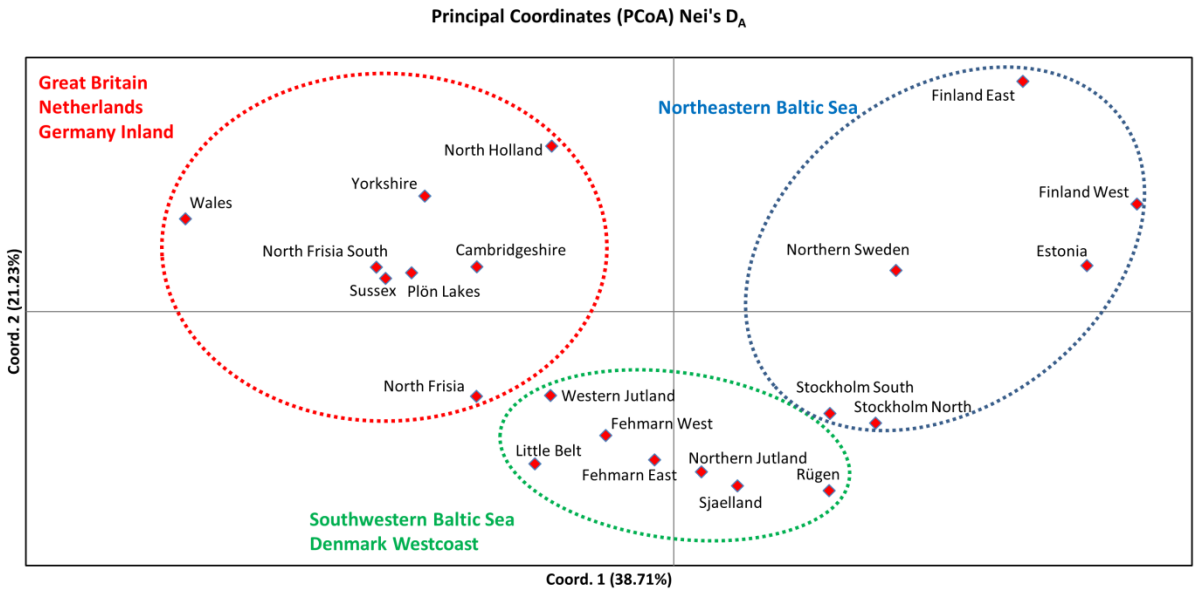


Figure 6. Principal Coordinate Analysis

Principal coordinate analysis cluster diagram based on pairwise distances (Nei's D_A) for 21 Northern European populations of *M. mutica*. The x-axis explains 38.71% and the y-axis 21.23% of variation. Broken lines represent geographic regions and not necessarily clusters suggested by the PCoA results.

A principal coordinate analysis (PCoA) computed on all 21 populations provided additional visual representations (Figure 6) of estimated genetic differentiation between populations based on Nei's chord distance D_A (Nei *et al.*, 1983). Results of the PCoA show the western populations in the top left corner, the south-western Baltic Sea populations in the bottom middle and the eastern Baltic Sea populations in the top-right corner.

Notably, the German populations from the Plön Lakes and Southern North Frisia were shown clustered with the English populations from Sussex and Cambridgeshire while showing considerable distance to geographically neighbouring populations in the Baltic Sea (Fehmarn and Little Belt). The population from North Frisia appeared close to geographically neighbouring populations from Western Jutland and the Little Belt. The two populations from Stockholm archipelago appeared slightly closer to the cluster of populations from the South-western Baltic Sea than to the (only loosely clustering) populations from Northern Sweden, Finland and Estonia.

3.1.2.3 - Testing the Relative Performance of F_{ST} and R_{ST} Estimates

The contribution of stepwise mutation relative to drift and migration on population differentiation was tested with 10,000 random permutations of allele sizes along allelic states with SPAGEDI 1.4 (Hardy & Vekemans, 2002). The results showed significant differences between pairwise and global R_{ST} values observed before and after (pR_{ST}) randomization.

Pairwise R_{ST} values were (on average and across all loci) significantly larger than pairwise pR_{ST} values (P (1-sided test, H1: $R_{ST} > pR_{ST}$) = 0.031), global R_{ST} values were (on average and across all loci) also significantly larger than global pR_{ST} values (P (1-sided test, H1: $R_{ST} > pR_{ST}$) ≤ 0.001). This implies that R_{ST} shows a lower square mean error than F_{ST} and therefore provides better representations of the studied population differentiation.

3.1.2.4 - Testing Genetic Distances for Congruence with AFLP Data

The results of the microsatellite analysis were tested for congruence with results of an AFLP analysis by Mende et al. (2010). Pairwise F_{ST} data for 16 populations that had been subject to both studies were compared with a Mantel test and Spearman's rank correlation test. The microsatellite analysis was validated showing very similar pairwise genetic differentiation (Mantel test, 10,000 permutations: r = 0.512; n = 16; P ≤ 0.001; Spearman's rho: 0.512; P ≤ 0.001).

3.1.2.5 - Estimation of the Number of Genetically Distinct Clusters

Bayesian inference of the optimal partition size resulted in concordant results for analyses with BAPS and STRUCTURE when run without considering geographic coordinate information and assuming a correlated allele model.

Using a spatial prior (spatial clustering with BAPS and GENELAND) and an independent allele frequency model (spatial clustering with GENELAND) resulted in lower estimates for the number of genetically distinct clusters "K".

Given a vector of possible upper limits for K from k=1 (one cluster / panmixis) to k=50 (50 genetically distinct groups) (k=1, k=2, k=3, k= 4, k=5, k= 10, k=20, k=50) a group level mixture analysis with BAPS resulted in an optimal partition with nine clusters. Out of the ten best visited partitions, five showed sizes of nine clusters and five showed sizes of ten clusters. According to the marginal likelihood values the overall probability was 0.88 (88 %) for an optimal size of nine clusters and 0.12 (12 %) for a size of ten clusters. Repeating the analysis considering spatial priors resulted in an optimal partition with only eight clusters (with a probability of 0.98 (98%) for K=8 and only 0.02 (2%) for K=9 (Figure 7 and Figure 8).

Concordantly with the BAPS mixture analysis, average likelihood values for 20 independent STRUCTURE runs for each K value from K=1 to K=25 (Figure 9) showed a size of nine clusters (K=9) for the partition with the highest average log. probability. Comparison of marginal likelihoods showed a probability of > 0.9999 (> 99.99%) probability for a partition

of K = 9. The probability for the partition size with the next highest average log. likelihood value (K=10) was insignificantly low (calculated as $1.201e^{-05}$ (< 0.01%)).

Spatially explicit Bayesian inference with five independent GENELAND runs of 100,000 MCMC iterations each with possible values for K between K=1 and K=25 and an *un*correlated allele model resulted in a predicted partition size of K=3. All five individual runs resulted in a partition size of K=3 (regardless of consideration of spatial data).

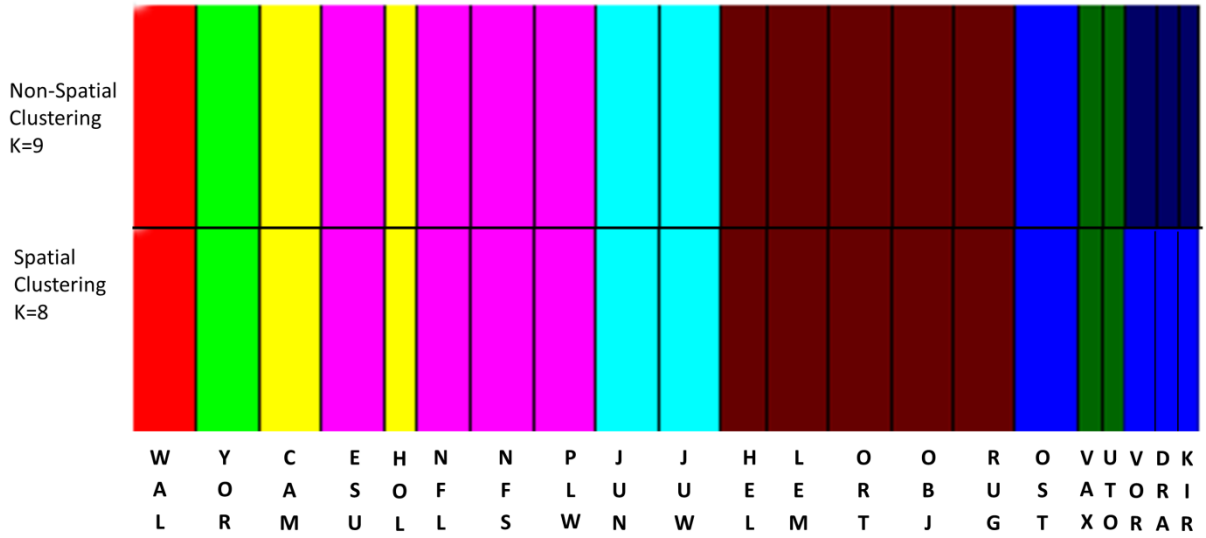


Figure 7. Clustering of Groups with BAPS (K=1-K=50)

Optimal partitions (K=8 / K=9) as results of clustering of groups of individuals with BAPS for possible values for K = 1 to K= 50 with (top half) and without (bottom half) consideration of spatial data. Different colours represent genetically distinct clusters.

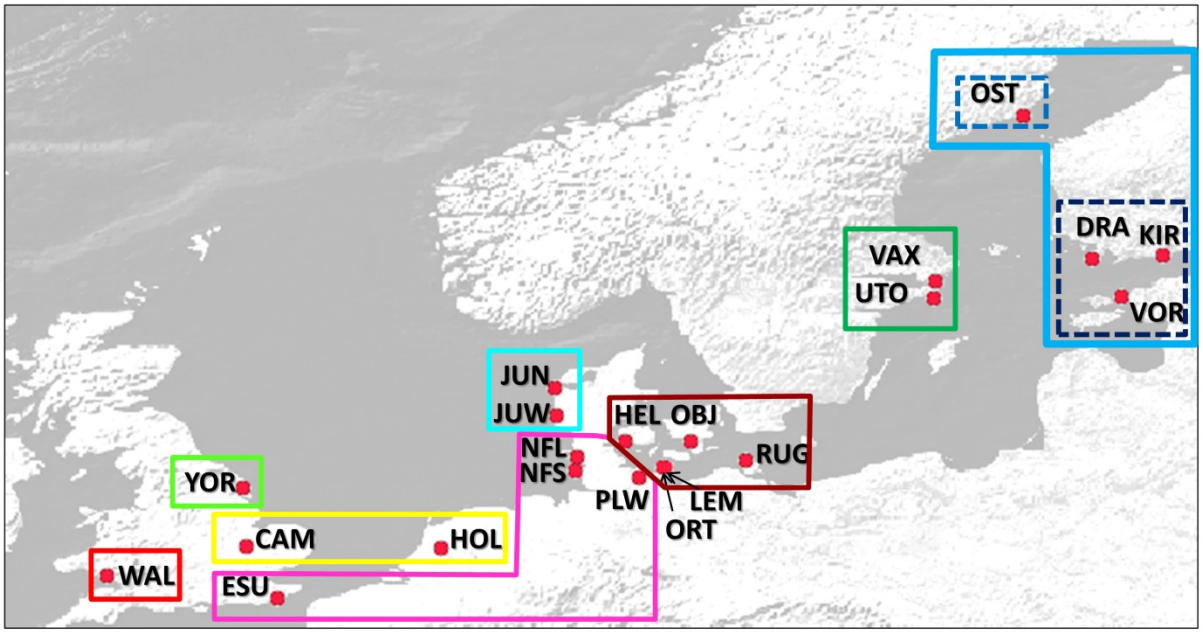


Figure 8. Map of BAPS Clustering Results

Results of (spatial) clustering of groups of individuals with BAPS.
Lines symbolize shared membership to genetic clusters. Broken lines represent cluster sub-division as only featured in the result of non-spatial clustering.

3.1.2.6 - Inferring the Uppermost Level of Population Structure

Although the highest average log likelihood after 20 independent STRUCTURE runs for each k value from K=1 to K=25 was observed for K=9, the modal value of ΔK (an ad hoc quantity based on the second order rate of change of the likelihood function with respect to K) suggested a ”true number of groups K” of K=2. The stepwise calculation of ΔK from L(K) is shown below (Figure 9- Figure 12). The increase of L(K) is strongest from K=1 to K=2 and the modal value of ΔK suggested that the “true number of K” or the uppermost level of structure is K=2. The modal value of ΔK (for K=2) is the critical value (Evanno *et al.*, 2005) but a mentionable, high ΔK value for K= 3 was observed.

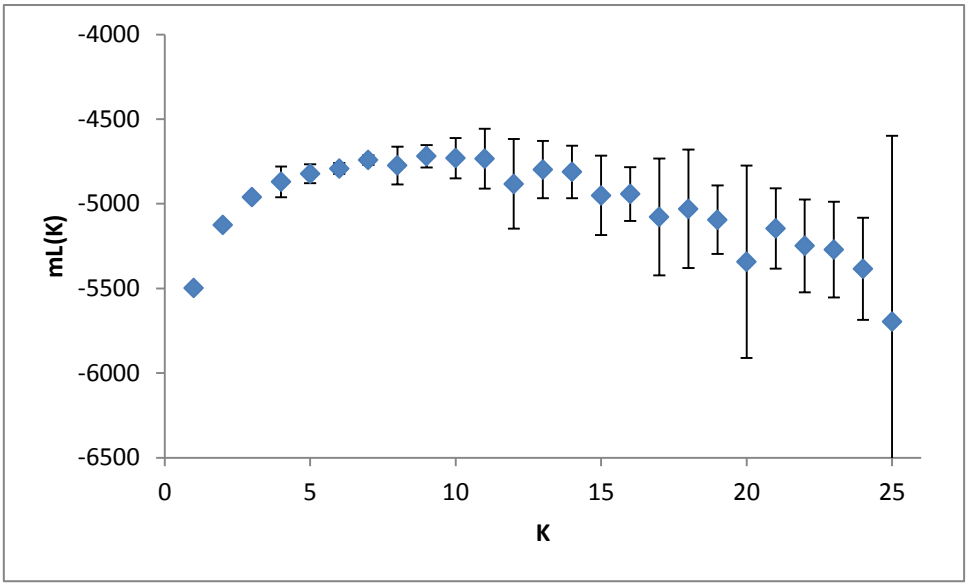


Figure 9. Mean Likelihood for Number of Clusters (K): mL(K)

Mean likelihood (mL(K)) for 20 runs. The number of clusters / populations is shown on the x-axis (K). Vertical error bars show standard deviation.

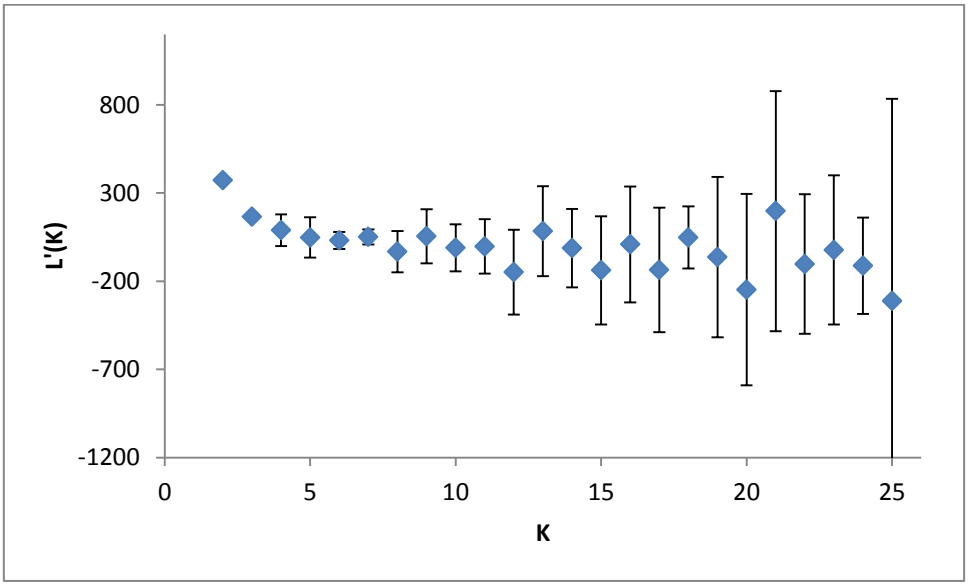


Figure 10. Rate of Change of the Likelihood Function with Respect to K

Mean difference (for 20 runs) between successive likelihood values of K: L' (K) = L(K)-L(K-1)
The number of clusters / populations (K) is shown on the x-axis. Vertical error bars show standard deviation.

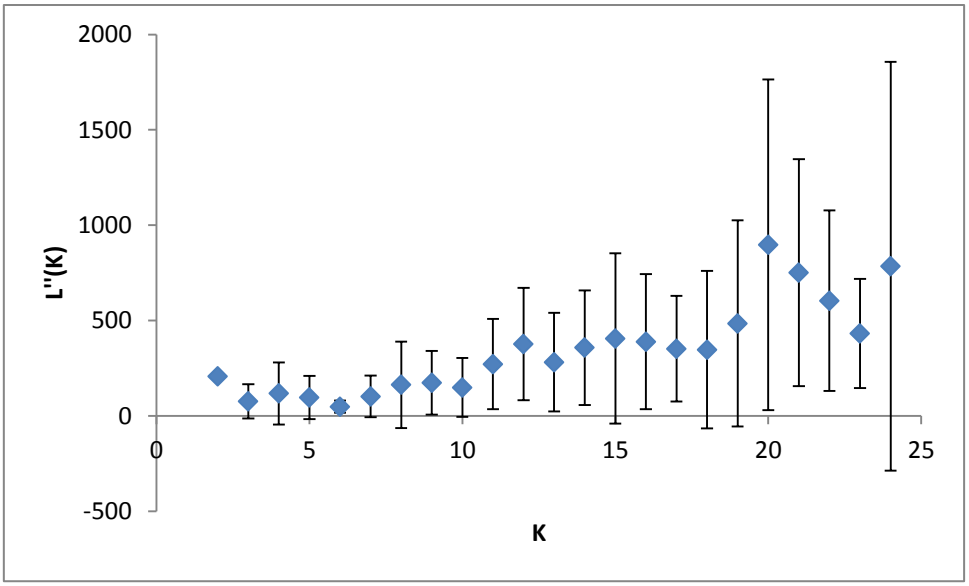


Figure 11. Second order rate of change of the likelihood function with respect to K

(Absolute value of) difference between successive values of K
$|L''(K)| = |L'(K+1)-L'(K)|$. Mean for 20 runs. The number of clusters / populations (K) is shown on the x-axis. Vertical error bars show standard deviation.

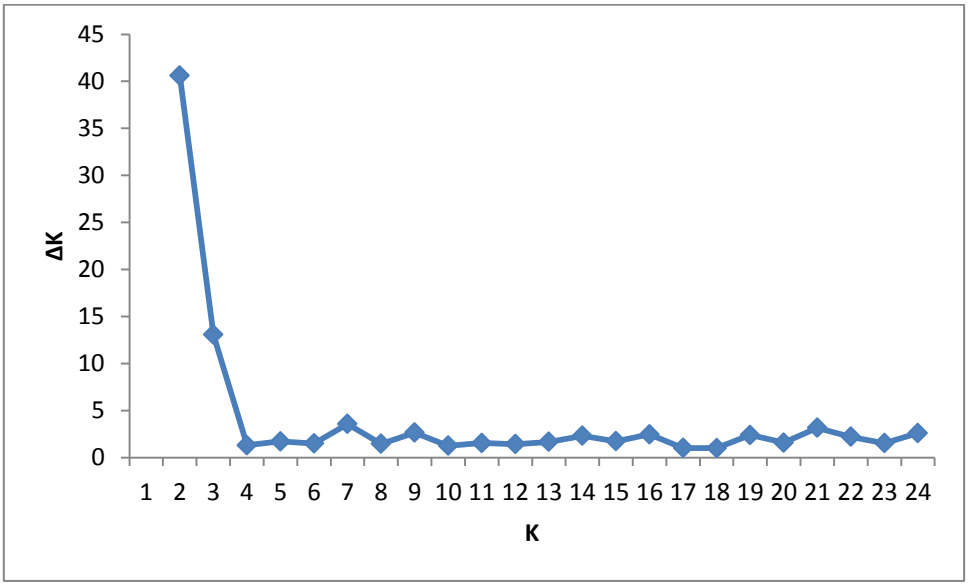


Figure 12. ΔK

ΔK or uppermost level of STRUCTURE (number of clusters)
$\Delta K = m|L''(K)|/s[L(K)] = (\text{mean}\ |L''(K)|)/s[L(K)]$. The number of clusters / populations (K) is shown on the x-axis.

3.1.2.7 - Allocation of Groups to Detected Clusters

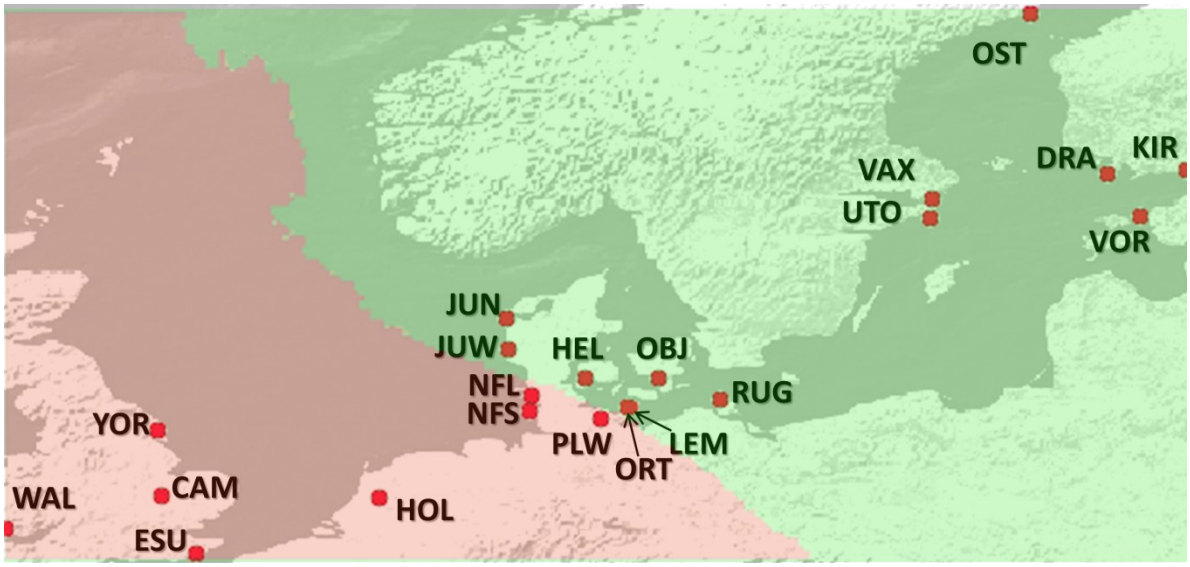


Figure 13. Spatial Clustering with Geneland (K=2)

Results of spatial clustering with GENELAND (assuming a partition of K=2).
Differently coloured zones are the result of a Voronoi tessellation and represent estimates of membership to two genetic clusters, but do not necessarily represent the geographic extent of these clusters.

While estimates for the most probable number of genetic clusters differed between methods of Bayesian inference, allocation of populations to a given number of clusters was largely congruent between analyses with different programs.

For a partition of K=2 (the uppermost level of population genetic structure according to the ΔK method) all used programs congruently allocate the 21 populations in the data set: The British populations, the Dutch population and the German inland populations form a western cluster and the populations from the west coast of Denmark and all Baltic populations form an eastern cluster (Figure 13).

Since the admixture model used with STRUCTURE does not assign individuals (or groups of individuals) to clusters, the average proportions of membership of each population to a given number of clusters were considered. Proportional membership to two assumed clusters (according to 20 STRUCTURE runs) is shown in Figure 14. Every population showed membership to one of two clusters with more than 55%. The most ambiguous case of proportional cluster membership (56.8% (±0.5%)) was observed for the West-Danish population "FT" in the Eastern cluster and the second most ambiguous case of proportional cluster membership was shown for the Northern German population "NFL" (74.4% (±0.7%)) in the Western cluster.

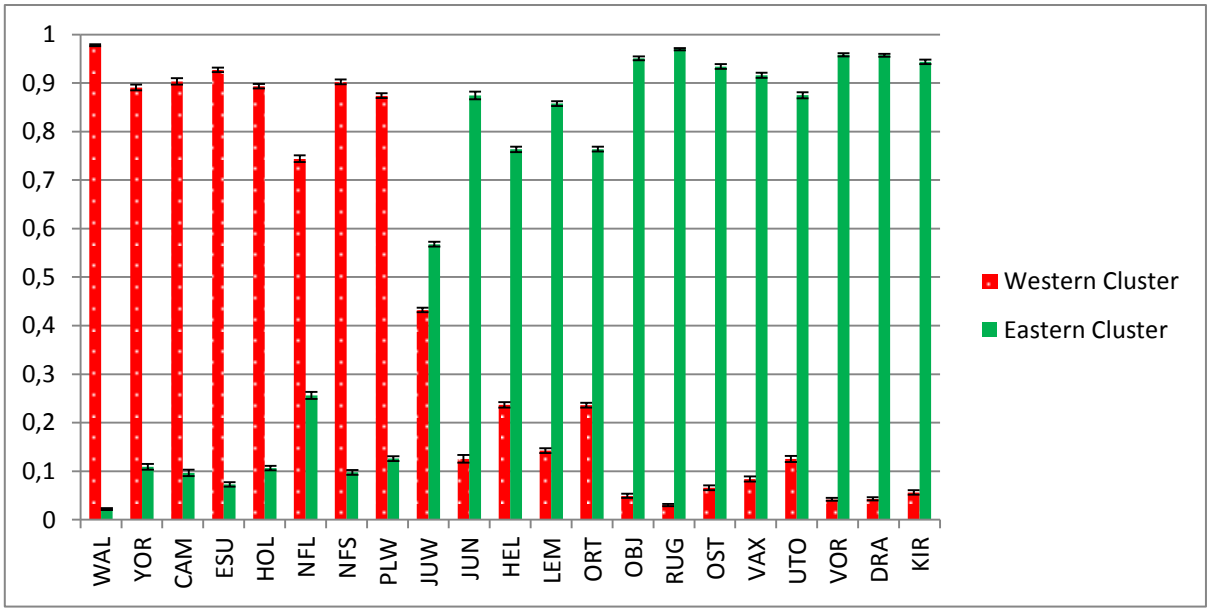


Figure 14. Proportions of Cluster Membership

Average proportion of membership of each (pre-defined) population in each of the two clusters according to 20 independent STRUCTURE runs. Error bars show standard deviation.

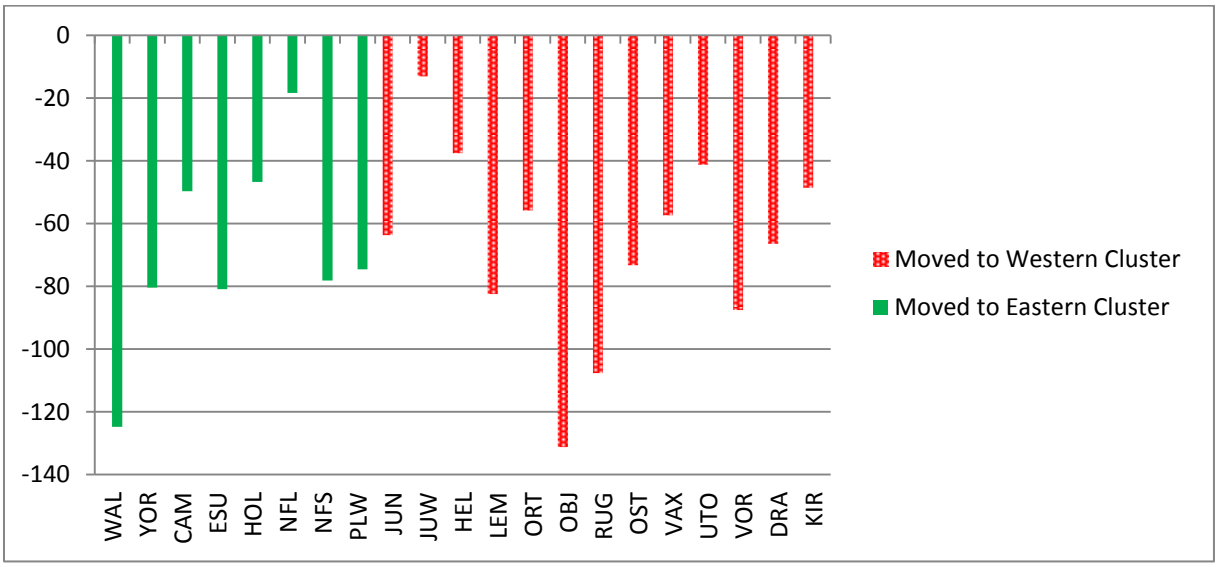


Figure 15. Change in Log Likelihood of Cluster Membership (K=2)

Change in log. likelihood when a population is moved to a (neighbouring) cluster. BAPS clustering of groups of individuals. 1000 runs for a fixed value of K=2.

Similarly, results of clustering with BAPS (20 runs for a fixed value of K=2) show changes of log. marginal likelihoods when a group is moved to another cluster (Figure 15) are smallest for "NFL" and "JUW" illustrating that allocation was less definite for these populations.

For a partition of K=3, allocation of populations by the programs BAPS (Figure 16 and Figure 19), STRUCTURE (Figure 18) and GENELAND (Figure 17) are again congruent in the divide between the "western" and the "eastern" populations (as in the partition for K=2). As the

delimitation of these three clusters is largely correspondent with the Atlantic-, the continental- and the boreal biogeographic regions (as defined in the European Council Directive 92/43/EEC, 1992) (see Figure 23) the western cluster will hence be referred to as the "Atlantic" cluster, the middle cluster will be referred to as the "continental" cluster and the eastern cluster will be referred to as the "boreal" cluster. While the first of three clusters (the "Atlantic" cluster) is in every instance identical with the "western" cluster (as in the partition for K=2), the "eastern" cluster is further sub-divided into two clusters. The westernmost cluster of these two (the "continental" cluster) is formed by the Danish and Baltic German populations and the easternmost ("boreal") cluster consists of the two Finnish, the Estonian and the three Swedish populations (although there was some ambiguity in case of the allocation of at least one of the Southern Swedish populations (Figure 18 and Figure 19).

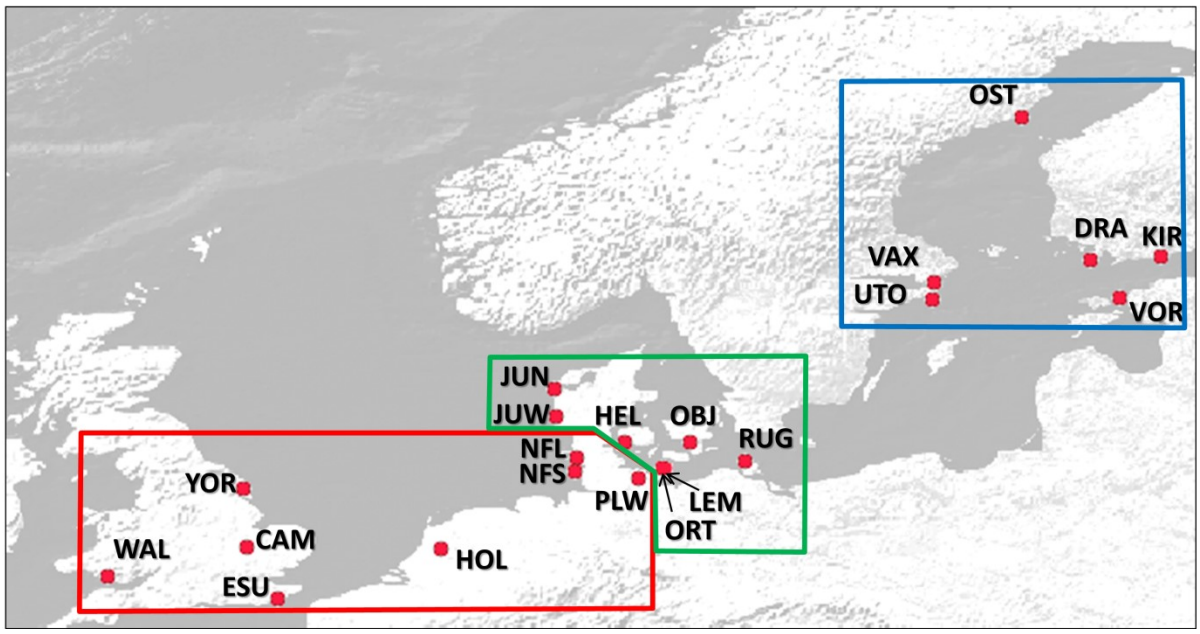


Figure 16. Clustering with BAPS (K=3)

Allocation of 21 populations to three clusters according to 1,000 BAPS runs of (spatial) clustering of groups of individuals with a fixed value of K=3. Lines symbolize shared membership to genetic clusters.

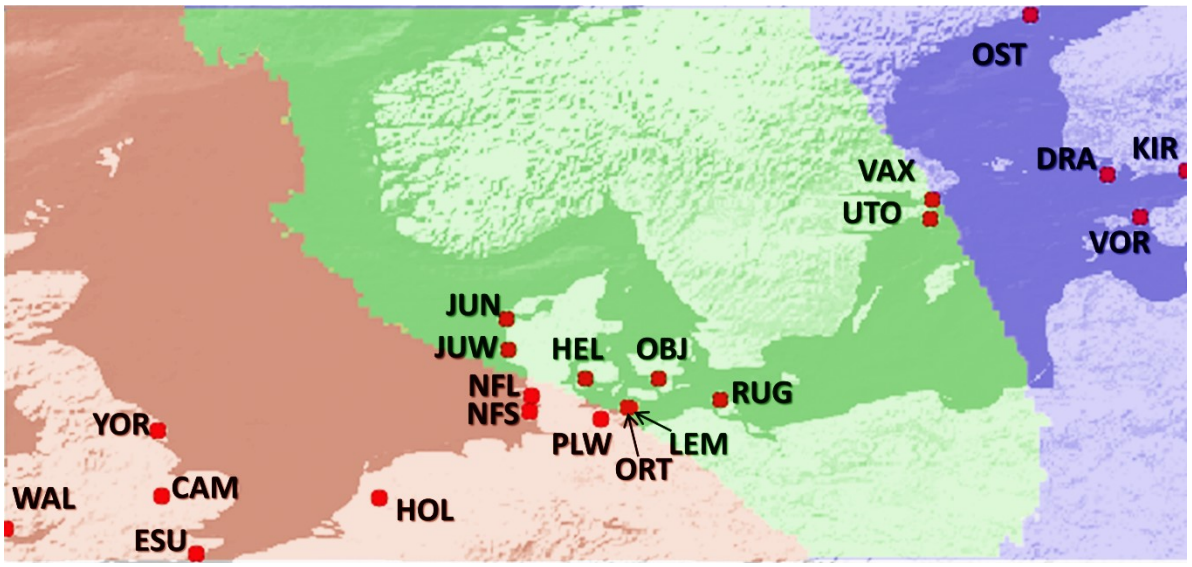


Figure 17. Spatial Clustering with Geneland (K=1 to 25)

Results of spatial clustering with GENELAND (for tested partition sizes from K=1 to K=25) with an uncorrelated allele model show an optimal partition of K=3. Differently coloured zones are the result of a Voronoi tessellation and represent estimates of membership to different genetic clusters, but do not necessarily represent the geographic extent of a cluster.

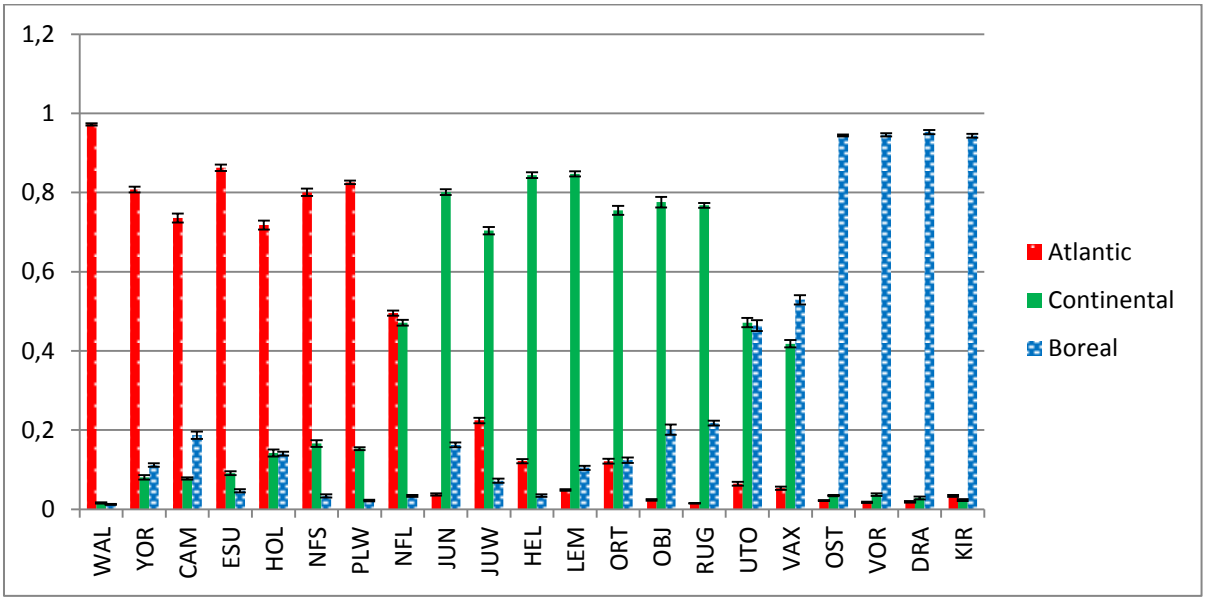


Figure 18. Proportions of Cluster Membership (K=3)

Average proportion of membership of each (pre-defined) population in each of the three clusters according to 20 independent STRUCTURE runs. Error bars show standard deviation.

Plotting the proportion of cluster membership (according to the results of 20 STRUCTURE runs for K=3) (Figure 18) suggested that most populations could be rather distinctly allocated to one of the three clusters with the possible exception of three populations: The North Frisian population “NFL” and two Swedish populations (“VAX” and “UTO”).

"NFL" is allocated to the Atlantic cluster by GENELAND and BAPS despite its geographical adjacency to populations of the continental cluster. The result showed a proportional membership of 55% (±0.9%) to the Atlantic cluster and 41% (±0.9%) to the continental cluster. This fits the representation in the neighbour joining tree (Figure 5) where "NFL" is positionend between eastern and western populations, near the Danish inland populations "JUN" and "JUW" in Jutland (which in turn GENELAND and BAPS assigned to the continental cluster). The results of a principal coordinate analysis also showed "NFL" close to the continental cluster (Figure 6).

The two closely neighbouring Swedish populations from Stockholm archipelago ("VAX" and "UTO") are the only populations for which assignment to a partition of K=3 is not congruent between the clustering results of GENELAND, STRUCTURE and BAPS. While "VAX" showed proportionally stronger membership to the Boreal cluster (52% (±1.0%)) than to the continental cluster (43% (±0.7%)), the closely neighbouring population "UTO" shows only 39% (±0.5%) membership to the boreal cluster and 55% (±0.6%) membership to the continental cluster according to STRUCTURE results. Clustering with GENELAND assigns both to the continental cluster (with a probability of membership of 0.9 (90%) for the continental cluster and 0.1 (10%) for the boreal cluster) while all other populations are allocated to their respective clusters with a probability of 1.0 (100%). In contrast, spatial and non-spatial clustering with BAPS (20 runs for a fixed value of K=3) assigns "VAX" and "UTO" to the boreal cluster. Changes of log. marginal likelihoods when a group is moved to another cluster (Figure 19) were smallest for "NFL", "VAX" and "UTO" illustrating that allocation was less definite for these populations.

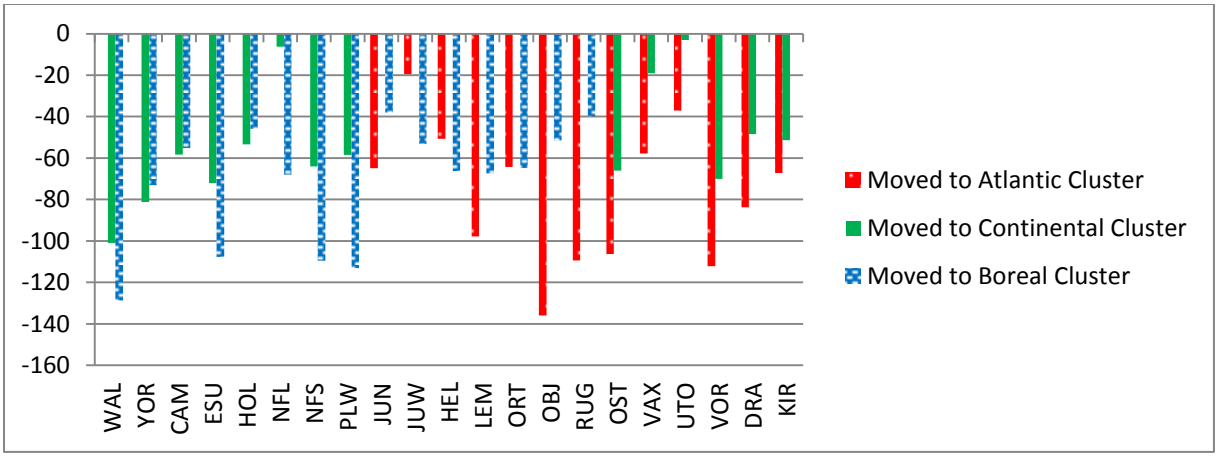


Figure 19. Change in Log Likelihoods of Cluster Membership

Change in log likelihood when a population is moved to a (neighbouring) cluster (BAPS clustering of groups of individuals. 1000 runs for a fixed value of K=3.

Allocating the (geographically pre-defined) populations to nine clusters without a spatial prior, BAPS assigns the population from Wales (“WAL”) and Northern England (“YOR”) to individual clusters. Notably, the population from Cambridgeshire (“CAM”) forms a cluster with the Dutch population (“HOL”) and the Southern English population (“ESU”) was allocated in a further cluster with the German populations “NFL”, “NFS” and “PLW”.
Two populations from the east coast of Denmark (“JUN” and “JUW”) form a cluster and the biggest cluster consists of five populations from the Baltic Belt Sea (“HEL”,”LEM” and ”ORT”) and off the coast of the island of Rügen (“RUG”). One population from the North of Sweden (“OST”) was allocated to an individual cluster, two populations of the coast of the Northern Baltic Proper form a second Swedish cluster (“VAX” and “UTO”). The last cluster consists of two Finnish populations (“DRA” and “KIR”) and a population from the Estonian coast (“VOR”). When spatial data is considered, assignment remains identical apart from one exception: The Northern Swedish population (“OST”) is clustered with the two Finnish populations and the Estonian population instead of being assigned into a cluster of its own. For a graphical representation see Figure 8.

3.1.2.8 - Admixture Based on Mixture Clustering with BAPS and STRUCTURE

Admixture in individuals that was detected by BAPS with a probability of $p \leq 0.05$ is plotted in the top half of Figure 20 (K=2) and Figure 21 (K=3). STRUCTURE was run with the admixture model (under the general assumption of admixed individuals) and the results of STRUCTURE runs displayed all individuals as to some degree admixed (Figure 20 and Figure 21).
Based on a BAPS mixture analysis for K=2, only two admixed individuals and one migrant were detected with a probability of $p \leq 0.05$. Samples from Western Jutland (“JUW”) contain one admixed individual that was allocated to the Western cluster with a proportion of 0.8 (80%) and one individual that appears to be a migrant that is fully assigned to the Western cluster with 100%. Furthermore, the samples from the Little Belt (“HEL”) contain an admixed individual that is assigned to the western cluster with 86%.
STRUCTURE results for K=2 showed samples from Northern Frisia (“NFL”), Western Jutland (“JUW”), Fehmarn West (“ORT”) and the Little Belt (“HEL”) as comparatively strongly admixed (Figure 20).

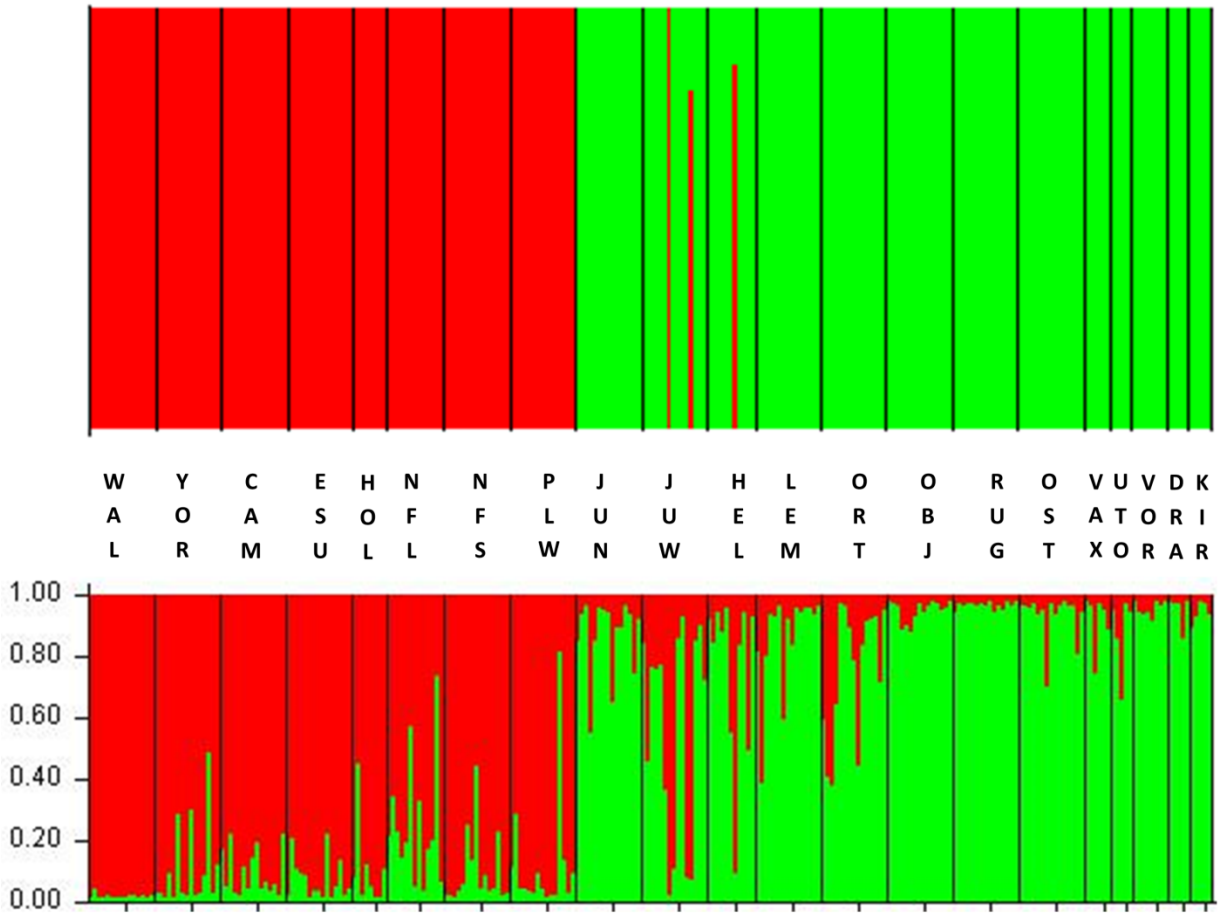


Figure 20. Admixture (K=2)

Result of a BAPS admixture analysis based on mixture clustering for K=2 (top) and a STRUCTURE admixture analysis for K=2 (bottom). Vertical bars represent individuals. Colors represent proportional membership to one of two genetically distinct clusters.

A BAPS admixture analysis based on a mixture analysis for K=3 resulted in four populations with significantly admixed individuals. The samples from Nord-Holland ("HOL") contained an individual that showed 38% membership to the boreal cluster and 1% membership to the continental cluster. One sample from Northern Frisia ("NFL") showed 60% membership to the continental cluster and 9 % membership to the boreal cluster. Samples from Western Jutland ("JUW") contained one individual that appears to be a migrant that is fully assigned to the Western cluster with 100% membership and one sample from Stockholm South ("UTO") was assigned to the continental cluster with 88 % and to the Atlantic cluster with 11%. Concordantly, STRUCTURE results showed comparatively strong proportional membership to the continental cluster for samples from Northern Frisia ("NFL") and the Stockholm populations ("VAX" and "UTO") and comparatively strong proportional membership to the Atlantic cluster for samples from Western Jutland ("JUW") while samples from Northern Jutland ("JUN") showed stronger proportional admixture with the boreal cluster than with the closely neighbouring Atlantic cluster (Figure 21).

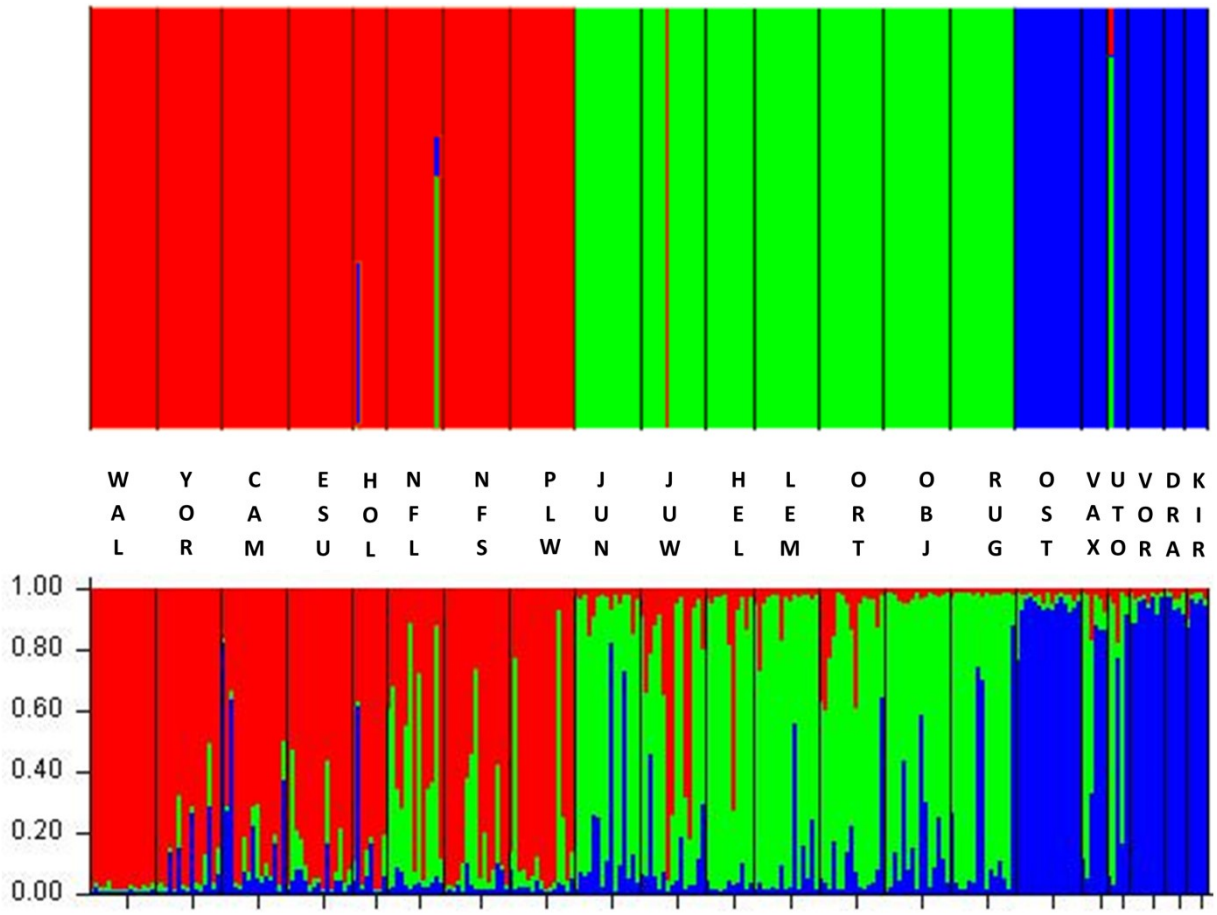


Figure 21. Admixture (K=3)

Result of a BAPS admixture analysis based on mixture clustering for K=3 (top) and a STRUCTURE admixture analysis for K=3 (bottom). Vertical bars represent individuals. Colors represent membership to one of three genetically distinct clusters.

3.1.2.9 - Testing for Isolation by Distance

A significant (positive) correlation between pairwise population genetic differentiation and geographic distance between populations (isolation by distance (IBD)) was detected for the whole dataset and eastern and Baltic Sea populations. Neither the western cluster populations nor the inland populations showed significant IBD. All significant results remained significant after adjusting for multiple comparisons with a sequential Bonferroni correction. Significance of results was not substantially altered when only considering population pairs showing geographic distances of at least 50 km (Table 9).

Table 9. Isolation by Distance

Results of Mantel test for correlation between geographic distances and pairwise genetic distances (R_{ST}) tested for the whole data set of population pairs (Total) and sub-sets of population pairs of inland populations (Inland), Baltic Sea populations (Baltic Sea), western cluster populations (West) and eastern cluster populations (East) considering all population pairs or only pairwise values for geographic distances of more than 50 km. The bottom line gives the number of population pairs for each category (n). P – values that remained significant after a sequential Bonferroni correction are printed in bold.

		Total	Inland	Baltic Sea	West	East
All population pairs	p	**≤0.001**	0.162	**0.002**	0.323	**0.003**
	r	0.586	0.132	0.441	0.044	0.582
	n	210	45	55	24	78
Population pairs > 50 km distance	p	**≤0.001**	0.263	**0.002**	0.314	**≤0.001**
	r	0.581	0.102	0.430	0.010	0.577
	n	208	44	54	23	77

3.1.2.10 - Calculating Residual R_{ST} values

Residual R_{ST} values were calculated for linear regressions of pairwise R_{ST} on geographic sampling distance. Pairwise R_{ST} values were calculated to rise with geographical distance as $R_{ST} = 0.0002*d+0.0297$ for the whole dataset, $R_{ST} = 0.0003*d+0.0114$ for the subset of population pairs within the eastern cluster and as $R_{ST} = 0.0002*d+0.0411$ for the Baltic Sea group (*d* being the geographical distance between populations in km).

Deviations of measured pairwise R_{ST} values (residual R_{ST} values) from the calculated R_{ST} values ranged from -0.474 to 0.614, showed no correlation with geographic distance and was further tested for correlation with average annual swan sightings and swan traffic counts (see 3.2.5.1). Residual R_{ST} values were not calculated for the Western cluster and the inland populations since Mantel tests results suggested no linear relationship between genetic and geographic distances for these groups (see 3.1.2.9).

3.1.2.11 - Analysis of mtDNA Haplotypes

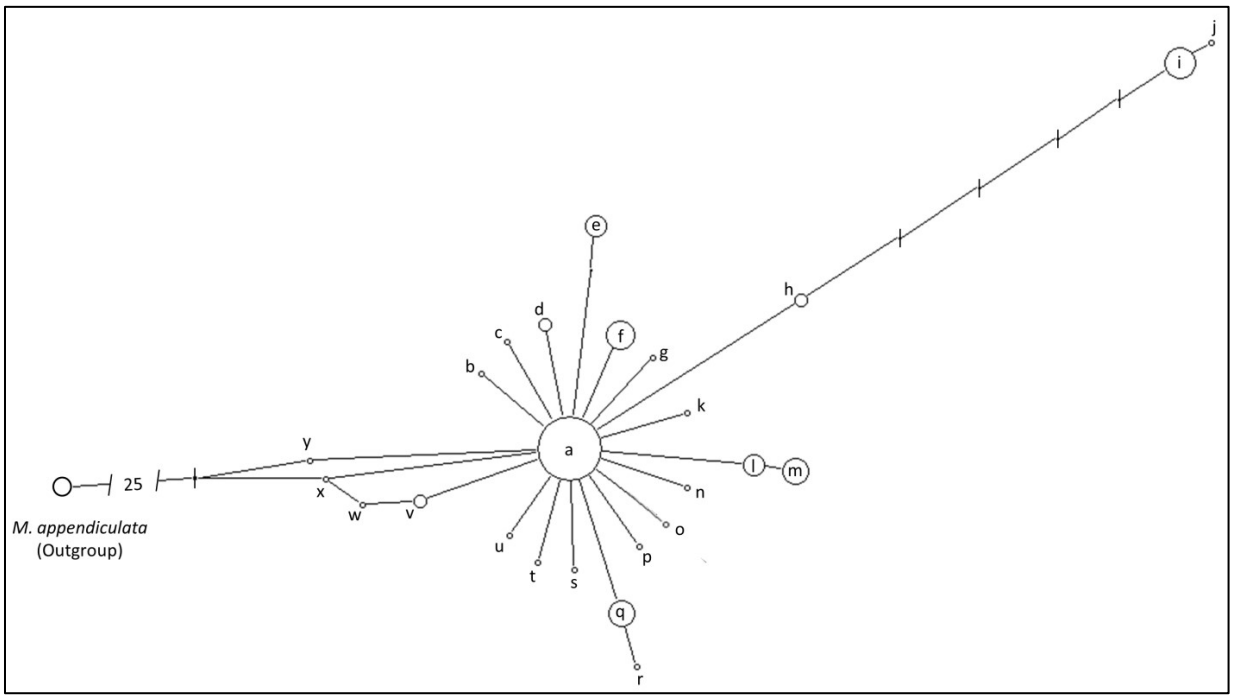


Figure 22. Haplotype Network for a 600bp Region of mtDNA (COI)

The figure represents the result of a median joining network (Bandelt et al. 1999) of 186 individuals of *Macroplea mutica* from 21 sampling sites and 5 individuals of *Macroplea appendiculata* (as outgroup) calculated with SPLITSTREE 4.12.8. (Huson & Bryant, 2006). Each haplotype is represented by a circle and labeled with a letter. Circle area is proportional to the number of samples within a given haplotype. Lines between haplotypes represent single mutational steps between alleles. Single mutational steps are assumed between adjacent haplotypes; the number in the line break indicates the number of mutational steps between *M. mutica* samples and outgroup samples (*M. appendiculata*).

List of samples within a given haplotype (numbers in parentheses indicate the proportion of samples from a given sampling site assigned to the respective haplotype).

H_a: Province of Rome, Italy (10/10) Wales, UK (9/10), Yorkshire, UK (2/2) Sussex, UK (10/10), North-Holland, The Netherlands (6/8), North Frisia South, Germany (4/10), Northern Jutland, Denmark (12/16), Western Jutland, Denmark (10/10), Plön Lakes South (9/10), Fehmarn West, Germany (11/16), Plön Lakes North, Germany (7/8), Rügen, Germany (3/10), Little Belt, Denmark (10/10), Stockholm South (3/5), Dragsfjärd, Finland (4/5), Northern Sweden (3/12), Kirkkonummi, Finland (4/5), Stockholm North, Sweden (4/6), China (7/8).

H_b: Lake Selent, Germany (1/8).

H_c: Stockholm South, Sweden (1/5).

H_d: Fehmarn West, Germany (2/16).

H_e: Lake Balaton, Hungary (4/4).

H_f: Northern Sweden (7/12).

H_g: Northern Sweden (1/12).

H_h: Fehmarn West, Germany (1/16), Dragsfjärd, Finland (1/5).

H_i: Sardinia, Italy (9/10).

H_j: Sardinia, Italy (1/10).

H_k: Stockholm North, Sweden (1/6).

H_l: Northern Jutland, Denmark (3/16), Plön Lakes South, Germany (1/10).

H_m: North Frisia South, Germany (6/10).

H_n: Stockholm South, Sweden (1/5).

H_o: Stockholm North, Sweden (1/6).

H_p: North-Holland, The Netherlands (1/8).

H_q: Rügen, Germany (6/10).

H_r: Rügen, Germany (1/10).

H_s: Wales, UK (1/10).

H_t: Fehmarn West, Germany (1/16).

H_u: Fehmarn West, Germany (1/16).

H_v: Northern Sweden (1/12), Finland East (1/5).

H_w: Northern Jutland, Denmark (1/16).

H_x: North-Holland, The Netherlands (1/8).

H_y: China (1/8).

In 186 sampled *M. mutica* specimens, 25 haplotypes were found, differing at one to six nucleotide sites (0.17-1.00% sequence divergence) (Figure 22). No indels were detected among the haplotypes and 26 of 600 (4.3%) sites examined were variable. Most haplotypes occurred in very low frequency.

The majority (71.9 %) of samples from Northern Europe and seven out of eight individuals from China showed one common haplotype (H_a).

Populations with fixed differences for single nucleotide polymorphisms (SNPs) compared to H_a were only found for a Sardinian population (ten individuals showing the same six substitutions) and four individuals from Lake Balaton, Hungary (one fixed substitution). Other than these two populations, only three populations showed a majority of individuals without the H_a haplotype: Six out of ten samples from southern North Frisia, Germany (H_m) showed the same substitutions at two positions. Six out of ten samples from Rügen, Germany (H_q) and seven out of twelve samples from Northern Sweden (H_f) showed shared haplotypes with one substitution respectively. Most haplotypes (19) showed only one nucleotide substitution in comparison to H_a, 14 haplotypes were only represented by single individuals.

Five outgroup samples of *Macroplea appendiculata* (one from Lake Selent (Germany) and four from Lake Naardermeer (Netherlands) showed the same haplotype with 4.5% sequence divergence to H_a (27 substitutions). Partial COI sequences for all detected haplotypes are presented in Appendix D.

3.1.3 - Bayesian Inference of Migration Rates with MIGRATE-N

The highest values for mutation-scaled migration rates (M) and effective (im-) migrants per generation (N_em) were calculated for migration from the continental cluster to the boreal cluster (M = 30; N_em = 0.915), followed by migration rates from the Atlantic cluster to the boreal cluster (M = 21; N_em = 0.641) and the Atlantic cluster to the continental cluster (M = 11.4; N_em = 0.564).

Calculated values were lower for migration from the continental cluster to the Atlantic cluster (M = 9; N_em = 0.191), from the boreal cluster to the continental cluster (M = 5; N_em = 0.248) and from the boreal cluster to the Atlantic cluster (M = 3.8 and N_em = 0.081). The highest mutation-scaled population size Θ (Theta) was inferred for the continental cluster (Θ = 0.198) followed by the boreal cluster (Θ = 0.122) and the Atlantic cluster (Θ = 0.085) (Table 10). The resulting effective population sizes (assuming a mutation rate of 10^{-4} per locus per generation

(Whittaker *et al.*, 2003)) were N_e = 212. 5 for the Atlantic cluster, N_e = 495.0 for the continental cluster and N_e = 305.0 for the boreal cluster.

Table 10. Posterior Distributions for a Bayesian Analysis with MIGRATE-N

The maximum posterior estimate (mode of the posterior distribution) and the borders of the 95 % credibility interval (2.5% and 97.5% quantile) are presented for the mutation-scaled effective population size Theta (Θ), the mutation-scaled immigration rate M and the effective numbers of immigrants per generation (N_em) Arrows symbolize the direction of migration between clusters.

Parameter	2,50 % Quantile	Mode	97,5 % Quantile
Θ Atlantic	0	0.085	0.217
Θ Continental	0	0.198	0.29
Θ Boreal	0	0.122	0.293
M Continental → Atlantic	5.4	9.0	17.8
M Boreal → Atlantic	1.0	3.8	11.4
M Atlantic →Continental	8.6	11.4	19.0
M Boreal → Continental	0.2	5.0	14.6
M Atlantic → Boreal	4.4	21.0	33.4
M Continental → Boreal	21.0	30.0	42.2
N_em Continental → Atlantic	0	0.191	0.964
N_em Boreal → Atlantic	0	0.081	0.618
N_em Atlantic →Continental	0	0.564	1.378
N_em Boreal → Continental	0	0.248	1.059
N_em Atlantic → Boreal	0	0.641	2.449
N_em Continental → Boreal	0	0.915	3.095

3.2 – GIS-Based Analyses

3.2.1 - Mapping Beetle Sampling Locations against Landscape Features

3.2.1.1 - Mapping Beetle Sampling Locations against Biogeographic Regions

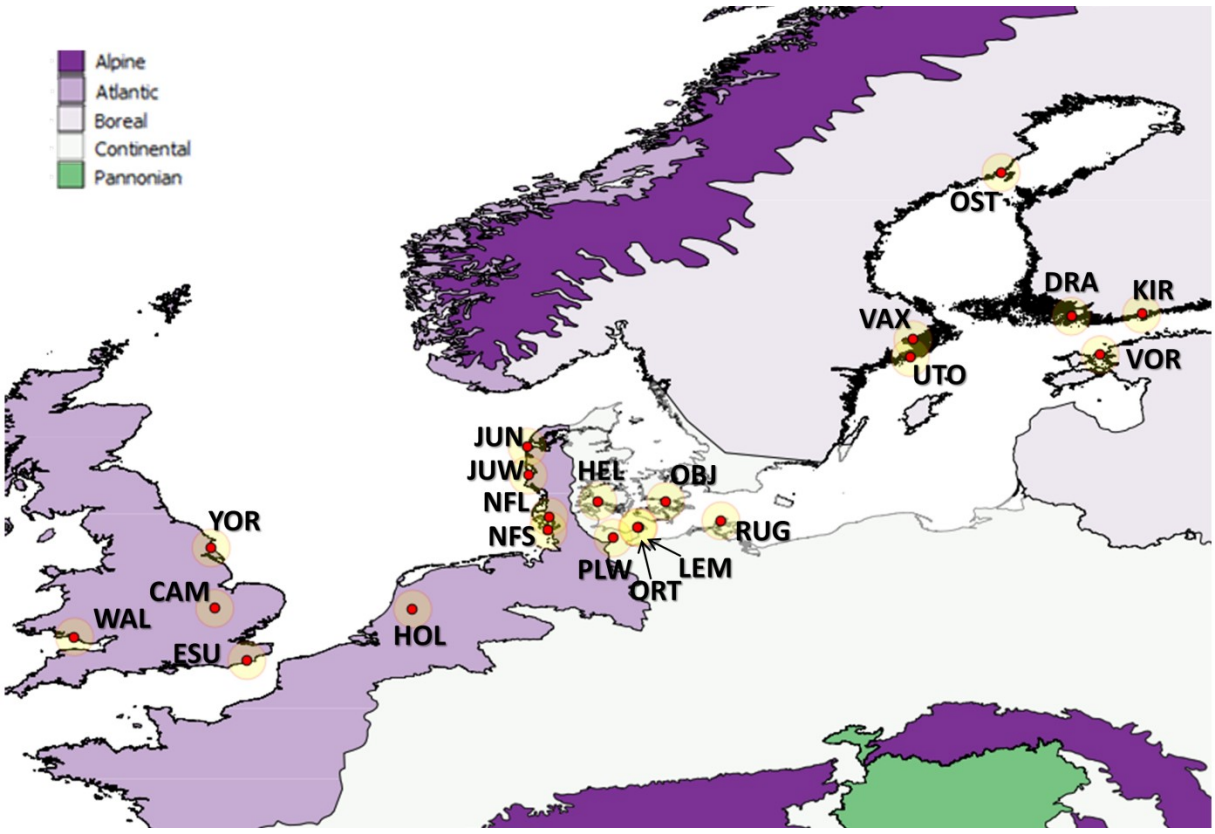


Figure 23. European Biogeographical Regions

Red dots represent sampled beetle populations. Colored areas represent different biogeographical regions.

Mapping the locations of 21 beetle sampling sites against a map of European biogeographic regions (Figure 23), the British, Dutch, Western German and Danish mainland populations ("WAL", "YOR", "CAM", "ESU", "HOL", "NFL", "NFS", "JUN" and "JUW") are shown to be located within the *Atlantic* biogeographic region. The Plön sampling site ("PLW") is situated just on the border between *Atlantic* and *continental* biogeographic region. The southwestern Baltic sampling sites (HEL", "OBJ", "LEM", "ORT" and "RUG") are located in the *continental* biogeographic region while the Swedish sampling sites ("VAX", "UTO" and "OST"), the Finnish and the Estonian populations ("DRA", "KIR" and "VOR") are located in the *boreal* biogeographic region. The map of biogeographic regions largely corresponds to the allocation of sampled populations to three clusters (see 3.1.2.7 and Figure 16).

3.2.1.2 - Mapping Beetle Sampling Locations against River Catchment Areas

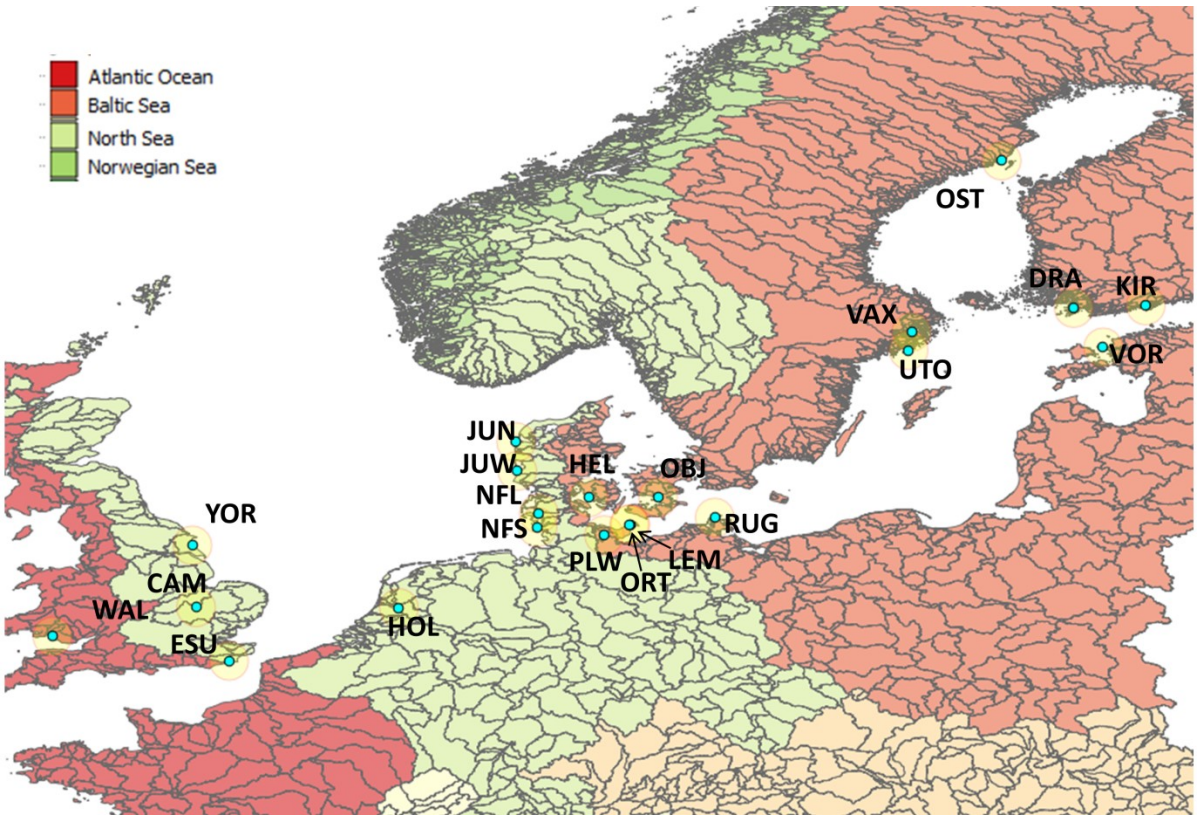


Figure 24. Main European River Catchments

Blue dots represent sampled beetle populations. Colored areas represent main European river catchment areas.

Mapping the locations of 21 beetle sampling sites against main European river catchments (Figure 24), the Welsh sampling site ("WAL") was found to be located in the Atlantic Ocean catchment area, the English, Dutch, Western German and Danish mainland populations ("YOR", "CAM", "ESU", "HOL", "NFL", "NFS", "JUN" and "JUW") are shown to be located within the North Sea catchment area. The Plön sampling site ("PLW") is situated just on the border between the North Sea catchment area and the Baltic Sea catchment area. All Baltic sampling sites ("HEL, "OBJ", "LEM", "ORT", "RUG", "VAX", "UTO", "OST", "DRA", "KIR" and "VOR") are located in the Baltic Sea catchment area. The mapped main river catchments largely correspond to the allocation of sampled populations to two clusters (see 3.1.2.7).

3.2.2 - Contributions of EURING Ringing Schemes to Obtained Mute Swan Sighting Data

A total of 264,430 records for geo-referenced sightings of mute swan were obtained from EURING. A total of 257,822 records dated from the period between 1974 and 2008 and 221,878 of these individual sighting records based on contributions from 18 EURING ringing schemes provided sufficient quality of information on sighting date and location to be used for mapping and further analyses. For a detailed list of contributing ringing schemes see Table 11.

Table 11. Contributions of Ringing Schemes

Contributions of individual ringing schemes to the mute swan data set obtained from EURING.

Ringing Centre	Country	Records	Remarks
Brussels	Belgium	254	
Praha	Czech Republic	2600	
Hiddensee	Germany	78019	
Radolfzell	Germany	1646	
Wilhelmshaven	Germany	6035	
Kopenhagen	Denmark	89	
Odense	Denmark	4	
Matsalu	Estonia	88	
Kaunas	Lithuania	8236	
Arnhem	The Netherlands	22938	
Oslo Museum	Norway	32	
Stavanger	Norway	896	
Helsinki Museum	Finland	195	
Bratislava	Slovakia	8	1960ies only
Moscow	Russia	236	
Goteborg	Sweden	92	1950ies only
Stockholm, Jager	Sweden	6	1960ies only
Olov Larsson	Sweden	4	
Stockholm, Ornis	Sweden	129	before 1960 only
Stockholm, Museum	Sweden	12795	
London (British Museum; Tring; Thetford)	UK	60287	
Gdansk; Varsovia	Poland	69841	*

*Contributions of the Polish ringing scheme included only intermediate sighting data (without data for ringing and last encounter) for each individual as the scheme feared potential conflict with another project that had purchased similar data.

3.2.3 - Mapping of Sighting Densities

Mapping mute swan sighting data with QGIS 2.0.1 (Quantum GIS Development Team, 2013) allowed for visualization of sighting data with different maps. Small dots as representation of individual sightings (Figure 25) showed sightings amass locally along landscape features such as the course of rivers (e.g. the rivers Ems, Weser and Elbe in Northwestern Germany (Figure 25, center)) and along coast lines. The presentation of sighting data with colored tiles (Figure 26) allowed for comparison of sighting densities (as many overlaid points representing sightings make it hard to assess sighting hot spots otherwise).

Maps of sighting data showed up to 21,436 mute swan sightings per square kilometer for the years 1974 to 2008 (an annual average of 612 sightings per km^2) and suggested main hot spots of swan sightings in western England, around the IJsselmeer in the Netherlands, the coast of Halland (southeastern Sweden), the Stockholm archipelago, eastern Germany (notably the Baltic Sea coast) and parts of Poland and Lithuania.

The number of (annual) mute swan sightings within a 50 km radius (an area of 7,850 km^2) around each beetle sampling site (Table 12) was inferred with QGIS (Quantum GIS Development Team, 2013). While the total number of sightings and the annual average was highest for the two sampling sites in the district of Stockholm (with more than 3000 recorded sightings each) the corrected annual average (see 2.4.3) was highest for the sampling site on the island of Rügen, followed by the Stockholm archipelago sites, the sites on the island of Fehmarn and in North-Holland. The fewest mute swan sightings were recorded for sampling sites at the coast of Estonia and Finland, the west coast of mainland Denmark (Jutland), and no sightings were recorded near the northernmost site in the dataset (west coast of the Bothnian Sea (Northern Sweden)). The average number of annual swan sightings for a pair of *M. mutica* sampling sites showed no significant correlation with geographic distance between sampling sites ($r = 0{,}194$; $p = 0.078$, Mantel test).

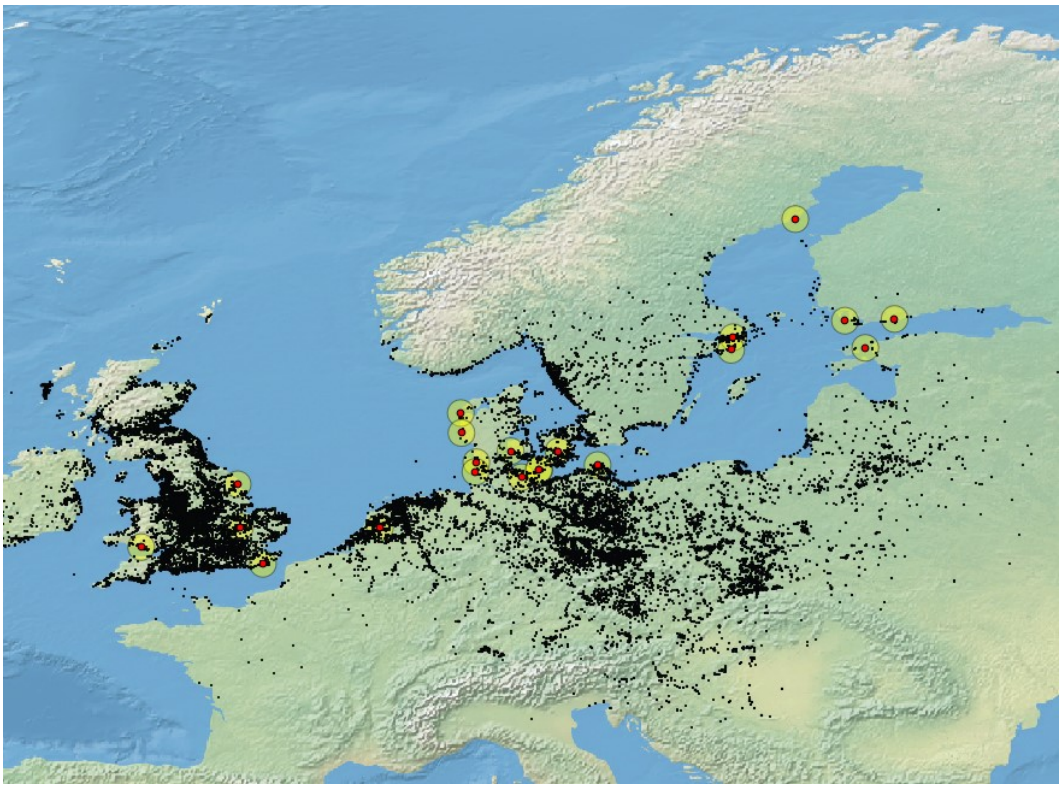


Figure 25. Mute Swan Sightings

Mute swan sightings in Northern Europe between 1974 and 2008 (according to the EURING data base) mapped with QGIS 2.0.1. Red dots represent sampled beetle populations. Black dots represent individual mute swan sightings.

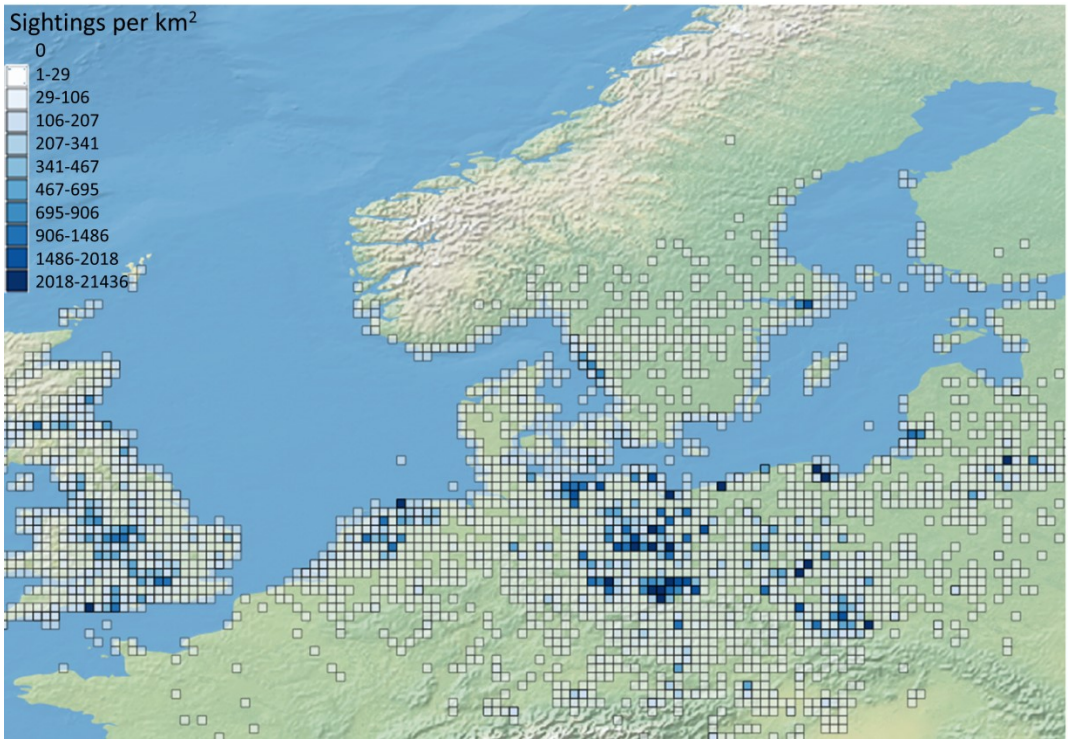


Figure 26. Mute Swan Sighting Densities

Mute swan sighting densities in Northern Europe between 1974 and 2008 (according to the EURING data base) mapped with QGIS 2.0.1. Each tile represents an area of 784 km^2.

Table 12. Swan Sightings at Beetle Sampling Sites

The total number and an annual average of individual swan sightings that dated from 1974 to 2008 mapped within a 50 km radius around a beetle sampling site (population) are shown. The corrected annual average considers only years with local abundances below the 95th percentile and above the 5th percentile for a given site.

Population	Swan Sightings 1974 - 2008	Annual Average	Corrected Annual Average
Rügen (RUG)	2033	58.1	97.8
Stockholm North (VAX)	3460	98.9	43.5
Stockholm South (UTO)	3098	88.5	39.1
Fehmarn West (ORT)	1964	56.1	30.4
Fehmarn East (LEM)	1964	56.1	30.2
North Holland (HOL)	1601	45.7	23.8
Plön Lakes (PLW)	1247	35.6	14.7
Sussex (ESU)	609	17.4	14.4
Sjaelland (OBJ)	1131	32.3	14.2
Cambridgeshire (CAM)	1692	48.3	5.8
Little Belt (HEL)	135	3.9	4
Yorkshire (YOR)	245	7	3.2
Wales (WAL)	137	3.9	2.3
North Frisia South (NFS)	129	3.7	1.6
North Frisia (NFL)	98	2.8	1.5
Finland West (DRA)	37	1.1	1.1
Estonia (VOR)	16	0.5	0.5
Northern Jutland (JUN)	15	0.4	0.1
Western Jutland (FT)	10	0.3	0.1
Finland East (KIR)	10	0.3	0.1
Northern Sweden (OST)	0	0	0

3.2.4 - Mute Swan Re-Sightings and Swan Movements

Geographic coordinate data for multiple sightings of the same individuals were used to work out which sites were connected by the movements / presence of individual swans. Lines connecting multiple sightings of the same individual (Figure 27) represent the shortest route between two sightings rather than the actual route travelled. But by considering dated spatial data for ringing and re-encounter (sighting- and re-sighting records) of individually marked mute swans, it was possible to infer the bearing of movements for given periods of time. Both total and relative swan traffic (see 2.4.3) showed highly significant negative correlation with geographic distance ($r = 0.190$; $P \leq 0.001$ and $r = 0.190$; $P \leq 0.001$ respectively, Mantel test).

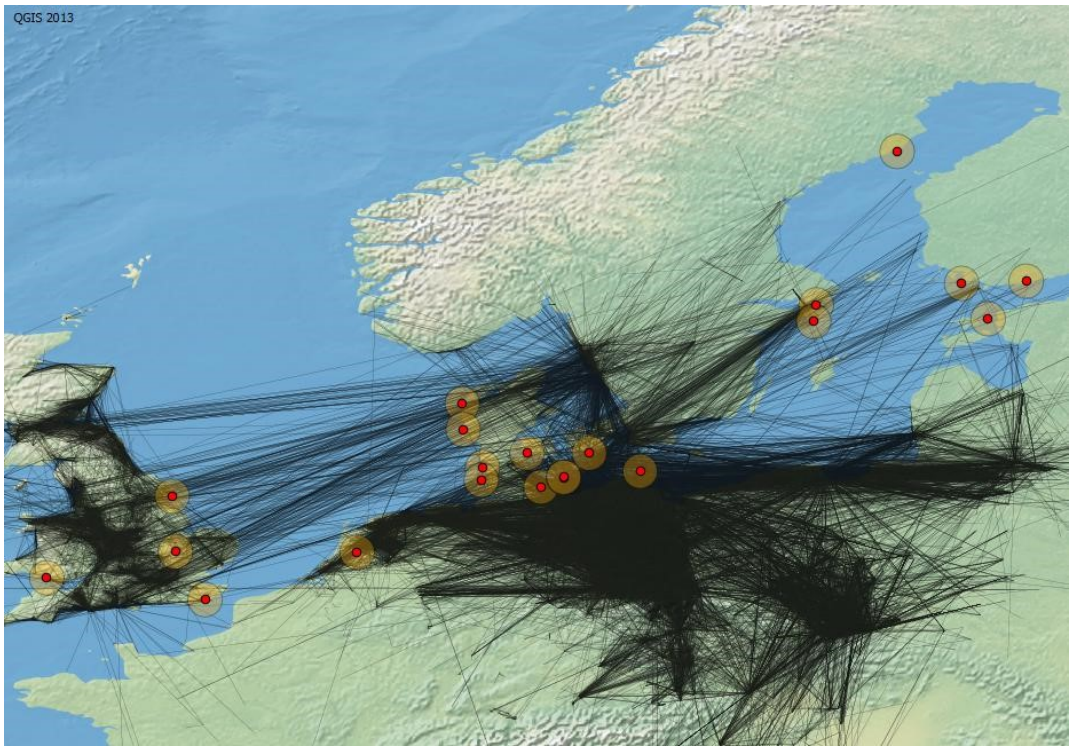


Figure 27. Mute Swan Movements

Mute swan re-sightings in Northern Europe between 1974 and 2008 (according to the EURING data base) mapped with QGIS 2.0.1. Red points represent sampled beetle populations. Black lines connect (re-) sighting locations of individual mute swans.

3.2.4.1 - Comparison of (Re-) Sightings during Summer and Winter

Sighting data (and re-sighting data) for the summer half-year (April-September) and the winter half-year (October-March) were compared (Table 13). The dataset contained almost twice as many re-sighting records for the winter half-years (22,403) as for the summer half-years (11,807). A 2-sample Z-test proved this to be a significant difference. The majority of re-sightings within six months showed moved distances of 10 km or less.

Movements in winter showed a more homogeneous directional distribution than movements in the summer half-year and short-distance movements (≤ 10 km) showed a more homogeneous directional distribution than long-distance movements (> 10 km). While the average geographical distance between two sightings of one individual was significantly greater in winter than in summer (regardless of considering or omitting data for stationary individuals), there was no significant difference between the average distance covered per day in summer and in winter (again, regardless of considering or omitting data for stationary individuals). Mute swans were recorded covering considerable distances within short periods of time. Sightings on the same day were as far apart as 120 km in summer and 301 km in

winter. Sightings on consecutive days showed covered distances of up to 355 km in summer and 426 km in winter.

Table 13. Seasonal Differences in Mute Swan Sightings

Comparison of sightings during the summer half-year (April-September) and the winter half-year (October-March). The table lists the number of mute swan sightings, the number of re-sightings of individuals, the average distance between re-sightings (average distance), the average distance individuals between re-sightings per day (average daily distance), the average distance between re-sightings when discarding data for stationary individuals (average moving distance), the average distance between re-sightings per day when discarding data for stationary individuals (average daily moving distance), the longest distance an individual swan moved between re-sightings on the same day (longest distance per day) and the longest distance between sightings on consecutive days (longest distance between two days). Differences between summer (April – September) and winter (October – March) were tested with a 2-sample Z-test. Significance levels of results are indicated as follows: (n.s.: $P > 0.05$, *: $P \leq 0.05$, ***: $P \leq 0.001$).

	Summer	Winter	Z-test
Sightings	85119	136759	-
Re-sightings	11807	22403	*
Average distance (km)	14.03	20.26	***
Average daily distance (km/day)	0.52	0.78	n.s.
Average moving distance_(km)	36.03	44.71	***
Average of moving distance covered per day_(km/day)	1.35	1.73	n.s.
Longest Distance per Day (km)	120	301	-
Longest Distance between 2 Days (km)	355	426	-

3.2.4.2 - Swan Traffic across Jutland

A two-proportion Z-test (Table 14) showed that the proportions of total mute swan sightings in mainland Denmark (Jutland) and mute swan movements intersecting with Jutland are significantly greater in winter (Figure 29) than in summer (Figure 28).

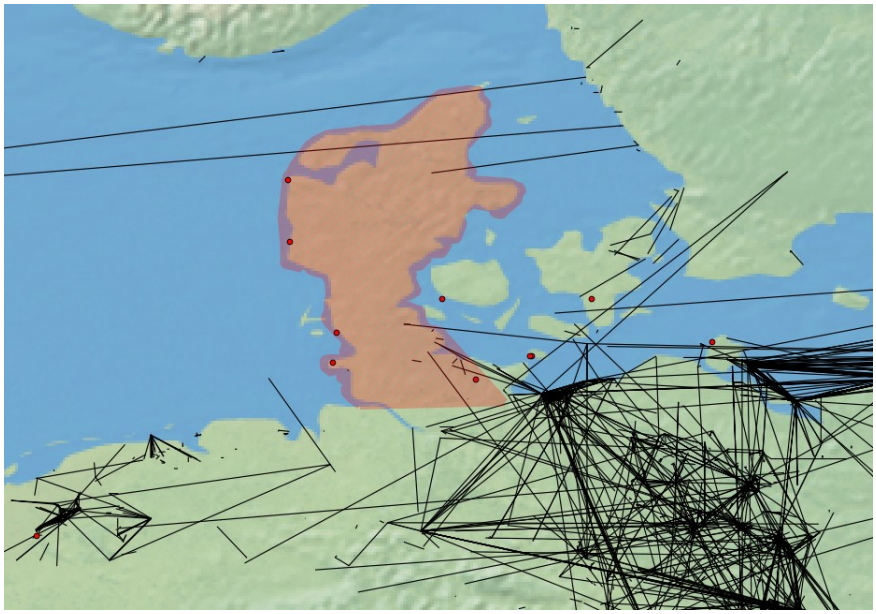


Figure 28. Mute Swan Traffic across Jutland - April to September

Black lines connect (re-) sighting locations of individual mute swans.
Red dots represent beetle sampling locations, Jutland is highlighted.

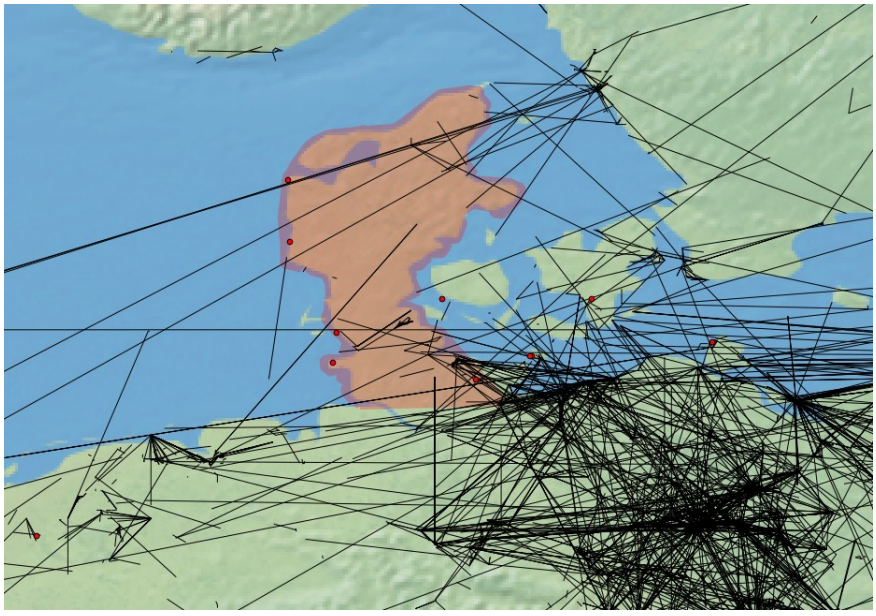


Figure 29. Mute Swan Traffic across Jutland - October to March

Black lines connect (re-) sighting locations of individual mute swans.
Red dots represent beetle sampling locations, Jutland is highlighted.

Table 14. Mute Swan Movements across Jutland

Recorded mute swan sightings and movements in Jutland (mainland Denmark) and a part of Northern Germany in the summer half-year and the winter half year in relation to the total number of recorded sightings and movements. The bottom line gives the probability for a significant difference according to results of a two-proportion Z-test.

	Total of mapped sightings	Sightings in Jutland (% of total)	Total of recorded movements	Recorded movements intersecting Jutland (% of total)
April - September	85119	229 (0.08%)	11807	10 (0.01%)
October - March	136759	735 (0.54%)	22403	64 (0.05%)
P-Value		≤ 0.001		≤ 0.001

3.2.4.3 - Directional Distribution of Total and Regional Mute Swan Movements

The directional distribution of mute swan movements showed regional and seasonal differences. In total 11,807 records were available for movements within six months between April and September and 22,403 for movements within six months between October and March. Based on data for seasonal directional distribution and distance of movements for the total Northern European dataset and eight regional subsets, mean and predominant directions of movements (Table 15) and distances travelled per cardinal direction were calculated and displayed as rose diagrams (Figure 34 and Figure 35 to Figure 38). Rose diagrams summarizing directional distribution of regional movements (covering at least 10 km) are pictured for summer (Figure 30) and winter (Figure 31). Rose diagrams summarizing directional distribution of *moved distances* (for covered distances greater than 10 km) are pictured for summer (Figure 32) and winter (Figure 33). In contrast to the analysis of directional distribution presented in Figure 30 and Figure 31, here each recorded movement was weighted by the distance covered. In the following, significant non-random distribution of movement directions with respect to the calculated mean angle (according to Rayleigh's Z equation) or further predicted directions of movement (according to a V test of circular uniformity) are referred to as *predominant* directions of movement.

The mean and predominant direction for all recorded movements of at least ten kilometers distance was northeastwards in the summer half-year and southwards in the winter half-year. When only considering movements over more than 100 km distance the mean bearing of movement shifted to northwestwards in summer but remained southwards in winter. Predominant bearing(s) of movements over more than 100 km were calculated as northwest-/ northwards in summer and southwards in winter.

Swans in Northern Great Britain (Scotland) (Figure 35a) showed significant predominance of northward movements during the summer half-year (April-September). The records also showed a noticeable tendency for eastward movements. While recorded eastward movements were less frequent, the distance covered is almost evenly split between east- and northward movements. Migratory individuals proved to travel between Great Britain and the European mainland (i.e. Denmark and even Sweden) (see Figure 27).

Sighting records for Norway and Northern Denmark (Figure 35b) were scarce and showed no significant predominance of directional distribution in neither summer nor winter. Distance travelled was almost exclusively distributed in an eastward direction in summer though (Figure 32).

Swans in Southern Sweden (Figure 36a) showed significant predominance of southwestward movements during the summer half-year (April-September) while the travelled distances were almost exclusively distributed eastwards in summer (Figure 32).

Movements in the Baltic States (Figure 36b) showed significant predominance of east-northeastward and westward movements during both the summer half-year (April-September) and winter half-year (the proportion of distance travelled westwards appears considerably greater though) (Figure 32 and Figure 33).

Swans in Southern Great Britain (Wales and England) (Figure 37a) showed significant predominance of northeast- and eastward movements during the summer half-year (April-September) but no significant predominance of directional distribution in winter. The proportion of travelled distance between April and September appeared greatest westwards (Figure 32).

Swans in the Netherlands and Northwestern Germany (Figure 37b) showed significant predominance of westward movements during the summer half-year (April-September) and southwards movements during the winter half-year (October to March).

Although records for movements in Denmark and Northeastern Germany (Figure 38a) were numerous in the dataset, they showed no significant predominance of directional distribution in neither summer nor winter.

Swans in Poland (Figure 38b) showed significant predominance of northwestward movements during the summer half-year (April-September) and eastward movements during the winter half-year (October to March). A greater proportion of travelled distance appeared to be distributed westwards in both the summer half-year (April-September) and winter half-year (October to March).

Table 15. Mean and Predominant Directions of Mute Swan Movements

Mean directions and predominant directions of regional mute swan movements.
The third column lists the calculated mean direction of movements (for movements during the summer half-year (April to September) and winter half-year (October to March) for different regions, and corresponding significance levels for a non-random distribution of movement directions with respect to the calculated mean angle according to Rayleigh's Z equation. The fourth column lists the number of mapped movements considered (n). The results of a V test of circular uniformity and respective significance levels with respect to predicted predominant routes of movement are listed in the fifth column (predominant direction). Significance levels of results are indicated as follows: (n.s. : $p > 0.05$, *: $p \leq 0.05$, **: $p \leq 0.01$, ***: $p \leq 0.001$).

	Season	Mean Direction		n	Predominant Direction	
Movements > 10 km						
Great Britain North	Summer	27.6°	n.s.	36	0°/360° *	
	Winter	224.2°	***	106	---	
Norway / Denmark North	Summer	97.6°	n.s.	3	---	
	Winter	257.2°	n.s.	31	---	
Sweden South	Summer	106.2°	n.s.	40	226° *	
	Winter	320°	n.s.	116	---	
Baltic States	Summer	243.6°	n.s.	32	67.5° *	269° *
	Winter	255.6°	n.s.	162	67.5° ***	255.6° ***
Great Britain South	Summer	60°	*	155	60° **	91° *
	Winter	199.7°	n.s.	352	---	
The Netherlands / Germany West	Summer	284.6°	***	131	284.6° ***	
	Winter	175.4°	***	155	175.4° ***	
Denmark / Germany Northeast	Summer	17.1°	n.s.	649	---	
	Winter	254.2°	n.s.	1147	---	
Poland	Summer	304.6°	***	335	304.6° ***	
	Winter	96.6°	***	820	96.6° ***	
Total	Summer	26.7°	*	1870	26.7° ***	
	Winter	184.8°	***	4814	184.8° ***	
Movements > 100 km						
Total	Summer	303.7°	***	352	303.7° ***	360° ***
	Winter	170.8°	***	977	170.8° ***	

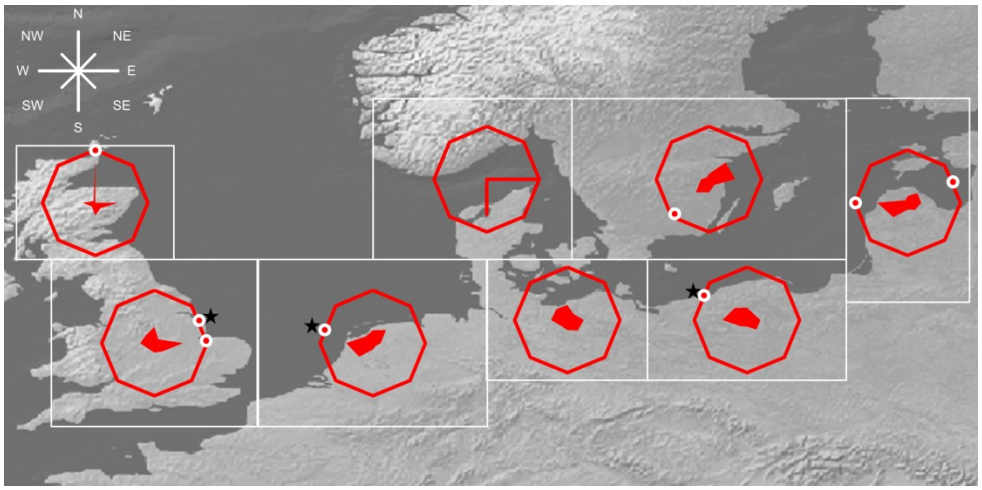


Figure 30. Directional Distribution of Mute Swan Movements (April to September)

Proportion of mapped swan movements (further than 10 km) within six month (April to September) are shown for the cardinal directions and for eight regions on the coast of Baltic and North Sea. The radius of the rose diagrams represent 50% of total mapped movements. White circles mark significantly predominant directions of movement (according to a V test of circular uniformity). Black stars mark significant non-random distribution of movement directions in respect to the mean direction of movement (according to Raleighs Z test).

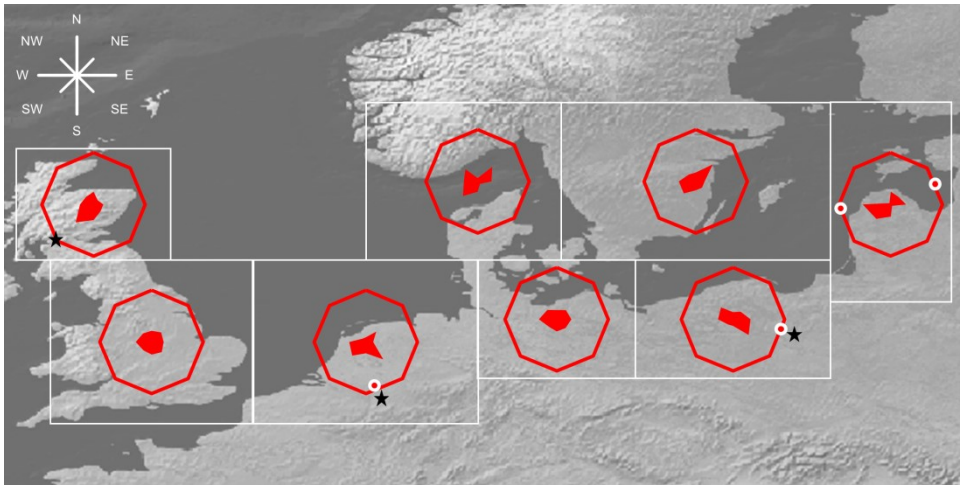


Figure 31. Directional Distribution of Mute Swan Movements between (October to March)

Proportion of mapped swan movements (further than 10 km) within six month (October and March) are shown for the cardinal directions and for eight regions on the coast of Baltic and North Sea. The radius of the rose diagrams represent 50% of total mapped movements. White circles mark significantly predominant directions of movement (according to a V test of circular uniformity). Black stars mark significant non-random distribution of movement directions in respect to the mean direction of movement (according to Raleighs Z test).

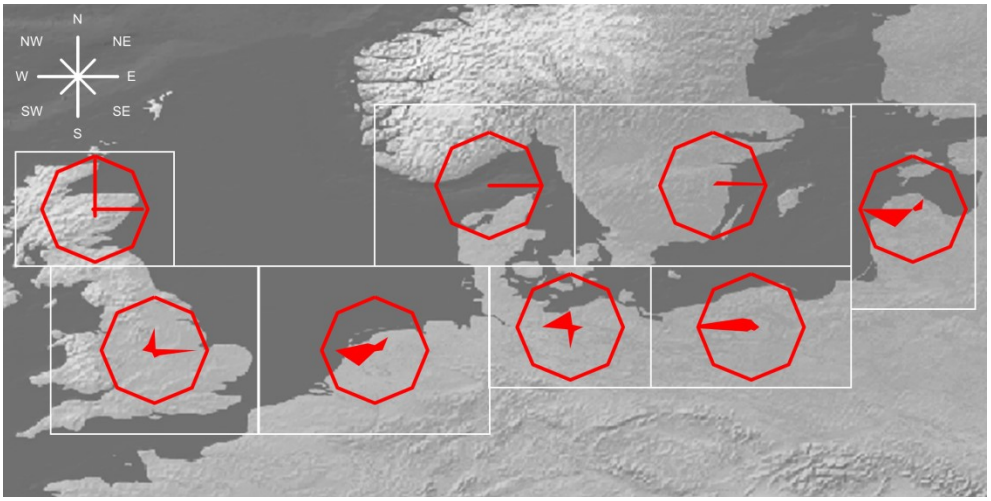


Figure 32. Directional Distribution of Distance Moved by Mute Swans between April and September

Proportion of distance covered by mapped swan movements (further than 10 km) within six month (April to September) are shown for the cardinal directions and for eight regions on the coast of Baltic and North Sea. The radius of the rose diagrams represent 50% of total moved distance.

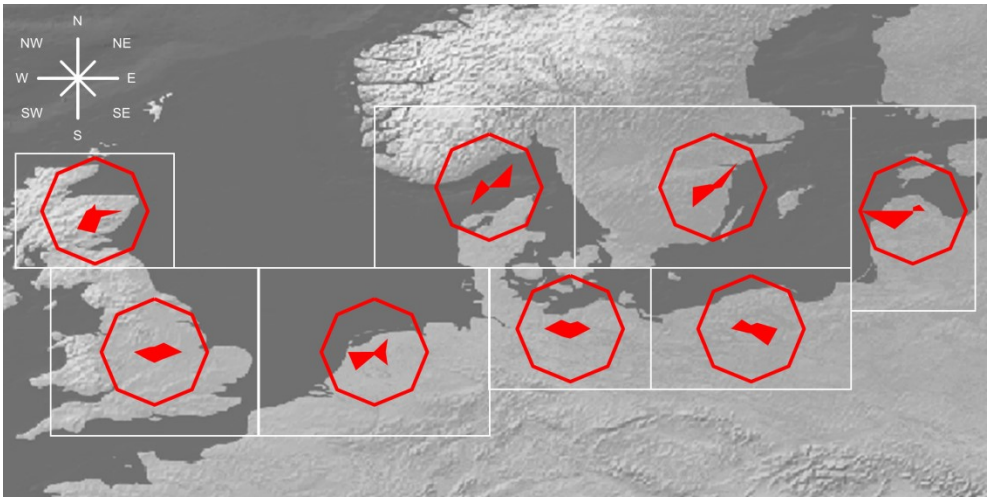


Figure 33. Directional Distribution of Distance Moved by Mute Swans between October and March

Proportion of distance covered by mapped swan movements (further than 10 km) within six month (between October and March) are shown for the cardinal directions and for eight regions on the coast of Baltic and North Sea. The radius of the rose diagrams represent 50% of total moved distance.

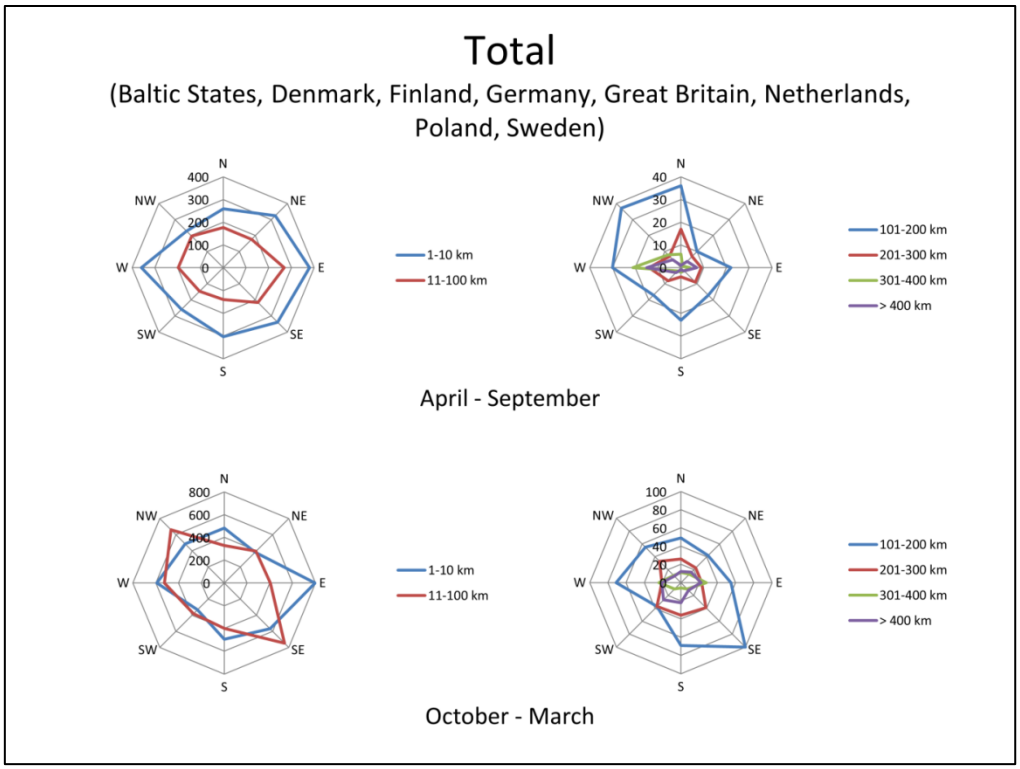


Figure 34. Directional Distribution of Northern European Mute Swan Movements

Movements covering distances of 1-10 km (blue line) and 11-100 km (red line) are presented on the left, movements of 101-200 km (blue line), 201-300 km (red line), 301-400 km (green line) and >400 km (purple line) are presented on the right. The top half shows movements between April and September, the bottom half shows movements between October and March. Number of movements is given on the y-axis.

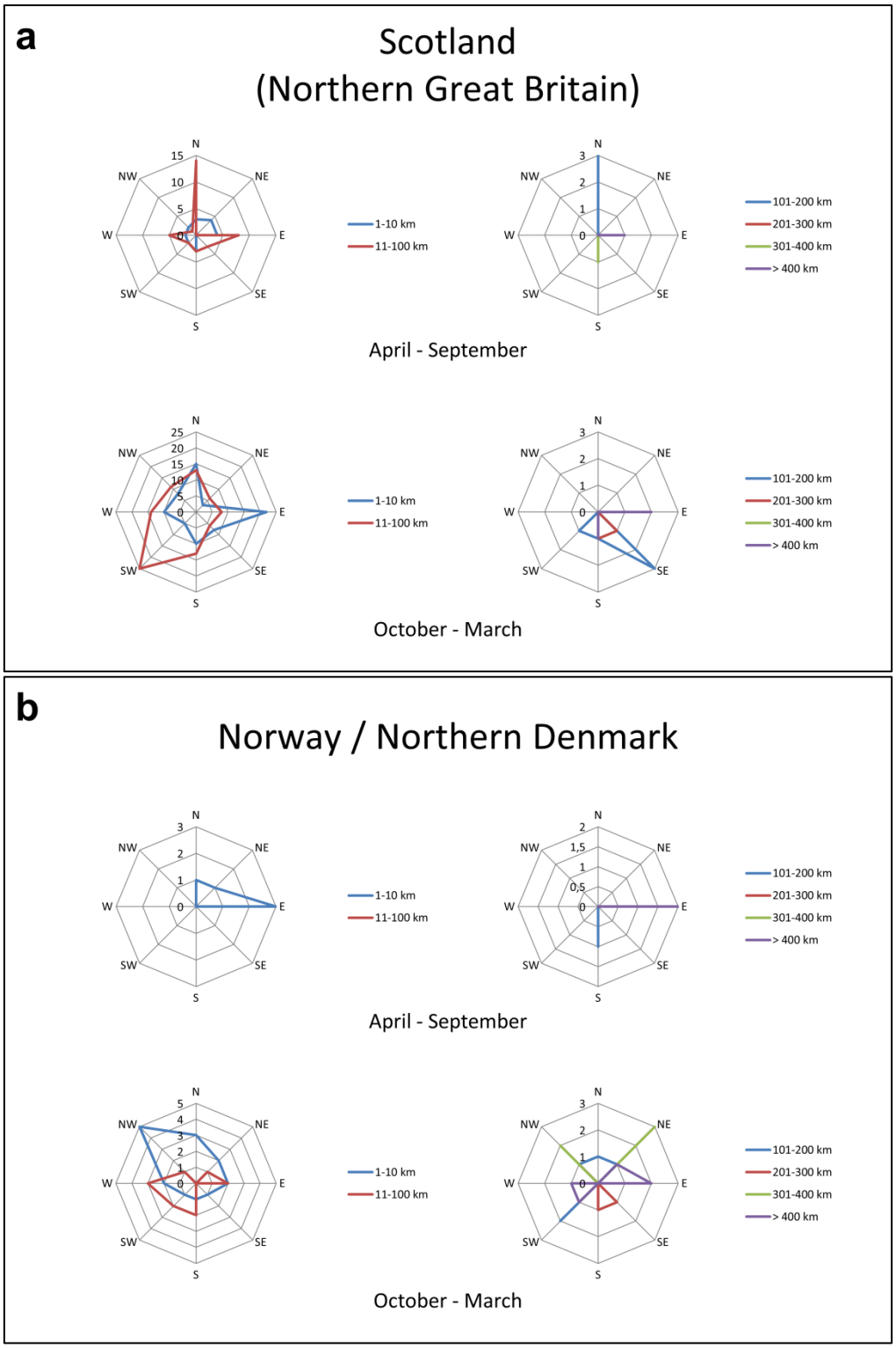


Figure 35. Mute Swan Movements in Scotland (a) and Norway and Northern Denmark (b).

Movements covering distances of 1-10 km (blue line) and 11-100 km (red line) are presented on the left, movements of 101-200 km (blue line), 201-300 km (red line), 301-400 km (green line) and >400 km (purple line) are presented on the right. The top half shows movements between April and September, the bottom half shows movements between October and March. Number of movements is given on the y-axis.

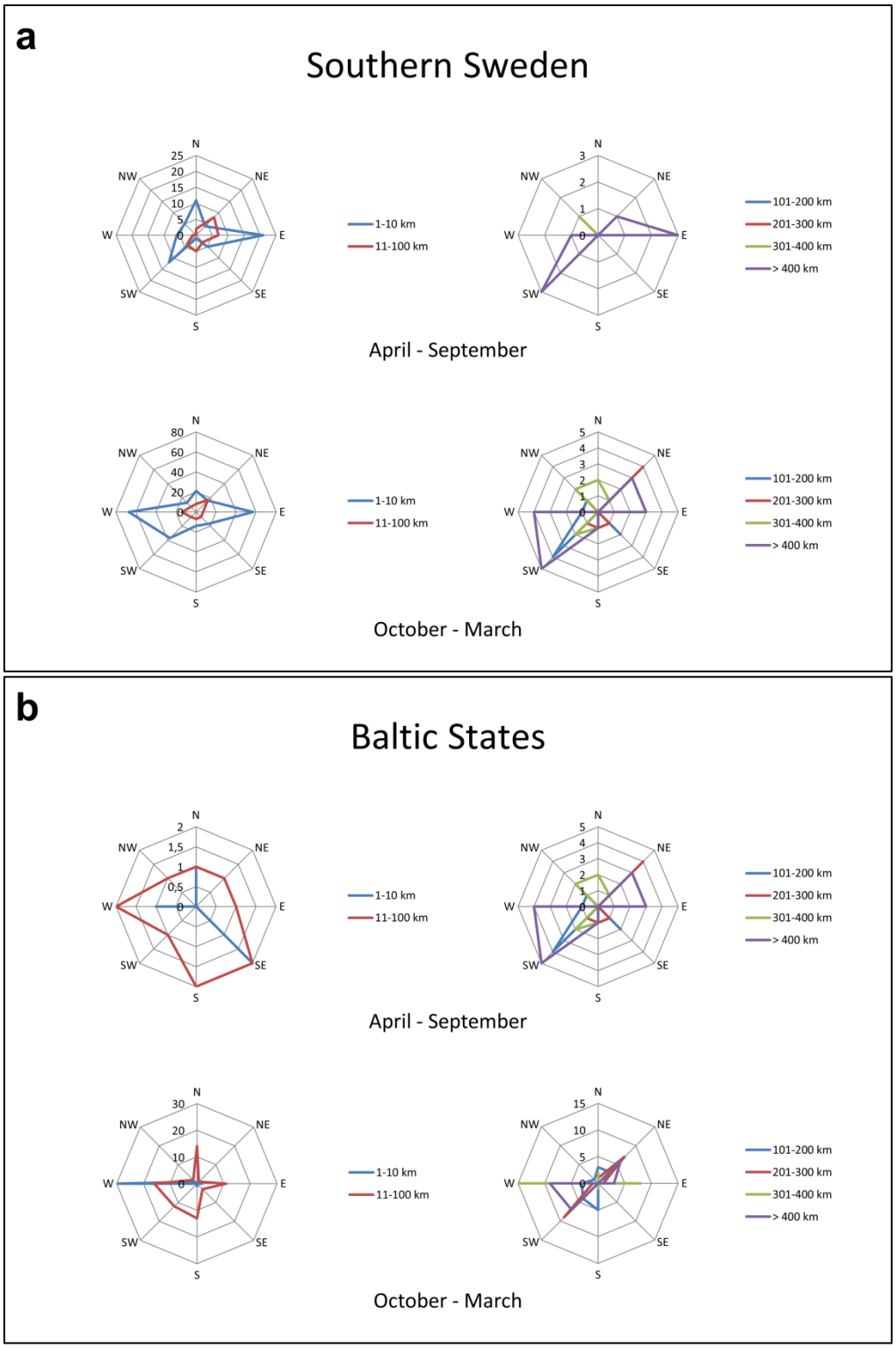


Figure 36. Mute Swan Movements in Southern Sweden (a) and the Baltic States (b).

Movements covering distances of 1-10 km (blue line) and 11-100 km (red line) are presented on the left, movements of 101-200 km (blue line), 201-300 km (red line), 301-400 km (green line) and >400 km (purple line) are presented on the right. The top half shows movements between April and September, the bottom half shows movements between October and March. Number of movements is given on the y-axis.

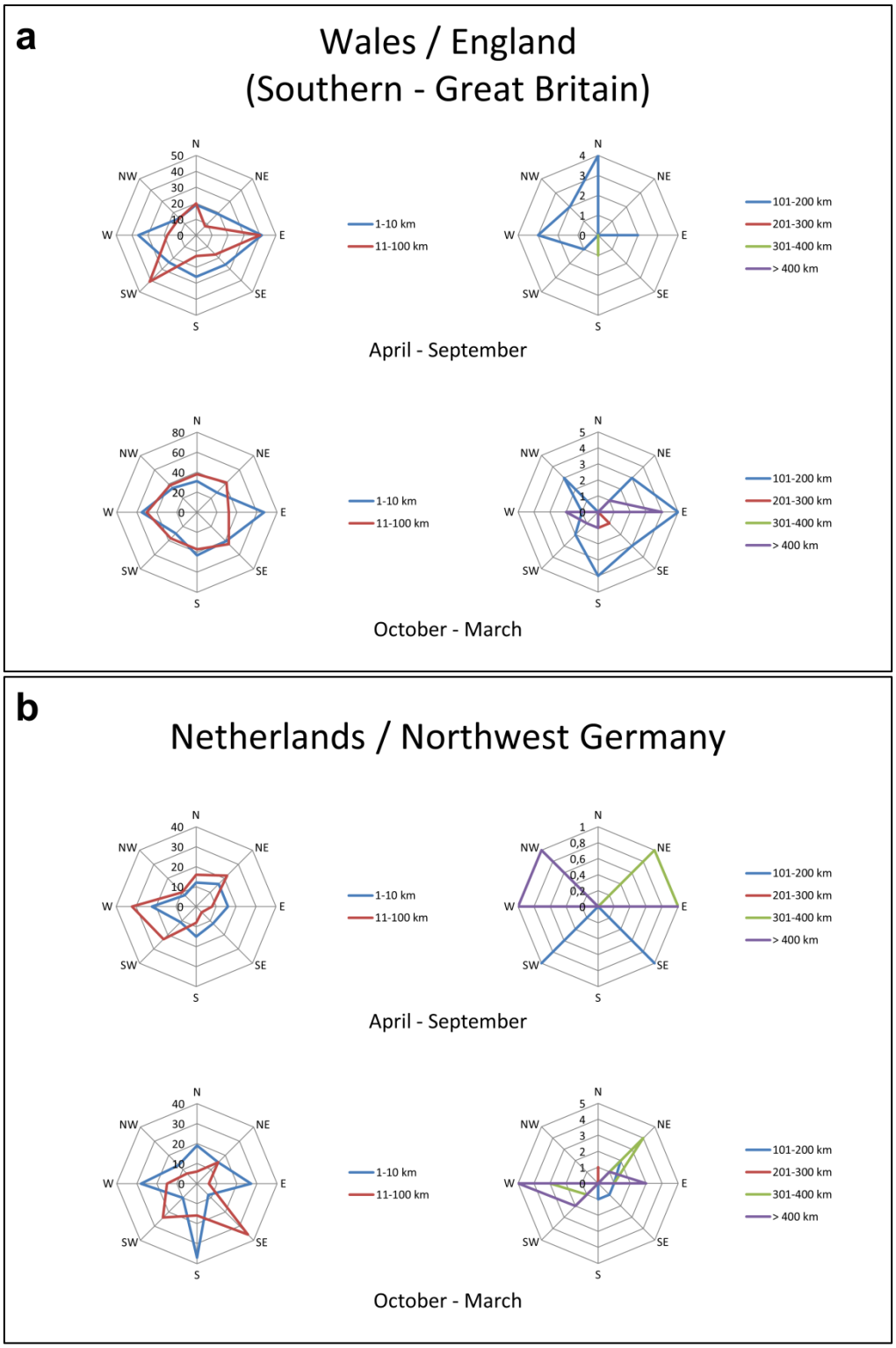


Figure 37. Mute Swan Movements in Wales and England (a) and Netherlands and Northwestern Germany (b).

Movements covering distances of 1-10 km (blue line) and 11-100 km (red line) are presented on the left, movements of 101-200 km (blue line), 201-300 km (red line), 301-400 km (green line) and >400 km (purple line) are presented on the right. The top half shows movements between April and September, the bottom half shows movements between October and March. Number of movements is given on the y-axis.

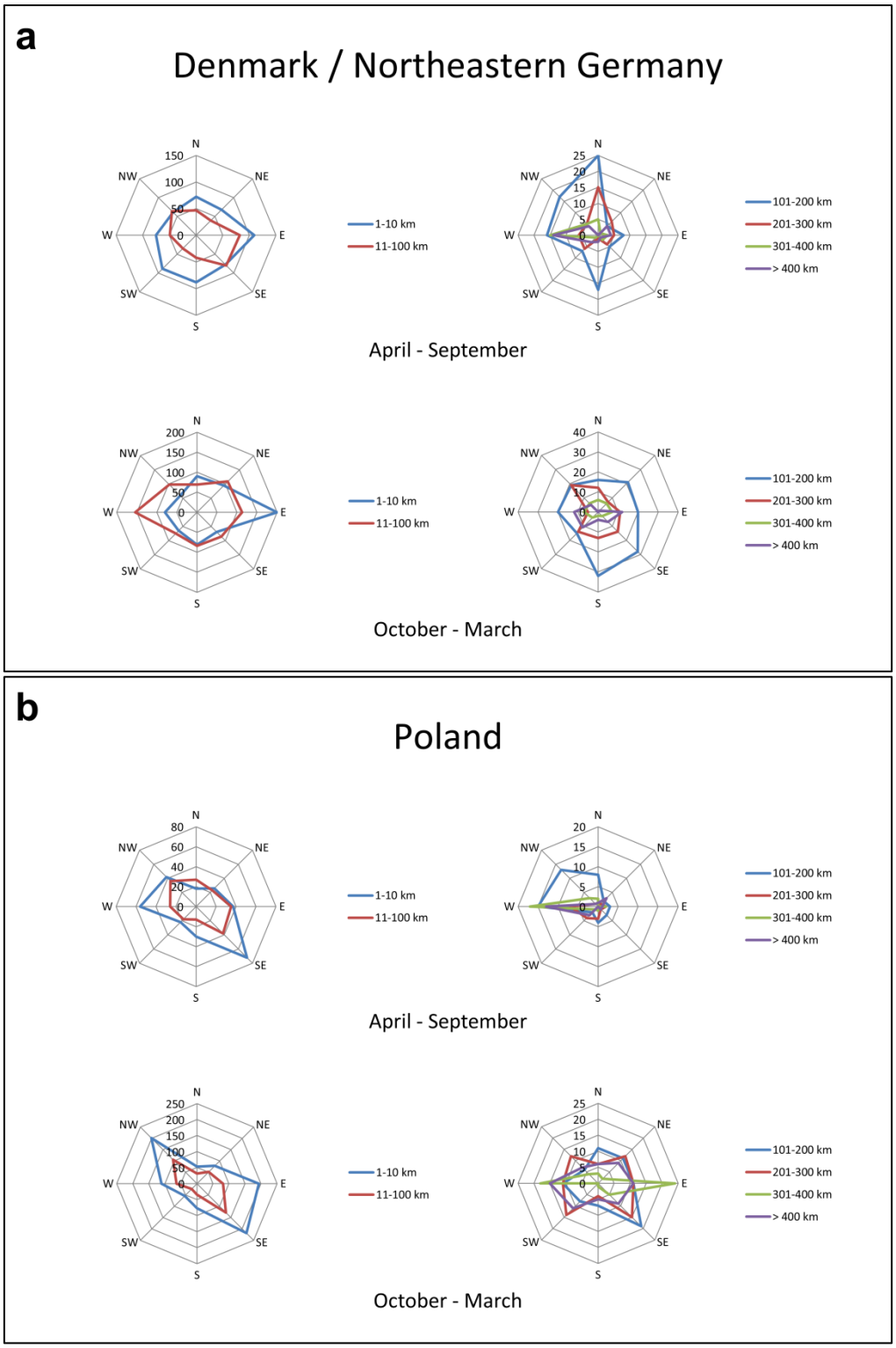


Figure 38. Mute Swan Movements in Denmark and Northeastern Germany (a) and Poland (b).

Movements covering distances of 1-10 km (blue line) and 11-100 km (red line) are presented on the left, movements of 101-200 km (blue line), 201-300 km (red line), 301-400 km (green line) and >400 km (purple line) are presented on the right. The top half shows movements between April and September, the bottom half shows movements between October and March. Number of movements is given on the y-axis.

3.2.5 - Testing for Correlation of Swan Movements and Abundances with Population Genetic Structure in Macroplea mutica

3.2.5.1 - Pairwise Genetic Differentiation in M. mutica and Swan Sighting Data

Table 16. Relationships between Genetic, Geographic and Swan Movement Matrices

Pairwise correlations between genetic distances (R_{ST} and Residual R_{ST}) and geographic distances between *M. mutica* populations or similarity in swan movements were calculated using Mantel tests. recorded swan Swan traffic between beetle sampling sites were tested as total of recorded movements (Swan Traffic) and swan traffic relative to recorded swan abundances (Relative Swan Traffic). Pairwise genetic distances were tested in the form of R_{ST}-values and residual R_{ST}-values based on linear regression on geographic distance. Residual R_{ST}-values were calculated for groups showing significant associations of genetic and geographic distances. P – values that remained significant after a sequential Bonferroni correction are printed in bold.

		Total	Inland	Baltic Sea	West	East	
R_{ST}	p	**≤0.001**	0.162	**0.002**	0.323	**0.003**	
	r	0.586	0.132	0.441	0.044	0.582	Geographic
Residual R_{ST}	p	0.713	-	0.461	-	0.347	Distance
	r	0.028	-	0.038	-	-0.047	
R_{ST}	p	**0.001**	**0.024**	**0.008**	0.068	**0.002**	
	r	-0.128	-0.271	-0.210	-0.181	-0.180	Swan
Residual R_{ST}	p	0.559	-	0.065	-	0.141	Traffic
	r	-0.018	-	-0.095	-	-0.026	
R_{ST}	p	**≤0.001**	**0.022**	**0.005**	0.072	**≤0.001**	
	r	-0.159	-0.191	-0.219	-0.195	-0.206	Relative
Residual R_{ST}	p	0.598	-	0.080	-	0.172	Swan Traffic
	r	-0.008	-	-0.087	-	-0.026	

A significant (positive) correlation between pairwise genetic distances and geographic distances (isolation by distance (IBD)) was detected in Mantel tests for the total data set and for the Baltic Sea- and eastern cluster groups, while neither the western cluster populations nor the inland populations showed significant IBD based on pairwise R_{ST} estimates (see 3.1.2.9). As expected and contrary to "raw" pairwise R_{ST}, residual R_{ST} (from linear regressions of pairwise R_{ST} on geographic distance) showed no significant correlation with geographic distances in Mantel tests (Table 16) and Spearman's rank correlation tests (Table 17).

Similarity matrices based on swan traffic between sampling sites showed significant correlation with pairwise R_{ST} values for all but the western cluster group in Mantel tests. Similarly, *relative* swan traffic between sites showed significant correlation with pairwise R_{ST} values for all but the western cluster group in Mantel tests. Spearman's rank correlation did

neither show significant correlation between R_{ST} and swan traffic (total or relative) for the inland sites nor for the western cluster but showed rather low p-values for the inland populations group ($r = -0.293$; $P = 0.050$ in case of R_{ST} versus swan traffic and $r = -0.293$; $P = 0.051$ for R_{ST} versus relative swan traffic).

Residual R_{ST} showed no significant correlation with total or relative swan traffic counts in Mantel tests. Spearman's rank correlation, however, initially showed significant correlations of both total and relative swan traffic counts with residual R_{ST} values in the Baltic Sea group, although this did not remain significant after a sequential Bonferroni correction in case of relative swan traffic. When gradually omitting sampling site pairs with geographic distances below a certain minimum, correlation of swan traffic with residual R_{ST} in the Baltic Sea group stayed significant for sampling site pairs with minimum geographic distances of up to 450 km ($n = 41$; $rho = -0.380$; $p = 0.014$; Spearman's rank correlation test). Total and relative swan traffic counts showed no significant correlations with residual R_{ST} for the total data set or the eastern cluster group.

The annual average of swan sightings within a 50 km radius around a beetle sampling site (averaged for population pairs) showed significant correlation with pairwise R_{ST} values and residual R_{ST} in the total data set, the eastern cluster and the Baltic Sea group in Spearman's rank correlation tests (Table 17). When gradually omitting sampling site pairs with geographic distances below a certain minimum, correlation of swan traffic with residual R_{ST} stayed significant for sampling site pairs with minimum geographic distances of up to 1,500 km ($n = 62$; $rho = -0.391$; $p \leq 0.001$; Spearman's rank correlation test) in the total data set and up to 400 km in the eastern cluster group ($n = 52$; $rho = -0.287$; $p = 0.039$; Spearman's rank correlation test). When only considering records for the summer half year (April to September), the annual average of swan sightings was also found to correlate significantly with residual R_{ST} in the total data set, the eastern cluster and the Baltic Sea group. Data for the winter half-year (March to October) only showed significant correlation with residual R_{ST} when testing the whole data set.

Neither the total annual swan average nor seasonal counts showed significant correlation of pairwise R_{ST} values in the western cluster or the inland populations group.

Table 17. Pairwise Correlations between Genetic Distances and Geographic Distances or Swan Abundances and Movements

Results of Spearman's rank correlation tests are shown for all populations (Total), the Baltic Sea sampling sites (Baltic) and the Eastern cluster sites (East) (a) and for the inland sites and western cluster groups (b). Pairwise population genetic data (pairwise R_{ST} or pairwise residual R_{ST}) were tested for correlation with total (Traffic) and relative swan traffic counts (Relative Swan Traffic), the annual average of swan sightings, averaged for a pair of beetle sampling sites for the whole year (Annual Swan Average), the half-year between April and September (Annual Swan Average Summer) and the half-year between October and March (Annual Swan Average Winter). The number of considered sample pairs (n), Spearman's correlation coefficient (rho) and determined probabilities (p) are given. Bold print marks result that remained significant after a sequential Bonferroni correction.

a	**Total**	**(n=210)**	**Baltic**	**(n=55)**	**East**	**(n=78)**
	rho	p	rho	p	rho	p
R_{ST} vs						
Annual Swan Average	-0.282	**≤ 0.001**	-0.400	**0.002**	-0.391	**≤ 0.001**
Residual R_{ST} vs						
Annual Swan Average	-0.245	**≤ 0.001**	-0.329	**0.014**	-0.295	**0.009**
Residual R_{ST} vs						
Annual Swan Average Summer	-0.244	**≤ 0.001**	-0.315	**0.019**	-0.277	**0.014**
Residual R_{ST} vs						
Annual Swan Average Winter	-0.275	**≤ 0.001**	0.037	0.787	-0.016	0.889
Residual R_{ST} vs						
Swan Traffic	-0.019	0.780	-0.331	**0.013**	-0.189	0.097
Residual R_{ST} vs						
Relative Swan Traffic	-0.030	0.670	-0.275	0.042	-0.144	0.208
Residual R_{ST} vs						
Geographic Distance	-0.065	0.346	0.005	0.972	-0.051	0.655

b	**Inland**	**(n=45)**	**West**	**(n=24)**
	rho	p	rho	p
R_{ST} vs				
Annual Swan Average	-0.105	0.492	0.174	0.416
R_{ST} vs				
Annual Swan Average Summer	-0.022	0.887	0.333	0.112
R_{ST} vs				
Annual Swan Average Winter	0.017	0.912	-0.279	0.187
R_{ST} vs				
Traffic	-0.293	0.050	-0.192	0.368
R_{ST} vs				
Relative Swan Traffic	-0.293	0.051	-0.191	0.373
R_{ST} vs				
Geographic Distance	0.094	0.538	0.201	0.347

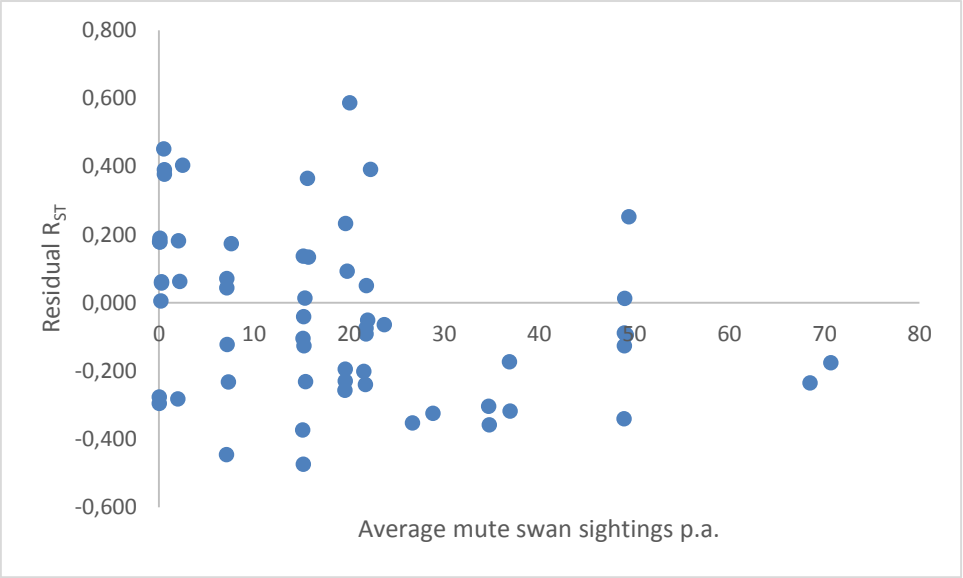


Figure 39. Residual Genetic Differentiation for Large Geographic Distances and Local Swan Abundances

Deviation from expected pairwise R_{ST} values (residual R_{ST}) (y-axis) and the annual number of swan sightings within a 50 km radius around sampling sites (averaged for population pairs) for population pairs in the eastern cluster with at least 400 km geographical distance (x-axis). Positive values on the y-axis imply greater genetic distance for a pair of (sub-) populations than would be expected given the geographic distance. Negative values on the y-axis imply less genetic distance for a pair of (sub-) populations than would be expected given the geographic distance.

3.2.5.2 - Swan Abundances and Breakdown of Isolation by Distance

When testing individual *M. mutica* populations for isolation by distance (IBD) towards all other populations (showing positive correlation between R_{ST} values and geographical distance in a Spearman's rank correlation test) 11 of 21 populations showed IBD after a Bonferroni correction. Five of eight populations in the western cluster and six of 13 in the eastern cluster showed IBD as well as seven of ten inland populations and four of eleven Baltic Sea populations (Table 18).

According to Fisher's exact test, IBD is significantly more frequent among populations with an average of less than ten annual swan sightings. This correlation was shown as significant for the whole data set, the Baltic populations and the eastern cluster but not for the western cluster or the inland populations group (Table 19).

Table 18. Isolation by Distance and Swan Sightings per Year

The table shows the results of a Spearman's rank correlation test (test for "isolation by distance") results are presented as probabilities (p-values) for correlation between pairwise R_{ST} values and geographical distance (IBD), results that proved to be significant after a sequential Bonferroni correction are printed in bold. Furthermore, the average annual number of mute swan sightings within a 50 km radius around each sampling site (for the years 1974-2008), the habitat type (Habitat) and membership to one of two genetic clusters (Cluster) are given.

Population (Abbreviation)	P (IBD)	Swans/Year	Habitat	Cluster
Wales (WAL)	**0.002**	2.3	Inland	Western
Yorkshire (YOR)	0.013	3.2	Inland	Western
Cambridgeshire (CAM)	**0.002**	5.8	Inland	Western
Sussex (ESU)	**≤ 0.001**	14.4	Inland	Western
North Holland (HOL)	0.009	23.8	Inland	Western
North Frisia (NFL)	**0.002**	1.5	Inland	Western
North Frisia South (NFS)	0.016	1.6	Inland	Western
Plön Lakes (PLW)	0.113	14.7	Inland	Western
Northern Jutland (JUN)	**≤ 0.001**	0.1	Inland	Eastern
Western Jutland (FT)	**≤ 0.001**	0.1	Inland	Eastern
Little Belt (HEL)	**0.003**	4	Baltic Sea	Eastern
Fehmarn East (LEM)	0.018	30.2	Baltic Sea	Eastern
Fehmarn West (ORT)	0.121	30.4	Baltic Sea	Eastern
Sjaeland (OBJ)	0.120	14.2	Baltic Sea	Eastern
Rügen (RUG)	0.069	97.8	Baltic Sea	Eastern
Norther Sweden (OST)	0.240	0	Baltic Sea	Eastern
Stockholm North (VAX)	0.018	43.5	Baltic Sea	Eastern
Stockholm South (UTO)	0.200	39.1	Baltic Sea	Eastern
Estonia (VOR)	**≤ 0.001**	0.5	Baltic Sea	Eastern
Finland West (DRA)	**≤ 0.001**	1.1	Baltic Sea	Eastern
Finland East (KIR)	**≤ 0.001**	0.1	Baltic Sea	Eastern

Table 19. Breakdown of Isolation by Distance (IBD)

Probabilities were calculated with a two-tailed Fisher's exact test of a 2x2 contingency table. Significant results imply that little-visited populations (with less than ten annual swan sightings on average) significantly more often show IBD toward the other populations. Bold print marks p-values that remained significant after a sequential Bonferroni correction.

Group	Probability
All Populations	**0.008**
Baltic Populations	**0.015**
Inland Populations	0.500
Eastern Cluster Populations	**0.005**
Western Cluster Populations	1.000

Table 20. Breakdown of Isolation by Distance - Mantel Test Results

For each group, populations with more than ten annual swan sightings on average (> 10 swan sightings p.a.) and less than ten annual swan sightings on average (< 10 swan sightings p.a.) were tested for isolation by distance, based on pairwise R_{ST} values and geographic distances. Probabilities were calculated with Mantel tests. The correlation coefficients (r), probabilities (p) and number of considered population pairs (n) are given. Bold print marks p-values that remained significant after a sequential Bonferroni correction.

	All populations		Western cluster		Eastern cluster	
	r	p	r	p	r	p
< 10 swan sightings p.a.	0.630	**≤ 0.001**	0.182	0.402	0.678	**0.022**
	n = 66		n = 15		n = 15	
> 10 swan sightings p.a.	0.442	0.066	0.998	1.000	0.196	0.858
	n = 36		n = 3		n = 15	

Mantel tests showed no significant spatial structure in pairwise genetic distances between populations with an average of more than 10 mute swan sightings per year. Populations with less than ten sightings per year did show significant isolation by distance in case of the total data set and the eastern cluster but not for the western cluster (Table 20). Considering only population pairs with geographic distances of more than 50 km did not alter the significance of the results.

Table 21. Residual Pairwise R_{ST} and Average Local Swan Abundances for Great Geographic Distances

Test for correlation between deviation from expected population differentiation (residual R_{ST}) and swan sighting densities. The table shows results of Spearman's rank correlation tests. The deviation from expected pairwise R_{ST} values for a given geographical distance between populations was tested for correlation with average annual swan sightings (averaged for each population pair). Considered were only population pairs with more than 300 km (>300 km) and more than 1500 km (>1500 km) geographical distance. Bold print marks p-values that remained significant after adjustment by a sequential Bonferroni correction.

	Spearman's rho	p	n
> 300 km	-0.249	**≤ 0.001**	175
>1,500 km	-0.391	**0.002**	62

3.3 – Feeding Trials and Simulation of Gut Passage

3.3.1 - Simulation of Gut Passage

No beetles hatched from 30 tested cocoons within four weeks after the treatment while beetles hatched from 24 (80%) of 30 untreated cocoons in the control group within the same time period. All treated cocoons showed signs of mechanical damage and resulting inflow of artificial digestive juices. 164 eggs were counted in the tested plant material after the treatment. Five larvae (3%) hatched in the three weeks following the treatment. Twelve larvae (40%) hatched from 30 untreated eggs in the control group within the same time period.

3.3.2 - Feeding Trials with Mallards

Spot checks of the number of beetle eggs within the plant material suggested an average of 1000 eggs ingested by the ducks per trial (i.e. 500 per duck). For each trial a total of approximately 300 ml faecal matter was collected from both ducks. Eggs were found in collected faeces after three out of four feeding trials (Table 22). All retrieved eggs were found attached to fragments of leaf sheath (Figure 40b) and / or stem in faecal lumps consisting of tightly clumped, comparatively little digested plant material defecated two to eight hours after feeding. Out of the approximately 4000 eggs fed to the ducks in total, 40 eggs (1%) could later be found in faecal samples. Out of 40 retrieved eggs 5 (12.5%) had been defecated within 2 hours after feeding, 29 eggs (72.5%) were found after retention times of 2-5 hours and 6 eggs (15%) were retrieved after retention times of 5-8 hours. Survival rates of retrieved eggs after gut passage ranged from 0% (second trial) to 60% (3 out of 5 retrieved after the first trial). The average survival rate of *retrieved* eggs after gut passage was 20%. Larvae (Figure 40c) hatched from 8 retrieved eggs (approx. 0.2 % of all eggs fed to the ducks during the trials). Twelve larvae (24%) hatched out of 50 eggs in an untreated control group kept under the same conditions as the retrieved eggs after gut passage.

Table 22. Results of Feeding Trials with Mallards

Number of plant fragments containing M. mutica eggs retrieved from mallard faeces, total number of retrieved eggs and number of larvae hatched from retrieved eggs.

	Retrieved 0-2 hours after feeding			Retrieved 2-5 hours after feeding			Retrieved 5-8 hours after feeding			Total		
Trial	Fragments	Eggs	Survival	Fragments	Eggs	Survival	Fragments	Eggs	Survival	Fragments	Eggs	Survival
1	0	0	-	1	3	2 (66%)	2	2	1 (50%)	3	5	3 (60%)
2	1	5	0 (0%)	4	8	0 (0%)	0	0	0 (0%)	5	13	0 (0%)
3	0	0	-	4	18	4 (22%)	1	4	1 (25%)	5	22	5 (23%)
4	0	0	-	0	0	-	0	0	-	0	0	0 (-)
Total	1	5	0 (0%)	9	29	6 (21%)	3	6	2 (33%)	13	40	8 (20%)
Control	-	-	-	-	-	-	-	-	-	5	51	12 (24%)

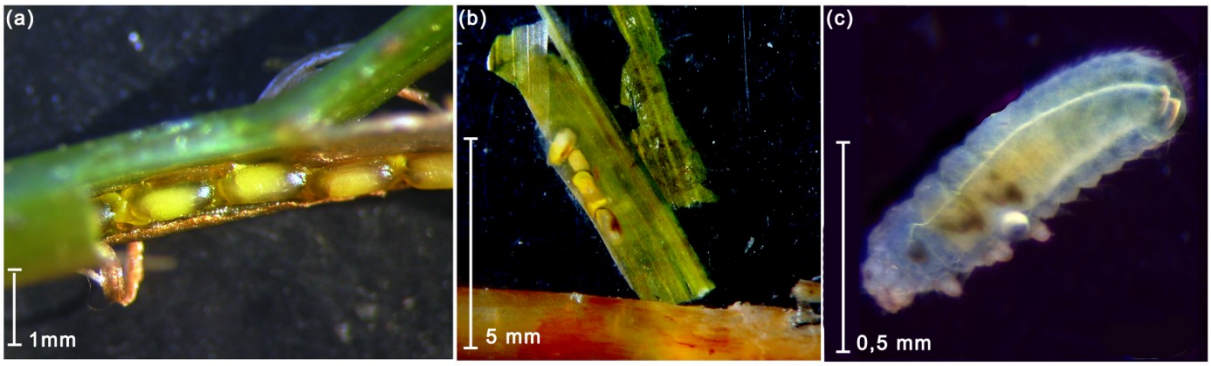


Figure 40. Eggs and Larva of *M. mutica*

(a) Row of *Macroplea mutica* eggs, laid under pondweed leaf sheath; (b) eggs of *M. mutica* attached to a leaf sheath fragment retrieved from duck faeces; (c) *M. mutica* larva that hatched after gut passage (bottom left, head capsule; top right, abdominal hooks).

4 - Discussion

The present study was dedicated to the investigation of the population genetic structure of *Macroplea mutica* and the examination of the possibility of transport by waterbirds. The following sections provide discussions on a) the development and characterization of microsatellite markers for *M. mutica*, b) the results of a population genetic analysis of *M. mutica,* in respect to proposed postglacial colonization history and landscape features acting as potential barriers to gene flow and c) implications of genetic and experimental evidence for the possibility of waterbird-mediated dispersal of *M. mutica*.

4.1 - Isolation of Microsatellites

At the onset of the present study, no microsatellite markers were available for *Macroplea mutica*. The isolation of microsatellites was therefore a fundamental first step towards the population genetic analysis.

The *de novo* isolation of microsatellite regions usually involves construction of a genomic library that is enriched for microsatellites. Squirrell *et al.* (2003) list the proportion of sequenced clones from an enriched library that has to be omitted on average during different steps of microsatellite isolation (attrition rate). A comparison showed relatively high attrition rates for the development of microsatellites for *M. mutica*. Isolation of microsatellites was based on an enriched library that contained a high proportion (95%) of contigs without microsatellites or contigs representing duplicate or chimeric sequences. Attrition for this step of the isolation process was therefore almost three times higher than the average reported by Squirrell *et al.* (2003) (36% attrition). Of 3182 unique SSRs only 21.5% contained sequences suitable for primer design (78.5% attrition). Average attrition for this step is 46 % according to Squirrell *et al.* (2003). Of 58 Loci for which primers were designed amplification of 89.7 % failed to produce interpretable polymorphic loci (the average is 50%).

The great redundancy of sequences in the enriched library might be result of relatively few probe motifs used in the enrichment procedure. The poor yield in the last step (monomorphic loci, no amplification products or multiple un-interpretable chromatogram peaks) might at least partially be the result of the screening procedure. Separation of amplification products on a 2.5 % agarose gel could be insufficient for clearly showing polymorphisms when alleles do not show obvious size differences. While this might in some cases save the expense of continuing with monomorphic loci, some loci might be prematurely omitted as monomorphic.

The screening procedure might have favored the development of highly polymorphic microsatellite markers though, as the average allele count per locus was high (17).
With the still increasing affordability of next generation sequencing technologies, construction of microsatellite libraries directly from sequencing reads might have rendered pre-sequencing enrichment procedures obsolete by now (Silva *et al.*, 2013). The procedure might then amongst other things benefit from the wider spectrum of available motifs since there is no attrition by pre-selection of probe motifs.

4.2 - Characterization and Validation of used Marker Set

Quality assessment and validation of the used marker set suggests that the developed microsatellite markers allow for credible estimates of population genetic differentiation among the sampled *M. mutica* subpopulations.
The analysis of six microsatellite loci (in combination with modest sample sizes per subpopulation) potentially provides limited reliability when representing (genome-wide) population differentiation. Nevertheless, the newly developed markers used in the present study have proven to be relatively unproblematic in amplification of loci. They showed neither problems in scoring (no signs for stuttering, large allele drop out or notable amounts of null alleles), nor linkage disequilibrium or problematic deviations from Hardy-Weinberg equilibrium. Furthermore, the marker set proved considerable resolution power according to simulations based on the number and distribution of detected alleles. Validation by comparison with results of an extensive AFLP study (Mende *et al.*, 2010) showed consistent estimates of genetic differentiation.

4.2.1 - Allele Scoring, Linkage Disequilibrium and Hardy-Weinberg Equilibrium

No evidence for scoring errors due to stuttering or evidence for large allele dropouts was found in an analysis with MICRO-CHECKER (Van Oosterhout *et al.*, 2004). The average frequency of potential null-alleles for a single locus was always below 8% and below 6% across loci which can be considered low enough not to risk considerable bias when estimating population differentiation without accounting for null alleles statistically (Oddou-Muratorio *et al.*, 2008). Ten out of 124 estimates showed significant single locus deviations from Hardy Weinberg equilibrium after a sequential Bonferroni correction. However, all these estimates showed homozygote excess and are likely to result from inbreeding rather than sampling

errors or subpopulation (sampling site) sub-structuring, considering the strong indication of inbreeding reported for *M. mutica* (Mende *et al.*, 2010). No interlocus associations remained significant after a sequential Bonferroni correction. Therefore there was no evidence for a non-random association of alleles over two or more loci.

4.2.2 - Resolution Power

The assessment of resolution power with POWSIM (Ryman & Palm, 2006) showed that for the used set of microsatellites, the detected allele frequencies and used sample sizes, F_{ST} values as low as 0.0129 were rightly detected in 100 % of all 10,000 simulation runs. Since all measured global and pairwise F_{ST} values were (mostly profoundly) higher, the power of the used marker set seems adequate for describing the population differentiation among the (rather strongly differentiated) *M. mutica* populations.

4.2.3 - Validation by Comparison with AFLP Data

Mantel's test and Spearman's rank correlation test showed that pairwise genetic distances (F_{ST}) were largely congruent with distances calculated by Mende *et al.* (2010) based on 251 AFLP loci (Mantel's test, 10,000 permutations: $r = 0.512$; $p \leq 0.001$; Spearman's rho: 0.512; $p \leq 0.001$). The strength of the observed relationship is consistent with results of other studies that compared AFLP- and microsatellite–based distances (using F_{ST} estimates for similar numbers of loci) for marker validation (Gaudeul *et al.*, 2004; Smee *et al.*, 2013).

4.3 - Population Structure and Genetic Differentiation

As expected due to the low mobility of *M. mutica*, detected genetic differentiation between subpopulations was high, suggesting low levels of gene flow (see 4.3.1). Significant patterns of isolation by distance and a clear signature of habitat structure (see 4.3.4.1 and 4.3.4.3) are further signs for low dispersal capacity (Miller *et al.*, 2002; Phillipsen & Lytle, 2013).
A clear hierarchical population structure (based on microsatellite data) with two large clusters divided by the main European water shed suggests that postglacial expansion in two hydrologically isolated habitats of different configuration (the large, continuous Baltic Sea habitat and highly structured inland habitats) might have played a major role in forming the

observed population structure (see 4.3.4.4). Low variability and geographically widespread mitochondrial DNA haplotypes suggest either a comparatively rapid colonization of Northern Europe (or even large parts of the Palearctic region) originating from an ancestral population with low genetic diversity or, alternatively, a mitochondrial selective sweep (see 4.3.3).

4.3.1 - Global and Pairwise Population Genetic Differentiation

Genetic (sub-) population differentiation was analyzed using the exact G test, estimates for fixation indices (using unbiased estimators of F_{ST} and the allele-size based analog measure R_{ST}) and classic measures of genetic distance (Nei's D_A). Allele size-based estimates of genetic differentiation proved to be preferable to allele identity-based estimates in case of the analyzed microsatellite data.

The exact G test showed highly significant global differentiation for the present dataset and only two population pairs for which differentiation was not significant. The exact G test is considered the most powerful test for differentiation of diploid organisms (Goudet *et al.*, 1996) especially when sampling is unbalanced. However, a test as powerful as the exact G-test might detect such fine differences in allele frequencies between subpopulations that significant differentiation might not necessarily be biologically meaningful (Hedrick, 1999; Balloux & Lugon-Moulin, 2002). The observation of highly significant pairwise population differentiation remains valid though and fits the assumptions of low potential for active dispersal in *M. mutica* (Mende *et al.*, 2010) and therefore suggest comparatively little gene flow between subpopulations.

Correspondently, the global average estimate (across all six loci and 21 (sub-) populations) for F_{ST} was 0.163, representing "great" differentiation (Hartl & Clark, 2007). The corresponding R_{ST} value was even higher (0.252). The inferred level of differentiation (global F_{ST} = 0.163) is greater than the estimate reported by Mende *et al.* (2010) based on AFLP markers (global F_{ST} = 0.135). These differences in F_{ST} estimates are not surprising since, despite many similarities, neither the geographic area nor the selection of considered sampling sites was identical between the present study and the study by Mende et al. (2010).

The ranking of population pairs based on pairwise F_{ST} and R_{ST} estimates showed little similarity, with the top and bottom ten percent of pairwise estimates showing only 23.8 % and 26.89 % identical composition (i.e. the same population pairs) respectively. Generally, calculated R_{ST}- and F_{ST} values for individual loci and across loci showed substantially different estimates for global and pairwise differentiation. Incongruence between F_{ST}

estimates and R_{ST} estimates based on microsatellite data is not uncommon (Lugon-Moulin *et al.*, 1999) and the reason for this might lie in the very different relative performance of both indices under certain circumstances (e.g. the high variance of R_{ST} estimates and the sensitivity of F_{ST} estimates to mutation (see 4.3.2)).

4.3.2 - Used Estimators of Genetic Differentiation

It is important to consider the relative performance of F_{ST} and R_{ST} when certain assumptions are met. For the present study estimates for both R_{ST} and F_{ST} are essentially unbiased, being based on a weighted analysis of variance, correcting for differences in sample sizes between sampling sites (sub-populations) and differences in variance between loci (Goodman, 1997). Both estimates can still be expected to differ in reliability though.

R_{ST}-estimates show a considerably higher associated variance than F_{ST}-estimates (Balloux & Lugon-Moulin, 2002). For this reason, F_{ST} estimates have been said to often represent differentiation more reliably than R_{ST}, especially when the numbers of loci and sampled individuals are small (Gaggiotti *et al.*, 1999). While R_{ST} estimates certainly tend to improve more strongly than F_{ST} with increasing numbers of loci and samples (Balloux & Goudet, 2002; Balloux & Lugon-Moulin, 2002), under certain circumstances (i.e. highly structured populations), R_{ST} should be preferred to F_{ST}, *especially* when the sample size is small (Balloux & Goudet, 2002).

Due to the high mutation rate of microsatellite loci, the main problem in analysis of microsatellite data with F_{ST} estimates is sensitivity to mutation (Balloux & Lugon-Moulin, 2002). R_{ST} estimates are independent of mutation as long the analyzed microsatellite loci perfectly follow a stepwise mutation model (SMM). The relative performance of R_{ST} over F_{ST} suffers whenever deviation from a strict SMM occurs. Since this is often the case, R_{ST} estimates are often likely to be an unknown function of migration and mutation (Balloux & Lugon-Moulin, 2002). When F_{ST}- and R_{ST} estimates show substantial differences it can therefore be crucial to assess the relative contribution of (step-wise) mutation versus drift and migration to differentiation in the studied population, to infer which statistic is likely to provide a better representation of population differentiation (Hardy *et al.*, 2003).

The results of a permutation test with SPAGEDI (Hardy & Vekemans, 2002) clearly suggested that R_{ST} shows a lower square mean error (the sum of the squared bias and variance) than allele identity-based estimates, due to the higher relative importance of (step-wise) mutation versus drift and migration for the observed population differentiation.

Therefore the results imply that R_{ST} performs better than F_{ST} and for this reason further analyses (i.e. tests for correlation between genetic differentiation and ecological variables) were based on R_{ST} estimates.

The use of different estimates of genetic distance was necessary due to the requirements of the respective analyses. F_{ST} estimates were presented mainly because F_{ST} is well-known and widely used as a measure of differentiation and therefore useful for the sake of comparison with other studies. For the present study, pairwise F_{ST} estimates were used to compare results to those of Mende *et al.* (2010). Nei's chord distance (D_A) was chosen for construction of a neighbor joining tree and a principal coordinate analysis, as it is generally held to be especially successful and reliable when estimating tree topologies based on microsatellite markers (Takezaki & Nei, 1996, 2008).

4.3.3 - Mitochondrial DNA Haplotypes

While microsatellite data showed clear phylogeographic structure (see e.g. the neighbor joining tree and results of the principal coordinate analysis (see 3.1.2.2; Figure 5 and Figure 6) the sequenced mtDNA showed almost no geographic structure and mostly weak divergence between haplotypes. The vast majority of sampled individuals showed one common haplotype (H_a) (see 3.1.2.1).

The apparent predominance of one haplotype is unlikely to be caused by PCR contamination, amplification of a pseudogene or insufficient sampling. PCR contamination can be ruled out due to the assiduous use of negative controls. Every PCR run contained at least one sample without template. Negative controls never showed a PCR product and therefore gave no reason to suspect contamination. The amplification of a non-functional pseudogene or numt (nuclear mitochondrial pseudogene) seems unlikely. Although it was not possible to check for a complete reading frame in the sequenced COI fragments, the sequences were free of indels and stop-codons (which would have suggested a pseudogene) and neither double bands on the PCR control gels nor notably ambiguous sequencing chromatograms suggested that primers amplified more than one locus. While sampling sizes per site were rather modest (and more samples per site could potentially have revealed more low frequency haplotypes), small sample sizes cannot explain why the vast majority of European samples (and the Chinese samples) share one common haplotype. Furthermore, the results are congruent with findings of Kölsch *et al.* (2006) who also report widespread mtDNA haplotypes and general lack of geographic structure for analyzed COI sequences of *M. mutica*.

Assuming a constant mutation rate for COI (molecular clock) allows for rough approximations of divergence times. Recent estimates for average substitution rates of insect COI assume average substitution rates of 3.54% My^{-1} (Papadopoulou *et al.*, 2010). Assuming the predominant haplotype H_a to be the ancestral haplotype, this would suggest maximum divergence times of 97,000 years between all *M. mutica* samples with exception of Sardinia (but including China and mainland Italy) and 286,000 years between the Sardinian population and all other samples. The strong predominance of one haplotype might be the result of a comparatively rapid colonization of the sampled range of *M. mutica* following a severe loss of genetic diversity (bottleneck). The level of divergence shown for the Sardinian samples suggests a comparatively long time of divergence and the population on Sardinia is therefore likely to have been largely isolated from the European (and Chinese) mainland population throughout the last glacial periods. The strong differentiation between the Sardinian population and samples from the Italian mainland might indicate additional influence of a persistent founder effect.

The divergence between COI haplotypes of *M. mutica* and *M. appendiculata* suggests a divergence time of only 1.3 My. This implies a much more recent split between the ancestors of both species than the previously calculated 2.5 My (Kölsch *et al.*, 2006). The difference between estimates is explained, however, by the fact that Kölsch et al. (2006) based their calculation on lower COI substitution rates (2.3% My^{-1}).

Reconstruction of historical patterns of population demography and biogeography based on mitochondrial DNA (mtDNA) is widely used but has to be interpreted with care, especially since the mitochondrial genome of arthropods is commonly subject to strong selective sweeps, often caused by vertically transmitted endosymbionts (see e.g. Hurst & Jiggins, 2005; Jäckel *et al.*, 2013). These processes can lead to rapid loss of mtDNA diversity and fixation of haplotypes. Although there is no further evidence for a possible selective sweep in *M. mutica* (besides a geographically widespread COI haplotype), it would be a possible explanation for the (lack of) structure observed for mtDNA haplotypes. In case of selective sweeps, mitochondrial DNA haplotypes often seize to be representative for population history and any estimates of divergence time or phylogeographic inferences based on mtDNA can be severely biased.

4.3.4 - Spatial Genetic Structure

Aquatic insects show varying patterns and scales of population structure, according to dispersal modes and dispersal abilities (Hughes *et al.*, 2009). Population structure of flightless aquatic insects is likely to show patterns of isolation by distance (Miller *et al.*, 2002) and strong signatures of habitat structure (Phillipsen & Lytle, 2013), reflecting low dispersal capacities. Considering the patterns of isolation by distance, the comparatively strong degree of subpopulation differentiation that was observed for *M. mutica* and the fact that detected clusters seem to reflect the hydrological connections and spatial structure of habitats therefore fits the picture of low active dispersal abilities in this aquatic beetle species.

Although potentially challenging, it is crucial to distinguish between different forms of spatial structure in genetic data. Hierarchical population structure and isolation by distance (IBD) are two sometimes confounded forms of spatial autocorrelation in genetic structure (Meirmans, 2012).

4.3.4.1 - Isolation by Distance

Testing for IBD is initially straightforward. A significant (positive) correlation between population differentiation and geographic distance between sampling sites can be interpreted as IBD. This regular increase in genetic differentiation (among individuals or subpopulations) with geographical distance – is mostly (but not exclusively) found in population genetic structure of species with comparatively limited dispersal abilities (Meirmans, 2012).

Hierarchical population structure shows two or more distinct clusters of (sub-) populations, often as the result of postglacial (re-) colonization from multiple glacial refugia. Since these clusters (like in the case of IBD) represent patterns of spatial autocorrelation, a hierarchical population structure can bias testing for IBD and vice versa; IBD can cause overestimation of hierarchical population structure (Frantz *et al.*, 2009; Meirmans, 2012).

To account for the possibility that detected hierarchical structure might bias IBD estimates, two detected genetic clusters (the eastern and the western cluster) were tested for IBD separately to account for the most distinct subdivision of sampled subpopulations, even though this meant loss of statistical power and the need to correct for multiple comparisons. This revealed that while the whole data set and the eastern cluster showed significant patterns of IBD the western cluster did not.

The lack of statistically significant patterns of IBD in the inland site group and the western cluster is unlikely to be a result of smaller sample sizes. While the high associated variance of

R_{ST} might cause estimates to suffer from smaller sample size (n = 24 sub-population pairs in case of the western cluster), R_{ST}-estimates have shown clearly significant IBD for even smaller subsets of the data (see 3.2.5.2). Since these groups consist solely of samples from hydrologically isolated inland waters, the weaker signature of geographical distance on genetic distance is unlikely to be the result of higher gene flow either. Instead, the lack of detectable IBD might be the result of different factors predominantly driving genetic differentiation across populations, according to different habitat structures. While patterns of genetic divergence across subpopulations in the Baltic Sea are likely to reflect limited dispersal capacity in a more continuously structured habitat (isolation by distance), genetic divergence across hydrologically isolated inland waters might reflect strong influences of colonization events (isolation by colonization) rather than contemporary patterns of migration (Spurgin *et al.*, 2014).

Concordantly, when testing inland samples and samples from the Baltic Sea separately, inland samples did not show IBD while the Baltic Sea populations did. The fact that the inland population group included the western cluster sites and two populations from the eastern cluster could potentially bias the test for IBD to some extent (Meirmans, 2012). But since excluding these sites (i.e. testing the western cluster) also showed no significant IBD, this potential effect seems negligible.

4.3.4.2 - Detection of Distinct Genetic Clusters

Testing for hierarchical structure, the number of distinct genetic clusters (K) and the boundaries between these clusters of subpopulations were inferred using different Bayesian assignment methods. Mapping the location of sampled *M. mutica* (sub-) populations (and detected genetic clusters) against landscape features proved to be useful for identification of possible barriers to gene flow and range expansion. This approach can, however, be suggestive when problems caused by spatial genetic structure and spatial heterogeneity of sampling efforts are not considered (Schwartz & McKelvey, 2008).

Besides bias due to patterns of IBD (isolation by distance) there are other known issues when interpreting clustering analyses results. And since even well-established methods for genetic cluster analyses can produce easily misleading results, it is advisable to base inference of the number of clusters K on more than one single method (Rowe & Beebee, 2007; Schwartz & McKelvey, 2008; Frantz *et al.*, 2009; Pedall *et al.*, 2011). Non-convergence between the clustering results of different programs can be expected to some extent (Frantz *et al.*, 2009)

and considering possible reasons for the differences in estimates admits feasible conclusions about the existing genetic structure.

While the ΔK method (Evanno *et al.*, 2005 (see below)) revealed a hierarchical population structure with two clusters (K = 2) of subpopulations as the uppermost level of population structure, STRUCTURE analyses using the likelihood approach and clustering analyses with BAPS using a high upper limit for K are likely to have overestimated K. This is known to occur in STRUCTURE analyses (Falush *et al.*, 2003; Evanno *et al.*, 2005), especially when isolation by distance influences population differentiation (Frantz *et al.*, 2009). Similarly, BAPS is known to overestimate weak differences in allele frequencies (and therefore overestimate K) when clustering individuals without spatial priors (Corander *et al.*, 2007; Rowe & Beebee, 2007). This fits the observation that using a spatial prior reduced the BAPS estimate for the number of clusters. Furthermore, BAPS might have overestimated *K* because subpopulations were well differentiated (Latch *et al.*, 2006). Additionally, the use of a correlated allele frequency model (as implemented in BAPS, STRUCTURE and optional in GENELAND) tends to underestimate admixture while overestimating K (Falush *et al.*, 2003).

To account for the problem of systematic overestimation of K, Evanno et al. (2005) developed the ad hoc statistic ΔK, based on the rate of change in the log probability of data between successive K values, to identify the uppermost or "true" level of population structure. The calculated ΔK value clearly supports a partition with two large clusters (*K*=2). The notable local ΔK maximum for *K*=3 and the optimal partition size according to clustering with GENELAND, however, suggest that the data might also support a partition of three clusters, possibly as the next-highest level of population structure.

4.3.4.3 - Uppermost Levels of Inferred Population Structure and Landscape Structure

Plotted ΔK values (see 3.1.2.6; Figure 12) notably showed a local maximum at K=3 and GENELAND runs invariably resulted in an optimal partition size of K=3 (see 3.1.2.6; Figure 17).

Consistently, when mapping the *M. mutica* sampling sites against landscape features, the *M. mutica* subpopulations largely seem to cluster genetically according to their geographic position within three biogeographic regions (Atlantic-, continental- and boreal biogeographic region, see 3.2.1.1). A possible cause for such a population structure is local adaptation to different ecological habitat characteristics (i.e. climatic differences across geographic regions) which can potentially lead to a reduction of gene flow across populations and influence population genetic structure (Orsini *et al.*, 2013b).

However, for a partition of *K*=3, allocation of the Stockholm Archipelago sites ("VAX and "UTO") between the two eastern clusters (continental and boreal cluster) was notably ambigous and not congruent between used clustering methods. Given the fact that sampling was relatively geographically disjunct between the north-eastern and south-western Baltic Sea, and that a significant pattern of isolation by distance (IBD) was detected among the Baltic Sea sites, it seems questionable if sites along the Baltic Sea coast really form two distinct clusters. These factors can easily lead to an overestimation of the number of clusters (K). The local maximum for K=3 in the ΔK plot and the GENELAND results favouring a partition of K=3 are hence likely to have resulted from a combination of spatially heterogenenous sampling and IBD in the Baltic Sea population(s).

For a predefined partition of K = 1 to K = 2, all used clustering programs congruently assigned various neighbouring sampling sites to two different clusters, independently of accounting for the geographical position of samples. This makes it very unlikely that the uppermost level of population structure (or the delineation between the two clusters) was falsely detected because of a patterns of IBD (Meirmans, 2012). Probabilities of cluster membership and proportional membership (see 3.1.2.5; Figure 14 and Figure 15) show a rather clear distinction between the eastern and the western cluster, in some cases notably separating geographically neighbouring sites (Figure 13). The inferred proportional cluster membership furthermore showed very little differences between different iterations of STRUCTURE runs (Figure 14). As a second order statistic the ΔK method cannot evaluate the probability for *K*=1 (Evanno *et al.*, 2005). However, given the clear differentiation between closely neighbouring sampling sites, it is highly unlikely that the data supports one continuous cluster (*K*=1). The partition of K=2 detected with the ΔK method is therefore likely to represents the most meaningful level of population structure within the analysed data and a genuine feature of population structure in *M. mutica*.

The two large genetic clusters correspond well to North Sea and Baltic Sea catchment areas, suggesting that the detected hierarchical population structure represents the results of hydrologic isolation between the eastern and the western cluster. The North Sea- and the Baltic Sea catchment areas are divided by the main European water shed which could be acting as a barrier for colonization and gene flow facilitated by transport with flowing water.

The fact that *M. mutica* samples from northern Germany and the Netherlands cluster tightly with British samples (see Figure 5, Figure 6 and Figure 8) might seem surprising at first. However, the separation of British Isles and mainland Europe followed a relatively recent,

postglacial flooding event (approximately 6,000 years ago). It is therefore plausible that the clustering of *M. mutica* samples from Great Britain with samples from the Netherlands and north-west Germany reflects a common ancestral population with a formerly more continuous distribution.

The Plön Lakes sampling site "PLW" is situated very close to the watershed separating the catchment areas draining into the North Sea and the Baltic Sea. Cluster analyses congruently assigned samples from the Plön Lakes to the western cluster. A recent study (Mende *et al.*, 2010) found *M. mutica* samples from waters just a few kilometers to the east of the Plön Lakes (Lake Selent) to show relatively strong genetic distance to the Plön Lakes samples and to cluster with Baltic Sea populations. Mende *et al.* (2010) attribute this to the lakes respective position relative to the watershed. The sampling site in Wales is situated in the Atlantic catchment area. A possibly resulting reduction of exchange with sites in the North Sea catchment area might be reflected in the relatively strong genetic differentiation between the Welsh samples and the other British samples (see Figure 6).

The only cases were the subdivision into clusters did not correspond well to mapped main catchment areas, were the two western Danish sampling sites in Jutland ("JUW", "JUN"). There is no evidence for current hydrologic connections of these sites to the Baltic Sea (and therefore to other sites within the eastern cluster). Both Jutland sampling sites are situated in North Sea coast lakes on the western side of the main European water shed (see Figure 24) but are congruently assigned to the eastern cluster in clustering analyses (see 3.1.2.5). The site in Western Jutland ("JUW") shows notably more ambiguous proportional cluster membership than other sites. This is probably caused by migration from the western cluster since one sample from western Jutland clearly represents a migrant from the western cluster (see 3.1.2.8). In case of both Jutland sampling sites, a former, indirect hydrologic connection to the Baltic Sea (i.e. the Kattegat) cannot be excluded for certain, since closely neighboring water bodies were directly or indirectly connected to the Kattegat in recent history. Given the rather modern origin of both Jutland sampling sites (both are derived from former North Sea fjords that have only very recently transformed into coastal lakes) colonization by *M. mutica* is likely to have occurred in the recent past and therefore possibly after the two genetic clusters formed to the east and to the west of the main European watershed.

Passive drifting with flowing water is the most important dispersal mechanism for many aquatic invertebrates (Bilton *et al.*, 2001; Van Leeuwen, 2012) and gene flow can be enhanced along currents and by flooding events (Kawata *et al.*, 2005). Mende *et al.* (2010)

highlight the likely importance of passive dispersal with flowing water for *M. mutica* and suggest that the main European water shed might explain the a strong genetic differentiation between Baltic Sea samples and Northern German samples. The presence of strong genetic differentiation between subpopulations along the main European water shed found for microsatellite data seems to support this interpretation.

It therefore seems straightforward to assume that the detected hierarchical population structure reflects different catchment areas in the studied range, given the low mobility of *M. mutica* and the probable importance of waterflow-mediated dispersal for this species. However, even in case of aquatic insects with very low active dispersal capacity, reduced gene flow across watersheds alone does not necessarily lead to a population structure that reflects the hydrogeology of their habitat (Miller *et al.*, 2002). Differences in postglacial (re-) colonization have therefore possibly added to the observed structure.

4.3.4.4 - Hierarchical Structure and Postglacial Range Extension

The (re-) colonization of suitable habitat in the course of de-glaciation after the last glacial maximum has greatly influenced the recent population genetic structure and distribution of many plant- and animal species (Hewitt, 1996, 1999, 2000; Taberlet *et al.*, 1998; Schmitt, 2007). Differences in post-glacial range expansion have recently been proposed as potential cause of strong regional genetic differentiation in *M. mutica* (Mende *et al.*, 2010). The genetic differentiation between the two large genetic clusters detected for *M. mutica* might therefore not only reflect low contemporary gene flow between hydrologically isolated habitats but also differences in route and timing of post-glacial range expansion.

Mende *et al*. (2010) reported a zone of disproportionately large genetic differentiation between neighboring *M. mutica* populations in Northern Germany. This putative suture zone corresponds to the border between the two large clusters detected for microsatellite data. Mende *et al*. (2010) discuss two (potentially confounded) explanations for the observed population genetic structure. Proposed causes are (1) different speeds of (postglacial) expansion over land and in the Baltic Sea that resulted in a merely *relatively* strong differentiation; and (2) postglacial colonization of the studied area by two genetic lineages from different glacial refugia, representing a less recent divergence of the two clusters.

The microsatellite analysis adds to the characterization of the putative contact zone between clusters in Northern Europe that was described by Mende *et al*. (2010). Concerted efforts to locate and sample yet unidentified *M. mutica* populations in Denmark, Great Britain, the Netherlands and Germany specifically allowed to verify the genetic similarity between

populations in north-western Germany, Great Britain and the Netherlands as well as the strong population genetic differentiation along a narrow zone in Northern Germany. Furthermore, extensive Bayesian clustering has corroborated the existence of a hierarchical population structure with a border between two large clusters situated in Northern Germany (and Denmark). The present results cannot, however, explain the origin of the observed structure with any kind of final certainty.

Microsatellites are quickly evolving markers and suitable for inferring levels of more recent gene flow while mitochondrial DNA (mtDNA) markers usually evolve more slowly and are therefore often used to study more ancient divergence of lineages. A deeper divergence between the two detected clusters based on mtDNA haplotypes would therefore support longer divergence times, consistent with a scenario of expansion from separate glacial refugia. COI haplotypes did not, however, mirror the divergence between the two large clusters for Northern European *M. mutica* subpopulations that was found for microsatellite data. Identification of lineages from different glacial refugia based on exclusivity of mitochondrial DNA (mtDNA) haplotypes is therefore not possible based on the present data. Considering the apparent lack of mtDNA variation in the ancestral population and assumed substitution rate estimates for COI, different glacial refugia during the last glacial maximum (Weichsel Glacial, 25,000 to 13,000 BP) would not necessarily have resulted in detectable divergence among Northern European *M. mutica* populations based on analysis of 600 bp fragments of COI. Therefore, different glacial refugia as explanation for the observed hierarchical population structure in Northern Europe can neither be confirmed nor excluded based on the present results.

Considering the complicated hydrogeographic history of the Baltic Sea with its considerable ecological transitions (e.g. changes in salinity) (Leppäranta & Myrberg, 2009), it is difficult to make assumptions about the point in time at which *M. mutica* might first have established in the Baltic Sea after the last glacial maximum. The Scandinavian ice sheet was, however, covering the full extent of the modern Baltic Sea area at the peak of the late Weichselian glacial maximum, while the area that corresponds to the modern range of *M. mutica* in northwestern Europe (i.e. the detected western cluster) was largely uncovered by the glacial ice sheet (Hewitt, 1999, 2004; Boulton *et al.*, 2001). The extent of the permafrost at the last glacial maximum is, however, likely to have rendered northwestern Europe largely uninhabitable for aquatic invertebrates (Mende *et al.*, 2010). Again, it is difficult to pinpoint at what time the ecological conditions permitted a range extension by *M. mutica* into northwestern Europe. It is, however, plausible to assume that the colonization of Northern

Europe to the west and to the east of the present main European water shed did not start at exactly the same time or occur at the same speed. After first colonization of the Baltic Sea, range extension of *M. mutica* might have occurred relatively quickly along the coast of an already large and continuous habitat. In contrast, postglacial range extension in northwestern Europe might have occurred in a comparatively more fragmented habitat structure, which more severely limited gene flow and expansion speed. In this sense, the zone of strong differentiation between neighboring populations in the eastern and the western cluster might indeed represent secondary contact between two postglacial expansion waves which, however, did not necessarily originate from separate glacial refugia.

Additionally (or alternatively), adaptive processes might have sped up differentiation upon colonization of the Baltic Sea. Nies & Reusch (2005) found strong reproductive isolation (and even incipient speciation) between neighbouring inland populations and Baltic Sea populations in a *M. mutica* host plant (*Potamogeton pectinatus* L.). Analogue to the findings for *M. mutica,* strong divergence found for microsatellite data was not reflected by data for a more slowly evolving phylogeographic marker (ITS), suggesting that the divergence is recent. Since differences in host plant use are well known to relatively quickly and effectively drive differentiation and even speciation in phytophagous insects (Feder *et al.*, 1988, 1994; Bush, 1994; Berlocher & Feder, 2002; Drès & Mallet, 2002) this developing divergence of host plants could well be mirrored in reduction of gene flow between *M. mutica* populations, resulting in accelerated differentiation between populations in inland lakes and the Baltic Sea. Therefore, relatively fast genetic differentiation between hydrologically isolated and differently structured habitats might have been sufficient to cause the pronounced differentiation between the Baltic Sea population and samples from north-western Europe, especially when considering that migration rates (and therefore gene flow) over land generally appear to be low in *M. mutica*. A parsimonious interpretation of the results would therefore not necessitate the assumption of separate glacial refugia.

4.3.5 - Migration Rates and Admixture

The admixture analyses with BAPS and STRUCTURE (see 3.1.2.8) as well as the migration rates calculated with MIGRATE-N (see 3.1.3) suggest that migration between the western and the eastern cluster in *M. mutica* predominantly (if not exclusively) occurred from west to east (in case of migration from the western cluster to the eastern cluster) or north-eastwards (in

case of migration from the south-western Baltic Sea to the north-eastern Baltic Sea), respectively.

The estimated effective population sizes (N_e) are rather rough estimates, since an average mutation rate was assumed for all loci (which are likely to mutate at different rates (Balloux & Lugon-Moulin, 2002)). Higher mutation rate estimates (e.g. 10^{-3} per locus per generation instead of the used estimate of 10^{-4} per locus per generation) have been reported for microsatellite loci (Balloux & Lugon-Moulin, 2002). The effective populations sizes (N_e) could therefore be even lower than calculated here. Low effective population sizes can to some extent be expected for *M. mutica* due to the high probability of strong founder effects (especially in inland habitats) and high reported levels of inbreeding (Mende *et al.*, 2010). Nevertheless, the calculated effective population sizes appear surprisingly low, which might reflect the low mobility of *M. mutica*. This is consistent with the observation that N_e is lower for inland sites (i.e. the western cluster) than for the Baltic Sea samples.

MIGRATE-N runs consistently showed comparatively high migration rates from the western cluster (Atlantic cluster) to the north-eastern Baltic Sea sites (boreal cluster). This was somewhat unexpected since the minimum distance between sampling sites in the Atlantic and the boreal clusters exceeds 1,000 km. Given the absence of any hydrologic connection, migration between these sites would therefore require considerable long-distance dispersal mediated by mobile vectors (i.e. waterbirds). The lack of samples from Southern Sweden might, however, suggest overly great distances between boreal and Atlantic sites, since yet undetected and un-sampled populations could provide "stepping stones" for gene flow between these clusters. Among the north-eastern subpopulations, Stockholm archipelago sites ("Vax" and especially "Uto") showed the least change in likelihood values when assigned to the western cluster (Figure 15) and the strongest signs of admixture with the western cluster (Figure 20) in BAPS- and STRUCTURE analyses. These sites are therefore the likely main cause for the strong migration rates detected by MIGRATE-N. Since the Stockholm Archipelago sites showed the highest local swan abundances of all sampling sites in the present data (see Table 12) this observation is therefore consistent with a potential role of waterbird-mediated dispersal.

The widely used and accepted "one-migrant-per-generation" rule-of-thumb (OMPG rule) states that a minimum of one migrant per generation is crucial to prevent loss of diversity within and divergence between subpopulations (Mills & Allendorf, 1996; Wang, 2004). The OMPG rule is based on inaugural theoretical work by Wright (1931). While some

assumptions made by Wright are likely to be violated in many cases, most departures from this ideal model can be accounted for by use of the *effective* number of migrants (Wang, 2004). The effective number of migrants per generation (N_em) therefore suggests that gene flow between the eastern and the western clusters is so low that further divergence seems more likely than genetic homogenization. The detected migration between the south-western Baltic Sea and the north-eastern Baltic Sea sites, however, is notably higher (N_em close to 1) and suggest that migration might be sufficient to prevent further divergence. These results corroborate the observation by Mende *et al.* (2010) who report considerably stronger admixture (caused by stronger gene flow) among Baltic Sea sites than among inland populations of *M. mutica*. Furthermore, the comparatively strong migration between north-eastern and south-western Baltic Sea sites suggests the existence of two (rather than three) large genetic clusters in the studied area and supports the interpretation that the distinction between two Baltic Sea clusters is an artefact of spatially heterogeneous sampling and isolation by distance (see 4.3.4.3).

4.4 - GIS Based Inference of Mute Swan Abundances and Movement Patterns

Mapping data for sightings of marked mute swans with QGIS provided the means to infer mute swan abundances around *M. mutica* sampling sites, to infer to what extent these sampling sites were connected by visitation of individual swans, and to visualize regional swan abundances and movement patterns. Data for regional directional distributions of mute swan movements were analysed to better visualize movements, to estimate potentially predominating routes of swan-mediated dispersal and to verify inferred movements by comparison with published data.

Due to the chosen seasonal distinction of movements and the large geographic scale of the analysis, comparison of inferred regional movements with previously published information on mute swan movements proved difficult. Inferred regional movements and hot spots of mute swan abundance are nevertheless consistent with published reports. Recorded movements suggest considerable mute swan movements between Great Britain and mainland Europe as well as between the south-western Baltic Sea and the northern Baltic Proper. Exchange between the North Sea coast and the Baltic Sea coast of Germany and Denmark, however, appeared notably low during spring and summer.

4.4.1 - Hotspots of Swan Abundance

Since the annual averages for mute swan sightings that were used in the present study were calculated for strictly defined areas around beetle sampling sites, these numbers are not easily comparable to count data presented in the literature, which are usually based on larger areas. Hotspots for mute swan sightings identified by mapping of re-encounters based on the EURING data set (see 3.2.3), however, are well consistent with important sites for mute swan populations reported by Atkinson *et al.* (2006).

4.4.2 - Seasonal Differences in Recovery Data

The much higher amount of sighting data for the half-year between March and October (winter) compared to the months between April and September (summer) and the significantly higher re-encounter (re-sighting) numbers in winter could be linked to higher ringing activity during the winter month or to the fact that waterbirds tend to aggregate in suitable areas (ice-free water bodies (Nilsson, 1975) or agricultural sites providing food) near human settlements during winter. Additionally, re-encounters representing recovery of dead birds might be more frequent in winter due to higher mortality rates during the cold months of the year. The comparison of re-encounters during winter and summer (see 3.2.4.1) at first appeared to deliver contradictory results: Relative to the total number of sightings, the number of individuals with more than one sighting (re-encounter / re-sighting data) is significantly higher in winter and the average distance between sightings is significantly higher between sightings in winter. Average distances travelled per day, however, suggest no significant differences between summer and winter. This would suggest that the longer average distance between sightings in winter is due to more time passing between encounters of individual swans during the winter and not because of stronger migratory activity during that period.
Movements for the month between April and September seem to show comparatively less homogenous directional distribution than movements between October and March (see Figure 30 and Figure 31). One possible reason is that, regardless of movements to established wintering sites, omnidirectional mass movements to avoid freezing waters can be expected in winter (Atkinson *et al.*, 2006), potentially leading to more homogenous directional patterns of recorded movement.

4.4.3 - Regional Mute Swan Movement Patterns

Information from ringing and recovery points transformed into linear data does not (necessarily) represent the exact route that was taken by a marked animal. But by considering data for a given period of time it can be used to statistically infer directional trends in (seasonal) movements.

The informative value of predominant movement directions inferred for the whole data set (Figure 34) is compromised by the wide geographical range of included data and the resulting variation in landscape- and habitat-structures. Recording movements based on regional data subsets allowed for more meaningful inference of movement patterns. But although query areas were chosen to include inferred hot spots of mute swan abundance and important mute swan populations according to Atkinson *et al.* (2006), query area delineation is inevitably still arbitrary to some extent. This is a general problem when defining sites for bird census data since definition procedures are rarely totally objective (Atkinson *et al.*, 2006).

Furthermore, directional distribution was recorded for movements in spring and summer (between April and September) and movements in autumn and winter (between October and March), mainly because the eggs of *M. mutica* (as a potential life stage for zoochorous dispersal by birds) are only present between April and September. Depending on the reproduction periods of different mute swan populations, recorded movements between April and September are likely to include dispersal of young birds and moult migration but might or might not include autumn migration and the return from moult sites (Rowell & Spray, 2004; Atkinson *et al.*, 2006) which makes verification based on published reports of seasonal migration difficult. Comparisons are further complicated by the large geographical scale of the analyzed data, since the beginning and end of seasonal mute swan movements (e. g. moult migration) are likely to vary with geographic latitude (Rowell & Spray, 2004; Atkinson *et al.*, 2006).

Atkinson *et al.* (2006) report that mute swans from the northernmost breeding areas in Sweden, Finland and the Baltic States move southwestwards to wintering areas in southern Sweden, Denmark and Northern Germany. Analysed re-encounter data for mute swans reflect these movements, the predominant direction for all recorded movements being north-east during summer and spring and south in autumn and winter. Swans in the south-western Baltic Sea (i.e. the Danish Straits and Southern Baltic Proper) showed considerable exchange with the Northern Baltic Proper (Figure 27).

Regional long distance movements for Southern Sweden between October and March were accordingly found strongly distributed towards the southwest (Figure 36a) but, perhaps contradictory, significant predominance of southwestward movements was also shown during the summer half-year (April-September) when a reverse movement (north-eastwards) during summer and spring would be expected. The reproductive season for mute swans in Sweden lasts until the middle of October (Atkinson *et al.*, 2006) but counts for summer through September might well have included beginning autumn migration in a south-western direction and moult migration by non-breeders, leading to the observed predominance of movements in north-eastern *and* south-western direction. When weighted by distance, however, the data strongly reflects movement distributed in an eastward direction in summer (Figure 32), which is more consistent with expectations for seasonal migration in this area. The recorded movements in a south-western direction were therefore mostly short-distance movements which is consistent with expectations for non-breeders and moult migration (G. Kölsch, personal communication).

A westward autumn migration is known for swans in the Baltic States from January to March (Atkinson *et al.*, 2006). Recorded movements consistently showed pronounced westward movements in winter (Figure 36b) but also showed significant predominance of east-northeastward *and* westward movements during both the summer half-year and winter half-year. Records for movements in September might in this case include beginning autumn migration, leading to the observed significant distributions in a westward direction for both summer- and winter half-year. Additionally, moult migration and the return from moult sites could cause a similar pattern. Perhaps more consistent with a predicted westward autumn migration, the proportion of westwards movements seems to be much greater between October and March (Figure 36b) while long distance movements show a stronger north-eastern movement distribution between April and September.

Despite being considered to show westward migration between January and March (Atkinson *et al.*, 2006) mute swans in Poland (Figure 38b) showed significant predominance of northwestwards movements between April and September and eastward movements during the winter half-year (October to March). A possible reason for the apparent contradiction might be the fact that little more than the coastal region of Poland was included for this analysis and predominance of movements along the coastline might have caused the recorded movement patterns.

For mute swans in Norway and Northern Denmark, wintering grounds are reported in the Netherlands, Germany, Poland, the Czech Republic and Slovakia as well as autumn migration

via east-southeast to south-southeast (Atkinson *et al.*, 2006). The scarcity of data for this region led to rather little informative value for analysis of predominant seasonal movements. It seems, however, that (similar to the Scottish and the Swedish population) the predominant direction for long distance movements (from April to September) is westward (Figure 32 and Figure 35b).

The population in England and Wales is described as "largely sedentary" with exchanges between Denmark, Netherlands, northern France, Sweden and Germany and southeastern direction of autumn migration (Atkinson *et al.*, 2006). While the Scottish population (mainland and Orkneys) is considered sedentary, swans from the Hebrides are considered "mainly sedentary" (Atkinson *et al.* (2006) report one ringing recovery from Norway). The present study found that British individuals could indeed be shown to travel between Great Britain and mainland Europe (within six months). Movements between April and September showed a strong tendency for westward migration, especially when weighted by travelled distances (Figure 32). The considerable number of long distance movements from northern Great Britain to Norway and Sweden in the analyzed data might be explained by migrating swans from the Hebrides. It does, however, seem more likely that the Scottish mainland- and Orkneys populations are not as sedentary as previously assumed and therefore at least add to the recorded movements between Great Britain and Scandinavia. Similarly, numerous re-encounters of ringed swans suggest considerable exchange of British mute swans with the Netherlands and Germany (Figure 27).

Mute swans in the Netherlands are held to be mainly sedentary but show autumn migration in north-northwestward to east-northeastward direction (Atkinson *et al.*, 2006). For swans ringed in western Lower Saxony, movements parallel to the North Sea coast are reported, with wintering sites either to the north-east or the south-west of moulting sites (Blüml *et al.*, 2012). For recorded movements in the Netherlands and Northwestern Germany, a north-eastward autumn migration might be traceable in recorded long distance movements between October and March (Figure 37b). Data analyzed for this study, however, mainly showed southward movements during the winter half-year (October to March) and significant predominance of westward movements during the summer half-year (April-September) (Figure 30). The predominance of westard movements in spring and summer appears even more striking when weighted by distance (Figure 32). These findings are consistent with reports of little exchange of populations breeding in Denmark and Eastern Germany with populations breeding in The Netherlands and Lower Saxony (Van Dijk, 1991; Van Dijk & Van Eerden, 1991; Blüml *et al.*, 2012). This observation is further substantiated by the significantly reduced numbers of mute

swan sightings in mainland Denmark (Jutland) and movements across Jutland in summer and spring (Figure 28), when compared to the rest of the year (Figure 29). The German North Sea coast is held to be of comparatively little importance for the German mute swan population (Harengerd *et al.*, 1990), possibly due to a lack of suitable habitats when compared to the Netherlands and the Baltic Sea. Combined with preferred movements along the course of coastlines or river systems (Rowell & Spray, 2004) and considerable moult site fidelity this might effectively limit exchange between mute swan populations to the east and to the west of Jutland.

Swans in East Germany and Schleswig-Holstein show wintering/ non-breeding grounds in Denmark and the Netherlands (Atkinson *et al.*, 2006). Danish and north-eastern German mute swans showed rather homogeneous directional distribution for movements in both summer and winter in the present study. This might at least partly be due to the complicated course of coastline included in the analyzed region, since mute swans can be expected to follow coast lines (Rowell & Spray, 2004) as recorded regional movements for southern Sweden and Poland (Figure 32) appear to reflect. Most of all, however, the apparent homogeneity of movement direction might reflect the importance of this region for mute swans and all-year omnidirectional movements of a large resident population.

4.4.4 - Potential Sources of Inherent Bias

When analysing ringing/ sighting data on large scales and over long periods of time it is inevitable to deal with the problem of bias. Ringing efforts, reporting probabilities and re-encounter probabilities will always be spatially and temporally heterogeneous. To reduce bias, local swan abundances were calculated as average number of sightings per year discarding records for years above the 95% percentile and below the 5% percentile for each site, thus removing records for years with exceedingly high or low ringing- or observer activity. Furthermore, calculating an annual average over 35 years should be less sensitive to temporal bias than e.g. a chronological analysis of population densities. Inferred swan abundances and movement patterns will to some extent nevertheless reflect differences in re-encounter probabilities. Additionally, mute swans have a long history of semi-domestication and the species is likely to frequent waters near human settlements and seek food from humans. Furthermore, while the practice of swan-keeping declined in recent centuries, swans escaping from semi-domesticated flocks have probably added to a recent increase of the wild

population (Rowell & Spray, 2004). It is difficult to tell to what extent this might influence how well mute swan sightings represent information about the wild population.

Data for mute swan sightings can, however, be generally expected to be considerably less affected by detectability-related observer bias than data for most other bird species, which are in most cases smaller and less conspicuous (Gayet *et al.*, 2011). The great size and distinct appearance of mute swans lead to a high detectability and therefore to high re-encounter probabilities of marked individuals. Furthermore, the practice of using highly visible neck collars instead of leg rings to mark individuals further improves re- encounter probability and therefore helps to increase reporting probability in case of re-encounters. As a result of the high detectability and high re-encounter probabilities, sighting data for mute swans are available in great quantities and often include several re-encounters for individuals. This improves the informative value for inference of movement patterns. Data for smaller species often provide little geo-referenced information beyond the ringing location and the location of the recovery of dead individuals (Atkinson *et al.*, 2006).

Furthermore, bias associated with ringing data are unlikely to compromise the credibility of tests that map sighting data against population genetic data for putatively transported aquatic organisms (Figuerola *et al.*, 2005; also see 4.5.2.6).

4.5 - Genetic Evidence for Swan-Mediated Dispersal of *Macroplea mutica*

4.5.1 - Consistence of Regional Mute Swan Movements with Putative Dispersal of M. mutica

Based solely on data on the ecology and distribution range of both species and data for predominant regional movements of swans, evidence for a potential role of mute swans for the dispersal of *M. mutica* is only correlative. Nevertheless, inferred regional movement patterns are consistent with a scenario of mute swan-mediated transport of *M. mutica*.

Assuming that the eggs are the *M. mutica* life stage best suited for (internal) dispersal by birds (see 4.6), vector movements and abundances between April and September (when eggs are present in the wild) should be most significant for dispersal. Consistently, the predominant direction of mute swan migration during spring and summer matches the predominant direction of gene flow detected in *M. mutica* (see Table 15 and 4.3.5). Furthermore, low exchange of swan populations in Denmark and Eastern Germany with swan populations in the Netherlands and Lower Saxony coincides with the border between two large genetic clusters

in *M. mutica*. Although this population genetic pattern is likely to have originated from differences in colonization and adaptation in hydrologically isolated habitats of different configuration (see 4.3.4.3), lack of bird-mediated dispersal between these clusters might add to its persistence. Similarly, while the partially tight genetic clustering of populations in Great Britain and on the European mainland (see e.g. Figure 5 and Figure 6) might reflect the distribution of a common ancestral population prior to the separation of the British Isles from the European mainland, the considerable recorded mute swan movements between Great Britain and mainland Europe suggest that waterbird-mediated dispersal could have contributed to the relatively low observed differentiation between certain *M. mutica* populations separated by the North Sea.

Tests for statistically significant associations between mute swan movements, mute swan abundances and pairwise genetic differentiation across *M. mutica* (sub-) populations provide less circumstantial evidence for a putative role of mute swans as dispersal vector for *M. mutica* (see 4.5.2).

4.5.2 - Pairwise Genetic Distances and Swan Movements and Abundances

If waterbirds are important vectors for the dispersal of *M. mutica*, observed population genetic structure (i.e. genetic distances between sampling sites) could reflect this mode of dispersal. If swans indeed act as dispersal vectors in *M. mutica*, transport events leading to more frequent exchange of individuals (beetles) between sites often visited by vectors should reduce genetic differentiation between populations (Mader *et al.*, 1998; Wada *et al.*, 2012; Van Leeuwen *et al.*, 2013).

Detecting undisputable genetic traces of bird-mediated dispersal is ultimately difficult. The inferred genetic evidence for waterbird-mediated dispersal of *M. mutica* is, however, comprehensive: Swan movements among sampled inland sites could be shown to be better predictors for genetic structure in *M. mutica* than geographic distance (see 4.5.2.1). Local swan abundances also showed significant correlations with pairwise genetic differentiation across *M. mutica* populations (see 4.5.2.3). Furthermore, the genetic data showed the breakdown of isolation by distance (IBD) between sampling sites with high swan abundances (see 4.5.2.4) and (residual) genetic distances decreased significantly with increasing swan abundances, even over large distances (see 4.5.2.2).

The fact that associations between swan sighting data and genetic differentiation in *M. mutica* were found for habitats with very different configurations (fragmented habitat structure

among inland habitats and more continuous habitat structure in the Baltic Sea) further supports a measurable impact of vector movements and abundances on the observed population structure. Habitat-related differences in the observed associations between mute swan abundances and movements with detected genetic differentiation are well consistent with expectations given the different habitat structures; recorded movements of a putative vector correlate most clearly with differentiation across inland habitats, where the degree of differentiation across populations is likely to be largely determined by colonization events rather than recent gene flow (also see 4.5.2.1). Across (sub-) populations within the more continuous Baltic Sea habitat, patterns of isolation by distance represent a comparatively stronger signature of dispersal distance limitation. Vector (i.e. mute swan) abundance along the Baltic Sea coast correlates negatively with the dependence of genetic distances on geographic distances (see 4.5.2.3 and 4.5.2.4), suggesting a homogenizing effect of present vectors on differentiation across populations.

To account for potential bias due to spatial auto-correlation, spatially heterogeneous sampling and analysis of hierarchically structured populations, statistical relationships were tested based on data subsets representing afore detected genetic clusters and correcting for geographic distance. Further sources of potential bias (see 4.5.2.6 – 4.5.2.8 and 4.5.3) are unlikely to favor a statistical relationship between swan abundances or movements with genetic differentiation in *M. mutica* as long as the dependence of genetic on geographic distances and significant hierarchical population structure is accounted for. Given the likeliness that discussed sources of possible bias (see 4.5.2.6 – 4.5.2.8 and 4.5.3) render tests more conservative rather than systematically favoring a statistical association between data for mute swan movement and abundance with the degree of differentiation across *M. mutica* populations, the detected correlations represent rather compelling evidence for bird-mediated dispersal of *M. mutica* and the underlying association might be even stronger than detectable based on the present data.

4.5.2.1 - Swan Traffic and Genetic Differentiation between Inland Populations of M. mutica

Total and relative swan traffic showed highly significant correlation with pairwise R_{ST} values for *M. mutica* populations in almost all tested groups. However, since in most cases geographic distance was significantly correlated with swan traffic between sites *and* genetic differentiation between beetle sampling sites, it is likely that the correlation between genetic

differentiation and swan traffic might be reflecting overlapping patterns of spatial auto-correlation rather than representing evidence for swan-mediated gene flow.

In case of the sampled inland sites and the western cluster, however, there is no statistical evidence for a systematic increase of genetic distance with geographic distance (isolation by distance, IBD). The inland sites group showed significant (negative) correlation between swan movements and genetic distance but *no* significant patterns of IBD. Associations between genetic differentiation across the western cluster sites and swan traffic (although largely similar to the inland sites group in composition) were not significant but showed comparatively low p-values. The lack of a statistically significant association might be caused by a comparatively small sample size, since the western cluster was by far the smallest tested group, with only 24 pairs of sampling sites.

Genetic differentiation among inland populations is significantly correlated with swan movements but not with local swan abundances, which might reflect a strong signature of colonization events in hydrologically isolated inland habitats. A comparatively strong effect of colonization patterns can be expected, since founder effects are more prominent in smaller and /or strongly fragmented habitats (De Meester *et al.*, 2002). This does, however, not compromise tests for correlation between swan movements and genetic distances across *M. mutica* populations. If anything, migration routes of vectors are likely to be even more strongly reflected by population structure dominated by colonization events than genetic differentiation more strongly affected by recent gene flow (De Meester *et al.*, 2002). This fact might explain why genetic differentiation among inland populations is significantly correlated with swan movements but not with swan abundances, since the local abundances of a vector species might rather influence the extent of contemporary gene flow, while predominant routes of vector movement might be more likely to mirror past colonization events. The reverse could apply to differentiation across subpopulations in the Baltic Sea. Here, genetic differentiation might more strongly reflect differences in contemporary gene flow related to local vector abundance rather than mirror vector migration routes, due to a weaker signature of colonization events in a more continuously structured habitat.

Theoretically, statistically non-significant (i.e. undetected) spatial structure in genetic differentiation could still lead to a bias towards small p-values, due to overlapping patterns of spatial autocorrelation between the R_{ST}-values and bird traffic counts. Because of this, results of this kind (when viewed isolated) have to be considered with some care when arguing for statistical evidence for bird-mediated gene flow.

Nevertheless, pairwise genetic differentiation among the inland (sub-) populations were better explained by swan traffic than by geographic distances between sampling sites. Cases like these, where "quantitative estimates of waterfowl movements provide(d) a better fit to genetic population structure than geographical distances" can be interpreted as evidence for waterbird-mediated zoochorous dispersal (Figuerola *et al.*, 2005).

4.5.2.2 - Swan Traffic and Residual Genetic Differentiation between M. mutica Populations in the Baltic Sea

The significant relationship between swan traffic and "raw" pairwise genetic differentiation across *M. mutica* subpopulations that was shown for the Baltic Sea might have resulted from bias due to overlapping spatial autocorrelation, since significant IBD was detected for genetic differentiation in the Baltic Sea population. However, *residual* R_{ST}-values showed a significant correlation between the amount of swan traffic and deviations from IBD. The extent to which sites are connected by recorded swan movements was hence significantly negatively correlated with the isolating effect of geographic distance. This suggests that dispersal limitations causing IBD might be reduced by bird-mediated transport along preferred routes of swan movement in the Baltic Sea area.

4.5.2.3 - Local Swan Abundances and Residual Genetic Differentiation between M. mutica Populations in the Baltic Sea

The deviation from spatially dependent population differentiation (i.e. residual R_{ST}, deviation from pairwise R_{ST} as expected for a given distance) correlates with the number of swan sightings in the whole data set and for the eastern cluster and Baltic Sea sites subsets; pairs of sites with high average swan counts showed lower differentiation than the geographic distance would lead to expect. This relationship stayed significant over geographic distances of hundreds of kilometers. When testing residual R_{ST} against seasonal data for average swan abundances, only swan sightings recorded for spring and summer (between April and September) showed significant correlation with residual R_{ST}- values while records for autumn and winter (October to March) showed no such association. The fact that population genetic differentiation of *M. mutica* is only mirrored by local abundances of potential vectors in spring and summer is consistent with experimental evidence that suggests that *M. mutica* eggs might be the life stage most important for dispersal by waterbirds (see 4.6.2), since eggs are only present in between April and September.

In a study that investigated the effect of bird-mediated dispersal on the population structure of pondweed, Mader *et al.* (1998) showed that the slope of geographic against genetic distance was lower for pondweed populations that were frequently visited by swans in comparison to those that were not visited by swans. Triest and Sierens (2013) criticized that in this case *higher* pairwise genetic differentiation at *small geographic scales* had falsely been interpreted as effect of zoochory on population structure and stated that bird-mediated dispersal should instead show an effect of *lower* differentiation over *long distances*.

Pairs of *M. mutica* subpopulations that were on average highly frequented by swans still showed significantly *lower* genetic differentiation than subpopulations at rarely visited sites when only considering greater geographic distances (up to minimal considered distances of 400 km within the eastern cluster). This suggests genetic homogenization by zoochory over longer distances while ruling out the effect of high differentiation at small geographic scales as sole reason for a significant correlation. Residual pairwise R_{ST} values for pairs of sites more than 1000 km apart might be compromised by the non-linearity of patterns of isolation by distance at large spatial scales (Bradbury & Bentzen, 2007). Calculated residual R_{ST} values might therefore tend to show small values for great geographic distances. Genetic homogenization by bird-mediated dispersal effective over a few hundred kilometers would, however, be consistent with published estimates. King *et al.* (2002) calculated that bird traffic along the Baltic Sea coast reduces genetic differentiation in the *M. mutica* host plant *Potamogeton pectinatus* over distances of 150 – 200 km.

The fact that the genetic signature of dispersal limitations (IBD) in *M. mutica* is reduced with increasing abundance of the putative vector therefore suggests that the presence of large number of swans might indeed increase possible dispersal distances, leading to genetic homogenization between *M. mutica* populations.

4.5.2.4 - Breakdown of IBD

Local swan abundances could also be shown to correlate with the spatial genetic structure of *M. mutica* in form of a breakdown of isolation by distance (IBD) between sites that are highly frequented by swans. Since strong patterns of IBD result from limitations in dispersal capacity (Miller *et al.*, 2002; Meirmans, 2012), zoochorous dispersal facilitating long distance dispersal can result in a breakdown of IBD wherever high numbers of vectors are present. Such breakdown of IBD has been reported as evidence for zoochorous dispersal of aquatic snails in hydrologically isolated ponds. Populations in ponds that were frequently visited by

vectors (large mammals in this case) showed a breakdown of IBD (while non-visited and rarely-visited neighboring sites showed significant IBD)(Van Leeuwen *et al.*, 2013).
M. mutica subpopulations were tested for significant patterns of isolation by distance, while accounting for recorded swan abundances. For the whole data set, the Baltic Sea subpopulations and the eastern cluster, Fisher's exact test showed that IBD is significantly less frequent among populations often visited by swans. Accordingly, Mantel tests for the total data set and the eastern cluster showed no IBD among often visited sites but showed significant IBD between little-visited sites. While the association between swan abundance and residual genetic differentiation shows that the spatially independent proportion of variation in genetic differentiation is correlated with local swan abundance, the breakdown of IBD shows that high swan abundance is also connected to reduced spatial dependence of differentiation across populations, suggesting that potential dispersal by swans might decrease dispersal distance limitations.

4.5.2.5 - Statistics

Pairwise genetic distances between *M. mutica* subpopulations were tested for correlation with geographic distances, average swan abundances and swan movements using simple Mantel tests and Spearman's rank correlation test.
In evolutionary biology and ecology Mantel tests are very popular for assessing significance of associations between matrices of distances and *partial* Mantel tests are often used to assess the relationship between two variables while correcting for some form of structure. The ubiquitous use of Mantel tests has, however, led to criticism and the validity of these tests has increasingly been challenged in recent years (Legendre & Fortin, 2010; Guillot & Rousset, 2013), especially in case of *partial* Mantel tests (Raufaste *et al.*, 2014; Rousset, 2014). To account for these concerns, residual pairwise differentiation (calculated by regression of genetic distance on geographic distance) were used to control for spatial structure instead of partial Mantel tests and the use of simple Mantel tests was limited to testing for the existence of structure of one variable in space (i.e. isolation by distance) or for dependence of two random variables that could be expressed as a matrices of distances (or similarities). Furthermore, great care was taken to ensure that spatial auto-correlation could be excluded for at least one of the two tested variables.
Spearman's rank correlation test was used to test for the significance of *monotonic* (rather than linear) associations in non-parametric data wherever compared variables could not be

expressed as matrices of distances or similarities for pairs of sites (e.g. in case of average swan abundances).

Residual R_{ST}-values were used in addition to "raw" R_{ST} values to account for the problem of spatial autocorrelation. Residual R_{ST} was calculated by regression on geographic distance and therefore only represents residual variance in the data (i.e. the proportion of variance that is not explained by geographic distance). Furthermore, the residual R_{ST}-values are likely to be imprecise to some extent since the calculation assumes perfect linearity of IBD while patterns of IBD might show non-linearity at very small or very large spatial scales (Bradbury & Bentzen, 2007). While this approach is therefore not without drawbacks, residual R_{ST}-values are advantageous in being demonstrably independent of geographical distance (see Table 16). This allows testing the deviation from an expected differentiation (for a given geographical distance) for correlation with recorded swan abundances and swan movements without having to fear bias due to overlapping patterns of spatial auto-correlation.

4.5.2.6 - Potential Bias Associated with Bird Ringing Data

Spatially heterogeneity of ringing activities and observation probabilities is the main bias connected to bird ringing data (Figuerola *et al.*, 2005; Korner-Nievergelt *et al.*, 2010; see also 4.4.4). However, when testing for correlations between ringing data and the genetic differentiation across populations of aquatic organisms, this kind of bias makes tests more conservative (making type II errors more likely and type I errors less likely) as there is no reason to expect associations between genetic differentiation of aquatic organisms and biases in ringing data (Figuerola *et al.*, 2005). Since observation probability for ringed birds increases mainly with observer density (i.e. with human population density (Korner-Nievergelt *et al.*, 2010)) the used tests could be systematically biased towards stronger statistical associations if human activity was strongly linked to dispersal of *M. mutica*. There is, however, no evidence for this connection. To the contrary, human activity might lead to regional reduction of gene flow in *M. mutica*, by increasing habitat fragmentation through anthropogenic wetland deterioration. Should this be the case, overlapping autocorrelation caused by human influence would systematically reduce statistical associations of bird sighting data and genetic differentiation in *M. mutica,* leading to more conservative tests.

4.5.2.7 - Used Measures of Pairwise Genetic Differentiation

The used estimators for R_{ST} (Rho_{ST}) and F_{ST} yield essentially unbiased estimates of population differentiation based on a weighted analysis of variance, correcting for differences

in sample size between populations and differences in variance between loci (Goodman, 1997). The main drawback of the calculated R_{ST}- (Rho_{ST}-) estimates is a high associated variance, especially when based on a modest number of loci and samples per population (Gaggiotti *et al.*, 1999; Balloux & Goudet, 2002). While this might lead to a loss of statistical power when testing for genetic evidence of bird-mediated dispersal, it is likely to make the used tests more conservative since the associated variance is unlikely to favor statistical associations with ecological variables. The microsatellite data could be validated by comparison with data for 251 AFLP loci (see 4.2.3) and showed considerable resolution power in simulations (see 4.2.2). Furthermore, the choice of R_{ST}-estimates (over F_{ST}-estimates) as measure of genetic differentiation is justified by a demonstrably smaller square mean error (see 4.3.2). It therefore seems that there is no reason to dispute conclusions based on the choice of markers and estimators of genetic differentiation.

4.5.2.8 – Dispersal and Gene Flow

Testing for genetic evidence of vector-mediated dispersal by comparing vector abundance and vector movement with genetic differentiation in *M. mutica* presupposes that the observed population genetic structure reflects the exchange of alleles as a result of dispersal events.

It has, however, been subject of recent debate to what extent dispersal and effective gene flow might under certain circumstances be mechanistically decoupled among populations of freshwater invertebrates (De Meester *et al.*, 2002; Bohonak & Jenkins, 2003; Orsini *et al.*, 2013a, 2013b). De Meester *et al.* (2002) describe a *dispersal - gene flow paradox*: high assumed dispersal capacity contrasts with strong genetic differentiation between neighboring populations in many aquatic invertebrate taxa. A proposed explanation states that local adaptation can lead to competitive superiority of resident genotypes over immigrant genotypes, which can severely limit effective gene flow between populations. Additionally, large resident population sizes often minimize the effective genetic contribution of immigrants. These mechanisms lead to resilient founder effects and patterns of population genetic differentiation that reflect colonization history much more strongly than contemporary dispersal events (De Meester *et al.*, 2002; Orsini *et al.*, 2013b).

These effects are, however, most prominent in species of cyclically parthenogenetic zooplankton and are unlikely to be of similar consequence in taxa with different life histories (De Meester *et al.*, 2002). The significance of local adaptive processes on population genetic differentiation should be much less pronounced for taxa like *M. mutica*, due to smaller

effective population sizes, less rapid growth and no asexual (clonal) phase of reproduction (De Meester *et al.*, 2002; Orsini *et al.*, 2013b).

Separately testing different habitat types and genetic clusters as well as explicitly accounting for geographic distances should further reduce potential bias due to influence of local adaptation on gene flow, since its impact is often strongly spatially auto-correlated (Orsini *et al.*, 2013b) and likely to be stronger between than within habitat types (Nies & Reusch, 2005).

Furthermore, while founder effects are indeed likely to show up in patterns of genetic differentiation in *M. mutica* (especially among inland habitats), population structure profoundly shaped by colonization events should reflect migration routes of vectors even more strongly than population structure predominantly influenced by contemporary gene flow (De Meester *et al.*, 2002; also see 4.5.2.1). The methods used in the present study should therefore be valid when testing for the genetic signature of dispersal.

4.5.3 - Other Dispersal Vectors

4.5.3.1 - Other Waterbird Species

The importance of *M. mutica* host plants (*Potamogeton* sp., *Zannichellia palustris, Ruppia* sp.) as food for numerous bird species suggest that several waterbird species, besides the mute swan, might potentially disperse *M. mutica*. Judging by overlaps in distribution, diet, the results of the feeding trials (see 4.6.2) and evidence for frequent excretion of viable macroinvertebrate propagules (Van Leeuwen *et al.*, 2012b) this includes mallards and other dabbling duck species as well as the Eurasian coot (*Fulica atra* L.).

Since waterbirds of different species strongly tend to congregate at suitable sites (Kirby *et al.*, 2008) recorded local mute swan abundances might represent valid information about dispersal vector abundances even if mute swans should not be the (single) most important vector species for *M. mutica*. Overlapping patterns of abundances and movement for mute swans and potentially more important vector species could therefore lead to overestimation of the putative role of mute swans as dispersal vectors but would not render the used tests invalid for assessing the significance of bird-mediated dispersal in *M. mutica*. Strongly different patterns of abundances and movements of other important vector species should cause used tests to be

more conservative rather than favouring statistical associations with population genetic data for *M. mutica*.

4.5.3.2 - Dispersal with Flowing Water

Transport with flowing water can certainly function as a dispersal mechanism for *M. mutica.* This is demonstrated by the fact that *M. mutica* specimens can be found washed ashore attached to floating host plant parts (Mende *et al.*, 2010). The fact that population structure reflects catchment areas (see 4.3.4.3) is also consistent with transport along hydrologic connections.

It can generally be assumed that dispersal of aquatic organisms with flowing water is more effective than dispersal by waterbirds due to a higher associated survival rate of transported organisms (Van Leeuwen, 2012). And since water will usually transports aquatic organisms throughout relatively suitable aquatic habitat, it can be considered a comparatively directional vector. Dispersal events (and especially colonization events) facilitated by waterbirds might, furthermore, be of relatively greater importance for organisms in comparatively short-lived, spatially highly structured and hydrologically isolated inland waters, whereas populations in more continuous and stable habitats (like the Baltic Sea) might be more strongly influenced by the water flow-mediated transport along predominant coastal currents. But even within large water bodies, water flow-mediated dispersal might be more strongly limited than it seems intuitively apparent. Hydrochorous dispersal in freshwater habitats can be restricted to downstream directions and (in the course of flooding events) to neighboring water bodies of similar elevation. And while currents in large water bodies can surely facilitate dispersal, they can also act as (directional) dispersal barriers. The high heterogeneity of coastal habitats leads to further restrictions of dispersal directionality (Srivastava & Kratina, 2013). Water-mediated dispersal across land during flooding events strongly limits the chances of ending in suitable habitats, therefore strongly reducing the directionality of transport.

Bird mediated dispersal can facilitate gene flow upstream, against predominant winds and surface currents and across topographic barriers. Waterbirds moving between ecologically similar water bodies can also provide dispersal with comparatively favorable directionality, unmatched by most other vectors (Van Leeuwen, 2012). Furthermore, observations that suggest population structure reflecting hydrologic connections do not oppose or contradict strong effects of proposed dispersal by waterbirds, since waterbird movements often follow catchment areas (preferring travel along coastal plains and avoiding higher grounds (Rowell

& Spray, 2004)). This may lead to slower genetic homogenization between hydrologically isolated populations of transported organisms.

Water flow- mediated dispersal might therefore not necessarily be of greater significance for the dispersal of *M. mutica* than bird-mediated dispersal, even within large continuous aquatic habitats.

The effects of both mechanisms are likely to be inevitably confounded in many cases and it is ultimately difficult to assess their relative contribution to observed genetic differentiation in many situations. It is, based on the present results, hardly possible to distinguish between effects of both mechanisms within the Baltic Sea. The higher genetic admixture among Baltic Sea populations of *M. mutica* compared to inland populations (Mende *et al.*, 2010) might reflect water flow-mediated gene flow. But due to high abundances of waterbirds along the Baltic Sea coast bird-mediated dispersal might at least add to this difference. Furthermore, the clear observed distinction between the eastern and the western cluster might to some extent also be the result of a confounded effect of (lacking) water- and bird-mediated dispersal. Swan movements and abundances seem to suggest that crossing the main European water shed in the area of Jutland might be avoided by swans, at least during summer (see 4.5.1). A hierarchical population structure that potentially originated by hydrological isolation might in this case persist due to this lack of seasonal exchange by waterbirds breeding to either side of the main European water shed.

To assess the relative importance of passive dispersal via flowing water in the Baltic Sea population of *M. mutica* directional gene flow between neighboring Baltic Sea coast populations would have to be quantified and tested against the direction of predominant surface currents along the coast.

It is theoretically possible that the spatial distribution of *M. mutica*, its host plants and the local abundances of waterbirds in the Baltic Sea are predominantly shaped by a similar combination of ecological factors (e.g. predominant wind direction determining coastal surface currents and wave exposure). This could potentially lead to overlapping patterns of autocorrelation that might cause the presented test results to represent the effect of surface currents rather than zoochorous dispersal. This would, however, not explain statistical associations of swan movement with genetic differentiation of *M. mutica* between inland waters. Furthermore, while bird migration is certainly influenced by wind direction, birds readily move against head winds and compensate for drift with winds (Krüger & Garthe, 2001). There is therefore little reason to expect systematic bias due to waterbird movements

and abundances predominantly reflecting the direction of surface currents that might add to determining the genetic differentiation across *M. mutica* populations.

The effect of additional dispersal vectors (besides bird-mediated dispersal) should therefore be more likely to weaken the statistical association between data for mute swans and genetic differentiation in *M. mutica*, instead of systematically biasing results in a way that favors a statistical association.

4.6 - Experimental Evidence for Potential of Internal Dispersal

4.6.1 - Digestion Simulation

The protocol for the laboratory simulation of waterfowl digestion was based on work by Furman *et al.* (2006) who had altered a procedure initially designed to simulate the human gastrointestinal tract in order to test lead bioaccessibility to waterfowl in mine-impacted soils. A similar protocol was used to study lead shot bioaccessibility in birds (Martinez-Haro *et al.*, 2009) Other studies using in vitro simulation of animal digestion are based on the "two-stage technique for the in vitro digestion of forage crops" (Tilley & Terry, 1963), e.g. the estimation of digestibility of oak browse diets for goats (Nastis & Malechek, 1988) or the simulation of grass seed passage through the digestive system of cattle (Ocumpaugh & Swakon, 1993).
The experimental setup might underestimate the physical forces applied in the gizzard of a bird and neglects simulation of potentially important stages of bird digestion like crop and cecae. As a result, this kind of simulation attempt will almost inevitably remain a rather crude approach to the real mechanical and chemical stress objects are exposed to after ingestion by birds. However, the method proved to be useful for initial testing of the potential of different life stages of *M. mutica* for endozoochorous dispersal before moving on to the more demanding feeding trials. *Macroplea* cocoons seemed to be exceedingly vulnerable to the physical forces applied during the simulation even though these might have been underestimated in the experimental setup. As all cocoons were clearly severely damaged after the laboratory simulations of gut passage, cocoons were excluded from further testing. Furthermore, *Macroplea* cocoons are considerably larger than 2 mm in diameter, which can be considered too large for internal transport (Green & Figuerola, 2005). Since a small percentage of eggs proved to be viable after the simulation, feeding trials were conducted to further test the potential of the eggs for internal transport.

4.6.2 - Feeding Trials

All recovered eggs were found *inside* of intact faecal lumps. This makes contamination with uningested eggs highly unlikely. While the offered diet was arguably easily digestible, it resembles diets chosen in other studies of this kind to represent a spring and summer diet (Charalambidou *et al.*, 2005).

Since digestive efficiency in birds varies considerably with species, age, and sex of the bird, season, activity, gizzard contents, seasonal diet and stress levels (Figuerola *et al.*, 2002; Charalambidou *et al.*, 2005; Van Leeuwen *et al.*, 2012a) the experimental setup does not allow addressing all factors necessary to quantify dispersal probabilities in detail. Nevertheless, the present results still represent a valid proof of concept for endozoochory. Recent publications have stressed the importance of extensive feeding studies to quantify the potential and limitations of *long distance* dispersal of putative propagules for endo-zoochorous dispersal (Clausen *et al.*, 2002; Green & Figuerola, 2005; Van Leeuwen *et al.*, 2012a). In the author's opinion, exact quantification of maximum dispersal distances is of subordinate importance in cases like that of *M. mutica*, as soon as ample evidence for a general capacity for internal dispersal is available. For a species as incapable of active dispersal as *M. mutica*, even rare events of *short distance* dispersal by waterbirds might already be of great significance.

The author proposes that eggs *of M. mutica* survive gut passage in the wild at least occasionally, thus facilitating internal dispersal. *M. mutica* eggs were found to survive retention times of up to 5-8 hours. And since retention in mallards may under certain circumstances well exceed 8 hours (Charalambidou & Santamaría, 2002; Charalambidou *et al.*, 2005) survival of even longer retention times cannot be excluded. Assuming that waterbirds reach average flight speeds of 75 km h^{-1} (Van Leeuwen *et al.*, 2012b), the presented results suggest potential dispersal distances of hundreds of kilometres. Eggs are present in the field from May until September. During this period there are considerable movements of waterfowl, encompassing moult migration and the onset of autumn migration. For the dispersal of otherwise immobile aquatic organisms, however, even non-seasonal local bird movements between neighbouring water bodies might be of importance (Green *et al.*, 2002).

Zoochory might accordingly be of particular significance to *M. mutica*, and the distribution and population genetic structuring of *M. mutica* in Northern Europe might to some extent

even reflect major routes of dispersal by migrating water fowl as a consequence (Mende *et al.*, 2010; see also 4.5).
External transport of *M. mutica* by waterbirds is a further dispersal possibility. There is, however, no evidence for this yet and - for organisms that can survive gut passage - internal transport might quantitatively be of much greater importance (Brochet *et al.*, 2010b; Sánchez *et al.*, 2012). Future research of both internal and external transport of aquatic insects is much needed.

Birds might show a tendency for reduced digestive efficiency and retention times in favour of maximized net energy intake, allowing internal transport of species which do not show obvious adaptations to endozoochorous dispersal (Van Leeuwen *et al.*, 2012b, 2012c). Thus, internal transport by waterbirds might be much more common and important than previously thought, particularly for freshwater species with little potential for active dispersal.

4.7 - Directions for Further Research

The present thesis presents first evidence for zoochorous dispersal in aquatic reed beetles of the genus *Macroplea*. There is much left know in order to fully understand how mechanisms of passive dispersal shaped the current distribution of *Macroplea mutica* and to what extent similar mechanisms might apply to other *Macroplea* species and other aquatic taxa hitherto not associated with zoochorous dispersal.

Experimental Evidence for the Potential of Internal Dispersal

Given the great plasticity of digestive processes in waterbirds (Charalambidou *et al.*, 2005; Van Leeuwen *et al.*, 2012a) and great temporal and interspecific differences of digestive efficiency (Figuerola *et al.*, 2002), it would be desirable to test if the results of feeding trials with mallards are reproducible, preferably including other waterbird species (e.g. mute swans).

Validating Genetic Evidence for Bird-Mediated Dispersal

In order to further validate the presented findings, it would also be desirable to extend performed analyses to ringing data for further waterbird species, further population genetic data for additional genetic markers and samples from further (yet unidentified) populations of

M. mutica and other *Macroplea* species. Testing for associations of genetic differentiation in *M. mutica* with abundances and movements of further bird species should include waterbird species with similar distribution and similar diet and habitat use (e.g. dabbling ducks (Anatinae) or coot (Rallidae)) to assess the relative importance of mute swans and other species as dispersal vectors for *M. mutica*. Additionally, comparison with data for bird species that are unlikely (internal) dispersal vectors due to their diet and feeding habits (e.g. wader species like godwits) could function as negative control to assure that detected associations between population genetic differentiation in *M. mutica* and waterbird abundances are truly independent of effects like observer density.

The microsatellite primers developed for *M. mutica* could potentially allow amplification of microsatellite loci in its sister species *Macroplea appendiculata* without further efforts in marker development. A comparison of dispersal ecology between the sister species *M. mutica* and *M. appendiculata* would be of interest, since their ecological differentiation is still poorly understood (Kölsch & Kubiak, 2011). An even more extensive comparison of dispersal mechanisms across the whole genus *Macroplea* might shed light on the notable differences in the extent of geographic distribution between *Macroplea* species (see Lou *et al.*, 2011).

Analysis of additional areas of the mitochondrial genome of *M. mutica* and testing for the presence of endo-symbionts that are generally associated with selective mitochondrial sweeps (e.g. *Wolbachia* (Hurst & Jiggins, 2005; Jäckel *et al.*, 2013)) would allow evaluating the informative value of mtDNA haplotype distributions for reconstruction of phylogeography in *M. mutica*.

Waterflow-Mediated Dispersal and Further GIS-Based Analyses of Environmental Variables

To better understand the relative importance of dispersal with flowing water and waterbirds, it seems recommendable to map genetic differentiation and predominant directions of gene flow across populations in the Baltic Sea against predominant coastal surface currents. Furthermore, in order to better account for potential influences of adaptation on gene flow and genetic differentiation, a number of environmental distances (e.g. climatic zones or latitudinal distances) as well as data for host plant use and population genetic data for host plans (e.g. *Potamogeton pectinatus*) could be mapped against population genetic data for *M. mutica*.

Other Aquatic Invertebrate Taxa

Finally, the evidence for waterbird-mediated dispersal in an aquatic leaf beetle implies that (internal) bird-mediated transport of aquatic invertebrates might be a more ubiquitous phenomenon than previously thought, even among volant aquatic insects (Green & Sánchez, 2006; Frisch *et al.*, 2007a). Looking into the possibility of bird-mediated transport of taxa not hitherto associated with this mode of dispersal could be a rewarding field of study and should help to further understand the ecological significance of migrating waterbirds for the world-wide biodiversity and distribution of aquatic freshwater invertebrates.

4.8 - Conclusion

Dispersal is one of the most essential ecological and evolutionary processes and of great significance for species distribution, biodiversity and long-term meta-population survival. Mechanisms of passive dispersal are therefore of great importance for species with low potential for active dispersal. For freshwater organisms dispersal is especially challenging and vital, and increasingly so due to recent anthropogenic wetland deterioration that further fragments already heterogeneously distributed habitats. Understanding how dispersal processes determine the connectivity among freshwater habitats might therefore be crucial for conservation efforts.

The fully aquatic leaf beetle *Macroplea mutica* shows an apparent contrast of a wide Palearctic distribution and strikingly low potential for active dispersal. It therefore represents an interesting and well suitable model to study potential mechanisms of passive dispersal in aquatic insects. The presented results represent rare evidence for waterbird-mediated dispersal of an aquatic insect.

Population genetic data based on six newly developed highly polymorphic microsatellite markers and *M. mutica* samples from 21 sampling locations in Northern Europe showed the strong genetic differentiation that could be expected given the low potential for active dispersal in this species. Results of extensive cluster analyses furthermore showed a strongly hierarchic population structure with a contact zone between two large genetic clusters running through Denmark and Northern Germany. The divergence between these clusters is likely to have originated from differential post-glacial range expansion in the Baltic Sea area and northwestern Europe. Due to a lack of phylogeographic structure, mitochondrial DNA haplotypes did, however, not support a hypothesis of separate glacial refugia.

Mapping population genetic data against ecological variables with GIS-based methods provides powerful tools to analyze associations between population genetic structure and ecological variables. GIS-based approaches are especially promising for the study of landscape features that might act as potential barriers to dispersal or spatial distribution of factors that might facilitate dispersal (like vector densities and hydrologic connections).
It is, however, crucial to conscientiously account for potential sources of bias (like spatial autocorrelation, spatially heterogeneity of collected data and potentially confounded factors driving the analyzed population genetic structure).
GIS-based mapping of sighting data for individually marked mute swans allowed inference of mute swan abundances and movement patterns on different spatial levels. It was therefore possible to map geo-referenced information for a potential vector species against (spatial) genetic structure of *M. mutica*. Significant correlations between genetic differentiation across *M. mutica* populations with mapped swan movements and abundances, suggest significance of waterbird-mediated transport for the dispersal of *M. mutica*.

Without the ability to fly or walk across land, passive dispersal by waterbirds might be of great importance to *M. mutica*, even though the species lacks the more apparent adaptations to zoochorous transport that are found in many physically and chemically resistant propagules of plants and zooplankton. Experimental evidence from feeding trials with mallards (*Anas platyrhynchos*) demonstrates that eggs of *M. mutica* are capable of surviving passage through waterbird gut (see 4.6.2 and Laux & Kölsch, 2014). This strongly suggests potential for internal waterbird-mediated dispersal. Given the fact that eggs must be regularly ingested by waterbirds with foraged plant material, at least occasional events of internal transport seem inevitable. These findings are therefore of far-reaching consequence for the evaluation of passive dispersal capacity in *M. mutica*, since internal transport by waterbirds is demonstrably a potent means of passive dispersal for a wide range of aquatic invertebrate taxa (see e.g. Green *et al.*, 2002; Green & Sánchez, 2006; Frisch *et al.*, 2007; Van Leeuwen *et al.*, 2012b).

As long as viable *M. mutica* specimens are not repeatedly found on the bodies or in the feces of waterbirds at the very instance of arrival after (migratory) movements between water bodies, thus indisputably proving frequent dispersal events, all evidence for importance of waterbird-mediated dispersal could be considered to be to some extent circumstantial or anecdotal. However, this kind of indisputable proof might be difficult to obtain in many

systems of zoochorous dispersal and the search for genetic evidence for this mode of transport is therefore an approach of obvious importance.

Detecting traces of zoochorous dispersal events in population genetic data of a putatively transported organism is ultimately a complex task. Given that the presented evidence for bird-mediated dispersal are based on data for sightings of only one potential vector species and that questions about the relative significance of dispersal with flowing water along recent or former hydrologic connections remain largely unanswered, detected associations might not represent indisputable evidence. However, considering the contrast of wide geographical distribution and low mobility in *M. mutica* and the presented evidence of potential for internal transport, the comprehensive significant associations of population genetic differentiation with abundances and movements of a suitable avian vector species provide ample reason to assume a considerable role of waterbird-mediated transport for the dispersal of *M. mutica*.

The presented evidence furthermore corroborates recent findings that suggest a far more ubiquitous importance of internal zoochory for organisms lacking clearly apparent adaptations to survival in the digestive tract of waterbirds (Figuerola *et al.*, 2005; Green & Sánchez, 2006; Frisch *et al.*, 2007a; Van Leeuwen *et al.*, 2012b, 2012c; Wada *et al.*, 2012) and highlights that bird-mediated dispersal of aquatic insects is a topic that deserves further attention.

References

Adams, S.M., Lindmeier, J.B. & Duvernell, D.D. (2006) Microsatellite analysis of the phylogeography, Pleistocene history and secondary contact hypotheses for the killifish, *Fundulus heteroclitus*. *Molecular Ecology*, **15**, 1109–1123.

Allin, C.C. & Husband, T.P. (2003) Mute Swan (*Cygnus olor*) Impact on submerged aquatic vegetation and macroinvertebrates in a Rhode Island coastal pond. *Northeastern Naturalist*, **10**, 305–318.

Altizer, S., Bartel, R. & Han, B.A. (2011) Animal migration and infectious disease risk. *Science*, **331**, 296–302.

Amezaga, J., Santamaría, L. & Green, A. (2002) Biotic wetland connectivity - supporting a new approach for wetland policy. *Acta Oecologica*, **23**, 213–222.

Askevold, I. (1990) Reconstructed phylogeny and reclassification of the genera of Donaciinae (Coleoptera: Chrysomelidae). *Quaestiones Entomologicae*, **26**, 601–664.

Atkinson, P.W., Clark, J.A., Delany, S., Hamalla Diagana, C., Feu, C. du, Fiedler, W., *et al.* (2006) Urgent preliminary assessment of ornithological data relevant to the spread of Avian Influenza in Europe. Report to the European Commission. (ed. Delany, S., Veen, J. & Clark, J.A.). *http://ec.europa.eu/environment/nature/conservation/wildbirds/birdflue/docs/rep_spread_avian_influenza_report.pdf (Accessed October 14th, 2014)*, p. 343.

Balloux, F. & Goudet, J. (2002) Statistical properties of population differentiation estimators under stepwise mutation in a finite island model. *Molecular Ecology*, **11**, 771–783.

Balloux, F. & Lugon-Moulin, N. (2002) The estimation of population differentiation with microsatellite markers. *Molecular Ecology*, **11**, 155–165.

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

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

Beerli, P. & Felsenstein, J. (1999) Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, **152**, 763–773.

Beerli, P. & Felsenstein, J. (2001) Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4563–4568.

Beladjal, L. & Mertens, J. (2009) Diaspore dispersal of Anostraca by flying insects. *Journal of Crustacean Biology*, **29**, 266–268.

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

Bilton, D., Freeland, J. & Okamura, B. (2001) Dispersal in freshwater invertebrates. *Annual Review of Ecology and Systematics*, **32**, 159–181.

Blüml, V., Degen, A. & Kruckenberg, H. (2012) Dispersal and choice of wintering- and moulting grounds of Mute Swans *Cygnus olor* ringed as juveniles in Western Lower Saxony. *Vogelwarte*, **50**, 155–168.

Bohonak, A.J. & Jenkins, D.G. (2003) Ecological and evolutionary significance of dispersal by freshwater invertebrates. *Ecology Letters*, **6**, 783–796.

Boulton, G.S., Dongelmans, P., Punkari, M. & Broadgate, M. (2001) Palaeoglaciology of an ice sheet through a glacial cycle : the European ice sheet through the Weichselian. *Quaternary Science Reviews*, **20**, 591–625.

Böving, A.G. (1910) Natural history of the larvae of Donaciinae. *Internationale Revue der gesamten Hydrobiologie und Hydrographie (Suppl.)*, **3**, 1–108.

Bradbury, I. & Bentzen, P. (2007) Non-linear genetic isolation by distance: implications for dispersal estimation in anadromous and marine fish populations. *Marine Ecology Progress Series*, **340**, 245–257.

Brochet, A.-L., Gauthier-Clerc, M., Guillemain, M., Fritz, H., Waterkeyn, A., Baltanás, Á., *et al.* (2010a) Field evidence of dispersal of branchiopods, ostracods and bryozoans by teal (*Anas crecca*) in the Camargue (southern France). *Hydrobiologia*, **637**, 255–261.

Brochet, A.-L., Guillemain, M., Fritz, H., Gauthier-Clerc, M. & Green, A.J. (2009) The role of migratory ducks in the long-distance dispersal of native plants and the spread of exotic plants in Europe. *Ecography*, **32**.

Brochet, A.-L., Guillemain, M., Fritz, H., Gauthier-Clerc, M. & Green, A.J. (2010b) Plant dispersal by teal (*Anas crecca*) in the Camargue: duck guts are more important than their feet. *Freshwater Biology*, **55**, 1262–1273.

Brochet, A.-L., Guillemain, M. & Gauthier-Clerc, M. (2010c) Endozoochory of Mediterranean aquatic plant seeds by teal after a period of desiccation: determinants of seed survival and influence of retention time on germinability. *Aquatic Botany*, **93**, 99–106.

Brochet, A.-L., Guillemain, M., Gauthier-Clerc, M., Fritz, H. & Green, A.J. (2010d) Endozoochory of Mediterranean aquatic plant seeds by teal after a period of desiccation: Determinants of seed survival and influence of retention time on germinability and viability. *Aquatic Botany*, **93**, 99–106.

Brown, C. (1933) A limnological study of certain fresh-water Polyzoa with special reference to their statoblasts. *Transactions of the American Microscopical Society*, **52**, 271–316.

Bush, G.L. (1994) Sympatric speciation in animals: new wine in old bottles. *Trends in Ecology & Evolution*, **9**, 285–288.

Cáceres, C. & Soluk, D. (2002) Blowing in the wind: a field test of overland dispersal and colonization by aquatic invertebrates. *Oecologia*, **131**, 402–408.

Charalambidou, I., Ketelaars, H.A.M. & Santamaria, L. (2003) Endozoochory by ducks: influence of developmental stage of *Bythotrephes* diapause eggs on dispersal probability. *Diversity and Distributions*, **9**, 367–374.

Charalambidou, I. & Santamaría, L. (2002) Waterbirds as endozoochorous dispersers of aquatic organisms: a review of experimental evidence. *Acta Oecologica*, **23**, 165–176.

Charalambidou, I., Santamaria, L., Jansen, C. & Nolet, B.A. (2005) Digestive plasticity in Mallard ducks modulates dispersal probabilities of aquatic plants and crustaceans. *Functional Ecology*, **19**, 513–519.

Cheng, L., Connor, T.R., Aanensen, D.M., Spratt, B.G. & Corander, J. (2011) Bayesian semi-supervised classification of bacterial samples using MLST databases. *BMC Bioinformatics*, **12**, 1–10.

Cheng, L., Connor, T.R., Sirén, J., Aanensen, D.M. & Corander, J. (2013) Hierarchical and spatially explicit clustering of DNA sequences with BAPS software. *Molecular Biology and Evolution*, **30**, 1224–1228.

Clausen, P., Nolet, B.A., Fox, A.D. & Klaassen, M. (2002) Long-distance endozoochorous dispersal of submerged macrophyte seeds by migratory waterbirds in northern Europe - a critical review of possibilities and limitations. *Acta Oecologica*, **23**, 191–203.

Cockerham, C.C. (1973) Analyses of gene frequencies. *Genetics*, **74**, 679–700.

Colbert, J. (2001) *Dispersal*. Oxford University Press, Oxford.

Corander, J., Marttinen, P. & Mäntyniemi, S. (2006) A Bayesian method for identification of stock mixtures from molecular marker data. *Fishery Bulletin*, **104**, 550–558.

Corander, J., Marttinen, P., Sirén, J. & Tang, J. (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC Bioinformatics*, **9**, 1–14.

Corander, J., Sirén, J. & Arjas, E. (2007) Bayesian spatial modeling of genetic population structure. *Computational Statistics*, **23**, 111–129.

Corander, J., Waldmann, P. & Sillanpää, M. (2003) Bayesian analysis of genetic differentiation between populations. *Genetics*, **374**, 367–374.

Cox, M.L. (2007) *Atlas of the Seed and Leaf Beetles of Britain and Ireland*. Pisces Publications, Newbury.

Darwin, C. (1859) *On the Origin of Species by Means of Natural Selection.* John Murray, London.

Darwin, F. (1909) *The Foundations of the Origin of Species. Two Essays Written in 1842 and 1844.* Cambridge University Press, Cambridge.

Dieckmann, U., O'Hara, B. & Weisser, W. (1999) The evolutionary ecology of dispersal. *Trends in Ecology & Evolution*, **14**, 88–90.

Dijk, K. Van. (1991) Herkomst en leeftijdssamenstelling van ruiende Knobbelzwanen *Cygnus olor* op het IJsselmeer. *Limosa*, **64**, 41–46.

Dijk, K. Van & Eerden, M.R. Van. (1991) Biometrics and timing of primary moult of non-breeding Mute Swans Cygnus olor at Lake lJsselmeer, the Netherlands. In *Proc. Third IWRB International Swan Symposium Wildfowl Supp. 1*. Oxford, pp. 296–303.

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

Egevang, C. & Stenhouse, I. (2010) Tracking of Arctic terns *Sterna paradisaea* reveals longest animal migration. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 2078–2081.

Ellegren, H. (2000a) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends in Genetics*, **16**, 551–558.

Ellegren, H. (2000b) Heterogeneous mutation processes in human microsatellite DNA sequences. *Nature Genetics*, **24**, 400–402.

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

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.

Faircloth, B.C. (2008) MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. *Molecular Ecology Resources*, **8**, 92–94.

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.

Fayard, J., Klein, E.K. & Lefèvre, F. (2009) Long distance dispersal and the fate of a gene from the colonization front. *Journal of Evolutionary Biology*, **22**, 2171–2182.

Feder, J.L., Chilcote, C.A. & Bush, G.L. (1988) Genetic differentiaton between sympatric host races of the apple maggot fly *Rhagoletis pomonella*. *Nature*, **336**, 61–64.

Feder, J.L., Opp, S.B., Wlazlo, B., Reynolds, K., Go, W. & Spisak, S. (1994) Host fidelity is an effective premating barrier between sympatric races of the apple maggot fly. *Proceedings of the National Academy of Sciences of the United States of America*, **91**, 7990–7994.

Figuerola, J., Green, A. & Santamaría, L. (2002) Comparative dispersal effectiveness of wigeongrass seeds by waterfowl wintering in south-west Spain: quantitative and qualitative aspects. *Journal of Ecology*, **90**, 989–1001.

Figuerola, J. & Green, A.J. (2002) Dispersal of aquatic organisms by waterbirds: a review of past research and priorities for future studies. *Freshwater Biology*, **47**, 483–494.

Figuerola, J., Green, A.J. & Michot, T.C. (2005) Invertebrate Eggs Can Fly: Evidence of Waterfowl-Mediated Gene Flow in Aquatic Invertebrates. *The American Naturalist*, **165**, 274–280.

Figuerola, J., Green, A.J. & Santamaria, L. (2003) Passive internal transport of aquatic organisms by waterfowl in Doñana, south-west Spain. *Global Ecology and Biogeography*, **12**, 427–436.

Finn, D.S., Theobald, D.M., Black, W.C. & Poff, N.L. (2006) Spatial population genetic structure and limited dispersal in a Rocky Mountain alpine stream insect. *Molecular Ecology*, **15**, 3553–3566.

Frankham, R. & Ralls, K. (1998) Conservation biology: Inbreeding leads to extinction. *Nature*, **392**, 441–442.

Frantz, A.C., Cellina, S., Krier, A., Schley, L. & Burke, T. (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, **46**, 493–505.

Freude, H., Harde, K.W. & Lohse, G.A. (1966) *Die Käfer Mitteleuropas*. Goecke & Evers, Krefeld.

Frisch, D., Green, A.J. & Figuerola, J. (2007a) High dispersal capacity of a broad spectrum of aquatic invertebrates via waterbirds. *Aquatic Sciences*, **69**, 568–574.

Frisch, D., Green, A.J., Figuerola, J., Group, W.E., Doçana, E.B. De, Maria, A., *et al.* (2007b) High dispersal capacity of a broad spectrum of aquatic invertebrates via waterbirds. *Aquatic Sciences*, **69**, 568–574.

Fronhofer, E.A., Kubisch, A., Hilker, F.M., Hovestadt, T. & Poethke, H.J. (2012) Why are metapopulations so rare? *Ecology*, **93**, 1967–78.

Furman, O., Strawn, D.G., Heinz, G.H. & Williams, B. (2006) Risk assessment test for lead bioaccessibility to waterfowl in mine-impacted soils. *Journal of Environmental Quality*, **35**, 450–458.

Gaggiotti, O.E., Lange, O., Rassmann, K. & Gliddon, C. (1999) A comparison of two indirect methods for estimating average levels of gene flow using microsatellite data. *Molecular Ecology*, **8**, 1513–1520.

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.

Gayet, G., Eraud, C., Benmergui, M., Broyer, J., Mesleard, F., Fritz, H., *et al.* (2011) Breeding mute swan habitat selection when accounting for detectability: a plastic behaviour consistent with rapidly expanding populations. *European Journal of Wildlife Research*, **57**, 1051–1056.

Goodman, S. (1997) RST Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology*, **6**, 881–885.

Goudet, J., Raymond, M., Meeüs, T. de & Rousset, F. (1996) Testing differentiation in diploid populations. *Genetics*, **144**, 1933–1940.

Green, A., Frisch, D., Michot, T., Allain, L. & Barrow, W. (2013) Endozoochory of seeds and invertebrates by migratory waterbirds in Oklahoma, USA. *Limnetica*, **32**, 39–46.

Green, A.J. & Figuerola, J. (2005) Recent advances in the study of long-distance dispersal of aquatic invertebrates via birds. *Diversity and Distributions*, **11**, 149–156.

Green, A.J., Figuerola, J. & Sánchez, M.I. (2002) Implications of waterbird ecology for the dispersal of aquatic organisms. *Acta Oecologica*, **23**, 177–189.

Green, A.J. & Sánchez, M.I. (2006) Passive internal dispersal of insect larvae by migratory birds. *Biology Letters*, **2**, 55–57.

Guedj, B. & Guillot, G. (2011) Estimating the location and shape of hybrid zones. *Molecular Ecology Resources*, **11**, 1119–1123.

Guerne, M. de. (1887) Remarques sur la distribution géographique du genre Podon, sur l'origine des Polyphémides pélagiques lacustres et sur le peuplement des lacs. *Bulletin de la Societe Zoologique de France*, **12**, 357–367.

Guerne, M. de. (1888) Sur la dissemination des organismes d'eau douce par les Palmipèdes. *Comptes Rendus des Séances de la Société de Biologie et de ses Fliiales*, **8**, 294–298.

Guillot, G. (2008) Inference of structure in subdivided populations at low levels of genetic differentiation-the correlated allele frequencies model revisited. *Bioinformatics*, **24**, 2222–2228.

Guillot, G., Estoup, A., Mortier, F. & Cosson, J.F. (2005a) A spatial statistical model for landscape genetics. *Genetics*, **170**, 1261–1280.

Guillot, G., Mortier, F. & Estoup, A. (2005b) Geneland: a computer package for landscape genetics. *Molecular Ecology Notes*, **5**, 712–715.

Guillot, G. & Rousset, F. (2013) Dismantling the Mantel tests. *Methods in Ecology and Evolution*, **4**, 336–344.

Guillot, G. & Santos, F. (2010) Using AFLP markers and the Geneland program for the inference of population genetic structure. *Molecular Ecology Resources*, **10**, 1082–1084.

Guillot, G., Santos, F. & Estoup, A. (2008) Analysing georeferenced population genetics data with Geneland: a new algorithm to deal with null alleles and a friendly graphical user interface. *Bioinformatics*, **24**, 1406–1407.

Guo, S.W. & Thompson, E.A. (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, **48**, 361–372.

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

Hamilton, W.D. & May, R. (1977) Dispersal in stable habitats. *Nature*, **269**, 578–581.

Haney, J. (1988) Diel patterns of zooplankton behavior. *Bulletin of Marine Science*, **43**, 583–603.

Hardy, O. & Vekemans, X. (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, 618–620.

Hardy, O.J., Charbonnel, N., Fréville, H. & Heuertz, M. (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics*, **163**, 1467–1482.

Harengerd, M., Kölsch, G. & Küsters, K. (1990) Dokumentation der Schwimmvogelzählung in der Bundesrepublik Deutschland 1966 - 1986. *Schriftenreihe des Dachverbandes Deutscher Avifaunisten*, **11**, 1–179.

Hartl, D.L. & Clark, A.G. (2007) *Principles of Population Genetics*. 4th edn. Sinauer, Sunderland, MA.

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

Hebert, P.D.N., Ratnasingham, S. & DeWaard, J.R. (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of The Royal Society B: Biological Sciences*, **270 Suppl.**, S96–599.

Hedrick, P.W. (1999) Highly variable loci and their interpretation in evolution and conservation. *Evolution*, **53**, 313–318.

Hernández, A. (2011) Internal dispersal of seed-inhabiting insects by vertebrate frugivores: a review and prospects. *Integrative Zoology*, **6**, 213–221.

Hernández, Á. & Falcó, J.V. (2008) Frugivorous birds dispersing braconid parasitoids via endozoochory. *Entomological Science*, **11**, 323–326.

Hewitt, G. (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.

Hewitt, G. (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.

Hewitt, G. (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–13.

Hewitt, G.M. (2004) The structure of biodiversity - insights from molecular phylogeography. *Frontiers in Zoology*, **1**, 1–16.

Higgins, K. & Lynch, M. (2001) Metapopulation extinction caused by mutation accumulation. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 2928–2933.

Holm, S. (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, **6**, 65–70.

Horn, M.H. (1997) Evidence for dispersal of fig seeds by the fruit-eating characid fish *Brycon guatemalensis* Regan in a Costa Rican tropical rain forest. *Oecologia*, **109**, 259–264.

Hughes, J.M., Schmidt, D.J. & Finn, D.S. (2009) Genes in streams: Using DNA to understand the movement of freshwater fauna and their riverine habitat. *BioScience*, **59**, 573–583.

Hurst, G.D.D. & Jiggins, F.M. (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of The Royal Society B: Biological Sciences*, **272**, 1525–1534.

Huson, D.H. & Bryant, D. (2006) Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.

Jäckel, R., Mora, D. & Dobler, S. (2013) Evidence for selective sweeps by Wolbachia infections: phylogeny of *Altica* leaf beetles and their reproductive parasites. *Molecular Ecology*, **16**, 4241–4255.

Kalinowski, S.T. (2009) How well do evolutionary trees describe genetic relationships among populations? *Heredity*, **102**, 506–513.

Kawata, M., Sawada, H., Yokoyama, J., Echenique-Dıaz, L.M. & Ishibashi, Y. (2005) The effects of spatial habitat structure on the evolution of density-dependent growth and reproduction in freshwater snails. *Hydrobiologia*, **533**, 229–241.

King, R.A., Gornall, R.J., Preston, C.D. & Croft, J.M. (2002) Population differentiation of *Potamogeton pectinatus* in the Baltic Sea with reference to waterfowl dispersal. *Molecular Ecology*, **11**, 1947–1956.

Kirby, J.S., Stattersfield, A.J., Butchart, S.H.M., Evans, M.I., Grimmett, R.F.A., Jones, V.R., *et al.* (2008) Key conservation issues for migratory land- and waterbird species on the world's major flyways. *Bird Conservation International*, **18**, 49 – 73.

Kokko, H. & López-Sepulcre, A. (2006) From individual dispersal to species ranges: perspectives for a changing world. *Science*, **313**, 789–791.

Kölsch, G. & Krause, A. (2011) Oxygen consumption of the aquatic leaf beetles *Macroplea mutica* and *Macroplea appendiculata* is low and not influenced by salinity. *Physiological Entomology*, **36**, 111–119.

Kölsch, G., Krause, A., Goetz, N. & Plagmann, S. (2010) The salinity preference of members of the genus *Macroplea* (Coleoptera, Chrysomelidae, Donaciinae), fully aquatic leaf beetles that occur in brackish water. *Journal of Experimental Marine Biology and Ecology*, **390**, 203–209.

Kölsch, G. & Kubiak, M. (2011) The aquatic leaf beetle species *Macroplea mutica* and *M. appendiculata* (Coleoptera, Chrysomelidae, Donaciinae) differ in their use of *Myriophyllum spicatum* as a host plant. *Aquatic Insects*, **33**, 13–26.

Kölsch, G., Matz-Grund, C. & Pedersen, B.V. (2009) Ultrastructural and molecular characterization of endosymbionts of the reed beetle genus *Macroplea* (Chrysomelidae, Donaciinae), and proposal of "*Candidatus Macropleicola appendiculatae*" and "*Candidatus Macropleicola muticae*." *Canadian Journal of Microbiology*, **55**, 1250–1260.

Kölsch, G. & Pedersen, B. V. (2010) Can the tight co-speciation between reed beetles (Col., Chrysomelidae, Donaciinae) and their bacterial endosymbionts, which provide cocoon material, clarify the deeper phylogeny of the hosts? *Molecular Phylogenetics and Evolution*, **54**, 810–821.

Kölsch, G. & Pedersen, B.V. (2008) Molecular phylogeny of reed beetles (Col., Chrysomelidae, Donaciinae): The signature of ecological specialization and geographical isolation. *Molecular Phylogenetics and Evolution*, **48**, 936–952.

Kölsch, G., Pedersen, B.V. & Biström, O. (2006) Species delimitation in the leaf beetle genus *Macroplea* (Coleoptera, Chrysomelidae) based on mitochondrial DNA, and phylogeographic considerations. *Insect Systematics & Evolution*, **37**, 467–479.

Korner-Nievergelt, F., Sauter, A., Atkinson, P.W., Guélat, J., Kania, W., Kéry, M., *et al.* (2010) Improving the analysis of movement data from marked individuals through explicit estimation of observer heterogeneity. *Journal of Avian Biology*, **41**, 8–17.

Kozak, K.H., Graham, C.H. & Wiens, J.J. (2008) Integrating GIS-based environmental data into evolutionary biology. *Trends in Ecology & Evolution*, **23**, 141–148.

Krüger, T. & Garthe, S. (2001) Flight altitudes of coastal birds in relation to wind direction and speed. *Atlantic Seabirds*, **3**, 203–216.

Latch, E.K., Dharmarajan, G., Glaubitz, J.C. & Rhodes, O.E. (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, **7**, 295–302.

Laux, J.-J. & Kölsch, G. (2014) Potential for passive internal dispersal: eggs of an aquatic leaf beetle survive passage through the digestive system of mallards. *Ecological Entomology*, **39**, 391–394.

Leeuwen, C.H.A. Van. (2012) Speeding up the snail's pace: bird-mediated dispersal of aquatic organisms. *PhD thesis, Radboud University Nijmegen, Nijmegen, The Netherlands*.

Leeuwen, C.H.A. Van, Huig, N., Velde, G. Van der, Alen, T.A. Van, Wagemaker, C.A.M., Sherman, C.D.H., *et al.* (2013) How did this snail get here? Several dispersal vectors inferred for an aquatic invasive species. *Freshwater Biology*, **58**, 88–99.

Leeuwen, C.H.A. Van, Tollenaar, M.L. & Klaassen, M. (2012a) Vector activity and propagule size affect dispersal potential by vertebrates. *Oecologia*, **170**, 101–109.

Leeuwen, C.H.A. Van, Velde, G. Van der, Groenendael, J.M. Van & Klaassen, M. (2012b) Gut travellers: internal dispersal of aquatic organisms by waterfowl. *Journal of Biogeography*, **39**, 2031–2040.

Leeuwen, C.H.A. Van, Velde, G. Van der, Lith, B. Van & Klaassen, M. (2012c) Experimental quantification of long distance dispersal potential of aquatic snails in the gut of migratory birds. *PloS ONE*, **7**, e32292.

Legendre, P. & Fortin, M.-J. (2010) Comparison of the Mantel test and alternative approaches for detecting complex multivariate relationships in the spatial analysis of genetic data. *Molecular Ecology Resources*, **10**, 831–844.

Leppäranta, M. & Myrberg, K. (2009) *Physical Oceanography of the Baltic Sea*. Springer, Berlin.

Lou, Q., Yu, P. & Liang, H. (2011) Two new species of *Macroplea* Samouelle (Coleoptera: Chrysomelidae: Donaciinae) from China, with a key to all known species. *Zootaxa*, **3003**, 1–21.

Lugon-Moulin, N., Brünner, H., Wyttenbach, A., Hausser, J. & Goudet, J. (1999) Hierarchical analyses of genetic differentiation in a hybrid zone of *Sorex araneus* (Insectivora: Soricidae). *Molecular Ecology*, **8**, 419–431.

Mader, E., Vierssen, W. van & Schwenk, K. (1998) Clonal diversity in the submerged macrophyte *Potamogeton pectinatus* L. inferred from nuclear and cytoplasmic variation. *Aquatic Botany*, **62**, 147–160.

Manel, S., Schwartz, M.K., Luikart, G. & Taberlet, P. (2003) Landscape genetics: combining landscape ecology and population genetics. *Trends in Ecology & Evolution*, **18**, 189–197.

Mann, J.S. & Crowson, R.A. (1983) Observations on the internal anatomy and classification of Donaciinae (Col., Chrysomelidae). *Entomologists' Monthly Magazine*, **119**, 17–27.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Martinez-Haro, M., Taggart, M.A., Green, A.J. & Mateo, R. (2009) Avian digestive tract simulation to study the effect of grit geochemistry and food on Pb shot bioaccessibility. *Environmental Science & Technology*, **43**, 9480–9486.

McKinnon, L., Smith, P.A., Nol, E., Martin, J.L., Doyle, F.I., Abraham, K.F., *et al.* (2010) Lower predation risk for migratory birds at high latitudes. *Science*, **327**, 326–327.

Medvedev, L.N. (2006) Contributions to the knowledge of the leaf beetles (Coleoptera, Chrysomelidae) of Russia. *Entomological Review*, **86**, 457–459.

Meester, L. De, Gómez, A., Okamura, B. & Schwenk, K. (2002) The Monopolization Hypothesis and the dispersal–gene flow paradox in aquatic organisms. *Acta Oecologica*, **23**, 121–135.

Meirmans, P.G. (2012) The trouble with isolation by distance. *Molecular Ecology*, **21**, 2839–2846.

Mende, M., Bistrom, O., Meichssner, E. & Kölsch, G. (2010) The aquatic leaf beetle *Macroplea mutica* (Coleoptera: Chrysomelidae) in Europe: Population structure, postglacial colonization and the signature of passive dispersal. *European Journal of Entomology*, **107**, 101–113.

Michalakis, Y. & Excoffier, L. (1996) A generic estimation of population subdivision using distances between alleles with special reference for microsatellite loci. *Genetics*, **142**, 5–8.

Miller, M., Blinn, D. & Keim, P. (2002) Correlations between observed dispersal capabilities and patterns of genetic differentiation in populations of four aquatic insect species from the Arizona White Mountains, U.S.A. *Freshwater Biology*, **47**, 1660–1673.

Mills, L. & Allendorf, F. (1996) The one-migrant-per-generation rule in conservation and management. *Conservation Biology*, **10**, 1509 – 1518.

Mohr, K.-H. (1985) Beiträge zur Insektenfauna der DDR: Coleoptera - Chrysomelidae: Donaciinae, Orsodacninae, Criocerinae. Clythrinae. *Beiträge zur Entomologie*, **35**, 219–262.

Nastis, S. & Malechek, J.C. (1988) Estimating digestibility of oak browse diets for goats by in vitro techniques. *Journal of Range Management*, **41**, 255–258.

Nathan, R. (2006) Long-distance dispersal of plants. *Science*, **313**, 786–788.

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

Nei, M., Tajima, F. & Tateno, Y. (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, **19**, 153–170.

Nies, G. & Reusch, T.B.H. (2005) Evolutionary divergence and possible incipient speciation in post-glacial populations of a cosmopolitan aquatic plant. *Journal of Evolutionary Biology*, **18**, 19–26.

Nilsson, L. (1975) Midwinter distribution and numbers of Swedish Anatidae. *Ornis Scandinavica*, **6**, 83–107.

Ocumpaugh, W.R. & Swakon, D.H.D. (1993) Simulating grass seed passage through the digestive system of cattle: a laboratory technique. *Crop Science*, **33**, 1084–1090.

Oddou-Muratorio, S., Vendramin, G.G., Buiteveld, J. & Fady, B. (2008) Population estimators or progeny tests: what is the best method to assess null allele frequencies at SSR loci? *Conservation Genetics*, **10**, 1343–1347.

Oosterhout, C. Van, Hutchinson, W.F., Wills, D.P.M. & Shipley, P. (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

Orsini, L., Mergeay, J., Vanoverbeke, J. & Meester, L. De. (2013a) The role of selection in driving landscape genomic structure of the waterflea *Daphnia magna*. *Molecular Ecology*, **22**, 583–601.

Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J. & Meester, L. De. (2013b) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, **22**, 5983–5999.

Page, R.D. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Computer applications in the biosciences : CABIOS*, **12**, 357–358.

Papadopoulou, A., Anastasiou, I. & Vogler, A.P. (2010) Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution*, **27**, 1659–1672.

Peakall, R. & Smouse, P. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics*, **1**, 6–8.

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.

Pearson, R.G. (2006) Climate change and the migration capacity of species. *Trends in Ecology & Evolution*, **21**, 111–113.

Pedall, I., Fritz, U., Stuckas, H., Valdeón, A. & Wink, M. (2011) Gene flow across secondary contact zones of the *Emys orbicularis* complex in the Western Mediterranean and evidence for extinction and re-introduction of pond turtles on Corsica and Sardinia (Testudines: Emydidae). *Journal of Zoological Systematics and Evolutionary Research*, **49**, 44–57.

Pérez-Espona, S., Pérez-Barbería, F.J., McLeod, J.E., Jiggins, C.D., Gordon, I.J. & Pemberton, J.M. (2008) Landscape features affect gene flow of Scottish Highland red deer (*Cervus elaphus*). *Molecular Ecology*, **17**, 981–996.

Phillipsen, I.C. & Lytle, D. a. (2013) Aquatic insects in a sea of desert: population genetic structure is shaped by limited dispersal in a naturally fragmented landscape. *Ecography*, **36**, 731–743.

Pimentel, D., Zuniga, R. & Morrison, D. (2005) Update on the environmental and economic costs associated with alien-invasive species in the United States. *Ecological Economics*, **52**, 273–288.

Pollux, B.J. a. (2011) The experimental study of seed dispersal by fish (ichthyochory). *Freshwater Biology*, **56**, 197–212.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Quantum GIS Development Team. (2012) Quantum GIS Geographic Information System. *http://qgis.osgeo.org*.

R Core Team. (2013) R: A language and environment for statistical computing. *http://www.r-project.org/*.

Raufaste, N., Rousset, F., Ii, U.M., Bataillon, P.E. & Cedex, M. (2014) Are partial Mantel tests adequate ? *Evolution*, **55**, 1703–1705.

Raymond, M. & Rousset, F. (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **68**, 248–249.

Robertson, A. & Hill, W.G. (1984) Deviations from Hardy-Weinberg proportions: sampling variances and use in estimation of inbreeding coefficients. *Genetics*, **107**, 703–718.

Rousset, F. (1996) Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics*, **142**, 1357–1362.

Rousset, F. (2008) GENEPOP'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.

Rousset, F. (2014) Partial Mantel tests : Reply to Castellano and Balletto. *Evolution*, **56**, 1874–1875.

Rowe, G. & Beebee, T.J.C. (2007) Defining population boundaries: use of three Bayesian approaches with microsatellite data from British natterjack toads (*Bufo calamita*). *Molecular Ecology*, **16**, 785–796.

Rowell, H.E. & Spray, C.J. (2004) *The Mute swan* Cygnus olor*: (Britain and Ireland populations) in Britain and Northern Ireland 1960/61-2000/01*. Waterbird Review Series, The Wildfowl & Wetlands Trust/Joint Nature Conservation Committee, Slimbridge.

Rozen, S. & Skaletsky, H. (2000) Primer3 on the WWW for general users and for biologist programmers. In *Bioinformatics Methods and Protocols: Methods in Molecular Biology* (ed. Krawetz, S. & Krawetz, S.A.). Humana Press Inc., Totowa (NJ), pp. 365–386.

Ryman, N. & Palm, S. (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes*, **6**, 600–602.

Saari, S. (2007) Meriuposkuoriaisen, *Macroplea pubipennis* (Coleoptera: Chrysomelidae), levinneisyys ja elinympäristövaatimukset Espoonlahdessa. *Master Thesis, University of Helsinki, Helsinki, Finland*, 55.

Saitou, N. & Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Sánchez, M.I., Hortas, F., Figuerola, J. & Green, A.J. (2012) Comparing the potential for dispersal via waterbirds of a native and an invasive brine shrimp. *Freshwater Biology*, **57**, 1896–1903.

Santamaría, L. (2002) Why are most aquatic plants widely distributed? Dispersal, clonal growth and small-scale heterogeneity in a stressful environment. *Acta Oecologica*, **23**, 137–154.

Santamaría, L. & Klaassen, M. (2002) Waterbird-mediated dispersal of aquatic organisms: an introduction. *Acta Oecologica*, **23**, 115–119.

Schmitt, T. (2007) Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in Zoology*, **4**, 11.

Schmitt, T., Gießl, A. & Seitz, A. (2002) Postglacial colonisation of western Central Europe by *Polyommatus coridon* (Poda 1761)(Lepidoptera: Lycaenidae): evidence from population genetics. *Heredity*, **88**, 26–34.

Schmitt, T., Röber, S. & Seitz, A. (2005) Is the last glaciation the only relevant event for the present genetic population structure of the Meadow Brown butterfly *Maniola jurtina* (Lepidoptera: Nymphalidae)? *Biological Journal of the Linnean Society*, **85**, 419–431.

Schwartz, M.K. & McKelvey, K.S. (2008) Why sampling scheme matters: the effect of sampling scheme on landscape genetic results. *Conservation Genetics*, **10**, 441–452.

Selkoe, K.A. & Toonen, R.J. (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters*, **9**, 615–29.

Silfverberg, H. (2010) Subfamily Donaciinae. In *Catalogue of Palearctic Coleoptera (Vol. 6)* (ed. Löbl, I. & Smetana, A.). Apollo Books, Stenstrup, pp. 354–358.

Silva, P.I.T., Martins, A.M., Gouvea, E.G., Pessoa-Filho, M. & Ferreira, M.E. (2013) Development and validation of microsatellite markers for *Brachiaria ruziziensis* obtained by partial genome assembly of Illumina single-end reads. *BMC Genomics*, **14**, 17.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. & Flook, P. (1994) Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.

Slatkin, M. (1995) A measure of population subdivision based on microsatellite allele frequencies. *Genetics*, **139**, 457–462.

Smee, M.R., Pauchet, Y., Wilkinson, P., Wee, B., Singer, M.C., French-Constant, R.H., *et al.* (2013) Microsatellites for the marsh fritillary butterfly: *de novo* transcriptome sequencing, and a comparison with amplified fragment length polymorphism (AFLP) markers. *PloS ONE*, **8**, e54721.

Sorensen, A. (1986) Seed dispersal by adhesion. *Annual Review of Ecology and Systematics*, **17**, 443–463.

Spear, S.F., Peterson, C.R., Matocq, M.D. & Storfer, A. (2005) Landscape genetics of the blotched tiger salamander (*Ambystoma tigrinum melanostictum*). *Molecular Ecology*, **14**, 2553–2564.

Spurgin, L.G., Illera, J.C., Jorgensen, T.H., Dawson, D.A. & Richardson, D.S. (2014) Genetic and phenotypic divergence in an island bird: isolation by distance, by colonization or by adaptation? *Molecular Ecology*, **23**, 1028–1039.

Squirrell, J., Hollingsworth, P.M., Woodhead, M., Russell, J., Lowe, A.J., Gibby, M., *et al.* (2003) How much effort is required to isolate nuclear microsatellites from plants? *Molecular Ecology*, **12**, 1339–1348.

Srivastava, D. & Kratina, P. (2013) Is dispersal limitation more prevalent in the ocean? *Oikos*, **122**, 298–300.

Stafford, J.D., Eichholz, M.W. & Phillips, A.C. (2012) Impacts of Mute Swans (*Cygnus olor*) on Submerged Aquatic Vegetation in Illinois River Valley Backwaters. *Wetlands*, **32**, 851–857.

Steiniger, S. & Hay, G.J. (2009) Free and open source geographic information tools for landscape ecology. *Ecological Informatics*, **4**, 183–195.

Steiniger, S. & Hunter, A.J.S. (2012) The 2012 free and open source GIS software map – A guide to facilitate research, development, and adoption. *Computers, Environment and Urban Systems*, **39**, 136–150.

Strayer, D.L. (2010) Alien species in fresh waters: ecological effects, interactions with other stressors, and prospects for the future. *Freshwater Biology*, **55**, 152–174.

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.

Takezaki, N. & Nei, M. (1996) Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics*, **144**, 389–399.

Takezaki, N. & Nei, M. (2008) Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. *Genetics*, **178**, 385–392.

Theissinger, K., Bálint, M., Feldheim, K.A., Haase, P., Johannesen, J., Laube, I., *et al.* (2013) Glacial survival and post-glacial recolonization of an arctic-alpine freshwater insect (*Arcynopteryx dichroa* , Plecoptera, Perlodidae) in Europe. *Journal of Biogeography*, **40**, 236–248.

Thomas, M.C. & Lana, P.C. (2011) A new look into the small-scale dispersal of free-living marine nematodes. *Zoologia*, **28**, 449–456.

Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Thorpe, W. & Crisp, D.J. (1949) Studies on plastron respiration IV. Plastron respiration in the Coleoptera. *Journal of Experimental Biology*, **26**, 219–261.

Tilley, J. & Terry, R. (1963) A two-stage technique for the *in vitro* digestion of forage crops. *Grass and Forage Science*, **18**, 104–111.

Triest, L. & Sierens, T. (2013) Is the genetic structure of Mediterranean *Ruppia* shaped by bird-mediated dispersal or sea currents? *Aquatic Botany*, **104**, 176–184.

Türkgülü, İ., Ekiz, A.N., Gök, A. & Şen, B. (2011) The first representative of the fully aquatic leaf beetle genus *Macroplea* Samouelle, 1819 (Coleoptera, Chrysomelidae) in Turkey: *Macroplea mutica* (Fabricius, 1792), with notes on its biology, habitat, host plant and distribution. *Zoosystematics and Evolution*, **87**, 291–295.

Vanschoenwinkel, B., Waterkeyn, A., Nhiwatiwa, T., Pinceel, T., Spooren, E., Geerts, A., *et al.* (2011) Passive external transport of freshwater invertebrates by elephant and other mud-wallowing mammals in an African savannah habitat. *Freshwater Biology*, **56**, 1606–1619.

Vilà, M., Espinar, J.L., Hejda, M., Hulme, P.E., Jarošík, V., Maron, J.L., *et al.* (2011) Ecological impacts of invasive alien plants: a meta-analysis of their effects on species, communities and ecosystems. *Ecology Letters*, **14**, 702–708.

Wada, S., Kawakami, K. & Chiba, S. (2012) Snails can survive passage through a bird's digestive system. *Journal of Biogeography*, **39**, 69–73.

Wang, J. (2004) Application of the one-migrant-per-generation rule to conservation and management. *Conservation Biology*, **18**, 332–343.

Waterkeyn, A., Pineau, O., Grillas, P. & Brendonck, L. (2010) Invertebrate dispersal by aquatic mammals: a case study with nutria *Myocastor coypus* (Rodentia, Mammalia) in Southern France. *Hydrobiologia*, **654**, 267–271.

Weir, B.S. & Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Westberg, E. & Kadereit, J.W. (2009) The influence of sea currents, past disruption of gene flow and species biology on the phylogeographical structure of coastal flowering plants. *Journal of Biogeography*, **36**, 1398–1410.

Whittaker, J.C., Harbord, R.M., Boxall, N., Mackay, I., Dawson, G. & Sibly, R.M. (2003) Likelihood-based estimation of microsatellite mutation rates. *Genetics*, **164**, 781–787.

Wiley, E. (1988) Vicariance biogeography. *Annual Review of Ecology and Systematics*, **19**, 513–542.

Wright, S. (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

Wright, S. (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Wright, S. (1977) *Evolution and the Genetics of Population, Experimental Results and Evolutionary Deductions*. The University of Chicago Press, Chicago.

Zane, L., Bargelloni, L. & Patarnello, T. (2002) Strategies for microsatellite isolation: a review. *Molecular Ecology*, **11**, 1–16.

Zar, J.H. (1999) *Biostatistical Analyses*. 4th edn. Prentice Hall, Upper Saddle River, NJ.

Zhang, J., Wheeler, G., Purcell, M. & Ding, J. (2010) Biology, distribution, and field host plants of *Macroplea japana* in China: an unsuitable candidate for biological control of *Hydrilla verticillata*. *Florida Entomologist*, **93**, 116–119.

*** European Commission, 1992 - Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora (online). Official Journal L 206, 22/07/1992 P. 0007-0050. Consolidated version 01.01.2007. Available at: http://ec.europa.eu/environment/nature/legislation/habitatsdirective/index_en.htm. Accessed October 9th, 2014.

Publication

Laux, J.-J. & Kölsch, G. (2014) Potential for passive internal dispersal: eggs of an aquatic leaf beetle survive passage through the digestive system of mallards. *Ecological Entomology*, **39**, 391–394

DOI: 10.1111/een.12097

Acknowledgements

This work was supported by the Estuary and Wetland Research Graduate School Hamburg (ESTRADE) as member of the State Excellence Initiative (LExI) funded by the Hamburg Science and Research Foundation.

Completion of this project would not have been possible without the support of many people. First and foremost the author wishes to express his gratitude to his supervisor, PD Dr. Gregor Kölsch, who was abundantly helpful and patiently offered invaluable assistance, support and guidance.

Deepest gratitude is also due to Prof. Dr. Susanne Dobler for providing the opportunity to work and graduate at the department of Molecular Evolutionary Biology and for constant kind and invaluable support.

The author furthermore wants to show greatest appreciation to Dr. Bernd Haenfling for his help with developing microsatellite markers and the opportunity to spend an immensely instructive time at Hull University.

The author is also very grateful to Dr. Veit Hennig for acting as a third graduation advisor and for his help with the GIS-based analyses.

Special thanks also go to everyone in the department of Molecular Evolutionary Biology for being such helpful and pleasant colleagues (in alphabetical order): Dr. Christian Baden, Michael Baum, Dr. Christiane Bramer, Safaa Dalla, Dr. Regina Jäckel, Juliette Kober, Saskia Liberei, Karin Meyer, Dr. Julia Offe, Dr. Georg Petschenka, Dr. Renja Romey-Glüsing, Semra Ünsal, Vera Wagschal and Samuel Waldron.

The author also wants to thank everyone who contributed to location and sampling of beetles in the field, especially Gert van Ee for his generous support in locating *M. mutica* in the Netherlands and Ad Bouman for invaluable help in sampling. The author is furthermore very grateful to Andrea Simon for helping with sampling in Yorkshire, UK.

The author also thanks everyone at ESTRADE graduate school, especially Prof. Dr. Kai Jensen and Dr. Tobias Gebauer.

Thanks also go to Dr. Gabriele Ismer and Dr. Silke Plagmann at Tierpark Gettorf, for their help with early attempts at conducting feeding trials and to Antje Schnasse und Dr. Anke Heisig from the department of Pharmaceutical Biology and Microbiology for help with the microsatellite electrophoresis.

The author furthermore thanks Chris du Feu from the European Union for Bird Ringing and all ringing schemes contributing to the used mute swan data set, especially the Dutch ringing centre in Arnhem.

Last but not least, the author wishes to express his love and gratitude to his friends and family.

Appendices

Appendix A - Detected Microsatellite Alleles

Table 23. Detected Microsatellite Alleles

Detected alleles for six microsatellite loci and 256 sampled *M. mutica* individuals. Each allele is represented by a three-digit number that equals detected PCR product lengths in base pairs. Sample abbreviations are listed in Table 2.

	loc16359		loc4107		loc12208		loc321		loc3012		loc1624	
Sample												
WAL4	151	151	183	185	204	204	389	389	163	163	386	389
WAL5	151	151	185	185	204	204	389	389	199	217	386	386
WAL6	151	151	185	185	204	204	389	395	163	229	386	386
WAL7	151	151	185	185	196	204	389	389	163	163	386	389
WAL8	151	151	183	185	196	196	389	398	163	241	386	389
WAL9	151	151	183	185	204	204	386	389	163	163	386	386
WAL10	151	151	185	185	196	204	389	389	229	247	386	386
WAL12	151	151	185	185	196	196	389	389	163	229	386	386
WAL13	151	151	185	185	196	196	389	389	163	163	386	386
WAL15	151	151	183	183	202	204	386	389	163	163	386	389
WAL16	151	151	185	185	196	204	386	386	163	163	386	386
WAL17	151	151	185	185	196	204	389	389	163	163	386	386
WAL18	151	151	185	185	196	204	389	389	241	250	386	389
WAL19	127	151	185	185	196	204	389	389	163	163	386	386
WAL20	151	151	185	185	196	204	386	386	163	163	386	389
YOR1	147	151	185	205	196	196	389	395	160	160	386	386
YOR2	147	151	185	185	196	204	395	398	160	244	386	389
YOR3	147	151	185	185	196	204	386	395	160	247	386	386
YOR4	147	151	185	185	196	204	275	386	160	247	386	386
YOR5	147	151	185	185	196	204	389	392	160	160	386	386
YOR6	131	151	185	185	194	196	398	401	160	160	398	401
YOR7	147	151	185	185	196	196	386	392	244	244	386	386
YOR8	147	151	185	185	196	196	398	401	160	244	386	386
YOR9	127	127	197	197	204	204	398	401	247	247	386	389
YOR10	147	151	185	185	194	196	389	398	160	244	386	389
YOR11	147	151	185	185	196	196	398	401	244	247	386	389
YOR12	127	127	433	433	196	196	392	398	160	247	386	389
YOR13	135	135	185	185	194	196	395	398	160	229	386	389
YOR14	147	151	185	205	196	196	386	389	160	247	386	389
YOR15	151	151	183	183	196	204	386	389	208	244	383	383
CAM1	147	151	183	183	194	194	383	383	226	226	386	389
CAM2	147	151	183	185	204	204	383	383	160	223	389	389
CAM3	151	151	185	185	194	204	383	386	223	226	389	389

Sample	loc16359		loc4107		loc12208		loc321		loc3012		loc1624	
CAM4	147	151	183	183	196	204	386	401	226	244	386	386
CAM5	151	167	183	185	196	196	389	398	160	226	386	386
CAM6	151	167	183	183	194	196	395	398	160	226	389	389
CAM7	151	151	183	183	194	196	392	392	160	226	386	389
CAM8	151	167	183	183	194	194	395	401	160	226	386	389
CAM9	127	151	183	183	196	196	386	398	226	226	386	389
CAM10	151	151	183	183	194	196	389	398	226	226	386	386
CAM11	151	151	185	185	196	204	392	398	226	226	389	389
CAM12	151	151	183	185	196	204	386	389	226	226	386	389
CAM13	151	151	183	183	194	196	383	389	226	226	386	389
CAM14	151	151	183	183	196	196	386	392	160	160	386	389
CAM15	151	151	183	183	194	204	395	395	223	226	386	389
ESU1	151	151	183	185	194	204	389	389	163	229	386	389
ESU2	151	151	183	185	194	194	398	398	163	232	386	389
ESU3	147	151	183	183	194	204	389	395	229	232	386	389
ESU4	151	151	183	183	194	204	395	395	229	229	386	389
ESU5	151	151	183	205	184	204	389	398	163	247	386	389
ESU6	151	151	183	183	196	204	386	389	163	163	386	389
ESU7	147	151	183	183	194	196	386	389	163	229	389	389
ESU8	147	147	183	183	194	204	392	392	163	229	389	389
ESU9	151	151	183	183	196	204	386	389	163	250	386	386
ESU10	151	151	183	183	194	196	383	398	229	232	386	389
ESU11	151	151	183	183	196	204	389	389	163	163	386	386
ESU12	147	147	183	185	202	204	398	398	229	229	389	389
ESU13	147	155	183	183	196	204	395	398	163	229	389	389
ESU14	151	151	183	183	196	196	389	398	163	163	389	389
ESU15	147	151	183	183	194	204	395	395	163	229	386	389
HOL2	127	151	183	205	204	204	398	398	226	226	386	386
HOL4	151	155	183	183	194	204	275	275	160	160	380	389
HOL13	147	151	183	183	194	204	398	404	160	160	380	389
HOL18	127	127	183	183	192	204	389	389	160	160	386	386
HOL21	147	151	183	205	194	194	383	404	160	160	386	389
HOL22	147	151	183	183	194	204	398	404	160	160	386	386
HOL31	151	151	185	185	196	204	389	389	160	160	380	380
HOL32	127	127	183	183	194	204	389	398	226	229	386	389
NFL1	151	151	185	205	180	196	386	398	163	163	389	389
NFL2	127	127	183	183	180	196	389	395	229	244	386	386
NFL3	147	147	183	183	196	196	395	395	193	226	389	389
NFL4	147	167	183	183	194	204	389	395	232	241	386	389
NFL5	127	131	183	183	196	204	395	395	163	163	389	389
NFL6	143	143	183	183	196	196	386	395	163	244	386	389
NFL9	151	151	183	183	194	204	398	398	163	226	389	389
NFL10	127	151	183	183	194	204	395	395	241	241	386	389
NFL11	151	167	183	183	204	204	395	395	163	226	380	386
NFL12	151	151	183	185	180	204	395	398	163	241	380	389

Sample	loc16359		loc4107		loc12208		loc321		loc3012		loc1624	
NFL13	127	155	183	183	196	204	395	395	247	250	389	389
NFL14	151	167	183	205	180	180	395	395	163	226	386	389
NFL15	151	167	183	183	194	194	398	398	163	229	386	389
NFS1	147	147	183	183	204	204	386	386	163	163	386	389
NFS2	147	151	183	183	204	204	389	398	226	244	386	389
NFS3	147	151	183	183	196	204	386	398	163	163	386	386
NFS4	151	151	183	183	196	204	389	395	226	226	386	386
NFS5	147	151	183	183	196	196	395	398	151	163	386	389
NFS6	127	151	183	183	196	204	383	404	268	286	386	389
NFS7	127	151	183	205	194	196	386	395	163	163	386	389
NFS8	127	127	183	183	196	204	386	389	181	181	389	389
NFS9	127	151	183	183	194	204	386	389	154	163	386	389
NFS10	127	127	183	183	196	204	386	404	190	190	386	386
NFS11	151	151	183	183	194	204	386	386	154	247	386	386
NFS12	147	151	183	183	194	204	389	395	241	241	386	386
NFS13	127	151	183	205	194	194	386	386	163	247	389	389
NFS14	151	151	183	183	194	204	383	383	247	247	386	389
NFS15	147	151	183	183	194	204	377	377	163	163	386	386
PLW1	147	151	183	183	194	194	389	407	229	247	386	389
PLW2	127	127	183	205	194	196	395	395	163	163	386	389
PLW3	151	151	183	205	194	196	395	404	163	247	386	389
PLW4	127	147	183	183	194	194	389	395	163	163	386	389
PLW5	127	147	183	183	194	204	389	395	226	247	386	386
PLW6	127	151	183	183	196	196	392	404	163	241	386	386
PLW7	147	147	183	183	194	196	389	398	193	220	386	386
PLW8	167	167	183	183	196	196	389	404	163	169	386	386
PLW9	147	151	183	183	204	204	389	395	163	247	386	386
PLW10	151	151	183	183	196	204	392	386	244	247	386	386
PLW11	127	151	183	183	194	204	404	404	163	247	386	389
PLW12	143	151	183	183	194	196	395	395	193	193	386	386
PLW13	127	147	183	183	196	204	389	395	193	193	386	386
PLW14	127	147	183	183	196	196	389	392	163	163	389	389
PLW15	127	151	183	205	194	194	401	404	163	247	386	389
JUN1	127	127	183	185	180	192	395	398	163	241	389	389
JUN2	127	135	185	185	180	194	395	401	241	241	386	389
JUN3	127	127	183	183	180	180	395	395	226	226	389	389
JUN4	127	127	183	183	194	194	386	395	163	229	389	389
JUN5	127	167	193	195	180	192	404	407	163	226	389	389
JUN6	127	151	185	195	192	206	398	395	226	226	389	389
JUN7	127	151	183	185	180	180	386	386	232	241	389	389
JUN8	127	151	185	185	192	192	395	395	241	244	389	389
JUN9	127	151	183	183	194	194	383	383	226	232	386	389
JUN10	151	155	183	195	180	192	395	398	163	241	377	389
JUN11	127	127	183	195	180	180	383	386	163	226	389	389
JUN12	139	179	193	203	190	202	395	395	226	226	380	389

Sample	loc16359		loc4107		loc12208		loc321		loc3012		loc1624	
JUN13	127	127	183	183	180	194	395	401	226	241	377	386
JUN14	127	127	183	185	192	202	395	395	163	226	386	389
JUN15	151	151	183	205	192	192	395	398	232	238	386	386
JUW 1	127	127	183	183	194	194	386	398	229	238	386	386
JUW 2	127	151	183	195	194	196	386	389	226	241	386	386
JUW 3	127	135	183	183	194	194	398	398	226	229	380	389
JUW 4	127	151	183	187	194	204	386	398	238	241	386	389
JUW 5	127	127	183	205	180	180	389	389	232	244	386	389
JUW 6	127	151	183	183	180	204	392	386	229	241	386	389
JUW 7	151	151	183	183	196	204	386	395	163	163	386	389
JUW 8	127	151	183	183	196	204	395	398	229	229	389	389
JUW 9	127	127	183	183	192	192	389	395	241	241	386	389
JUW 10	155	155	183	183	194	194	395	398	226	241	389	389
JUW 11	127	151	183	183	196	196	389	392	241	244	389	389
JUW 12	127	127	183	183	194	204	389	398	163	229	389	389
JUW 13	127	127	183	183	192	204	395	398	229	238	389	389
JUW 14	127	127	183	195	192	194	386	386	163	226	389	389
JUW 15	127	167	183	183	190	194	386	398	229	244	386	389
HEL1	127	151	185	205	194	196	386	386	232	238	386	389
HEL2	151	151	183	205	180	196	395	398	232	238	386	386
HEL3	127	151	185	205	194	196	386	395	238	238	386	389
HEL4	151	151	183	183	180	196	395	398	232	238	389	389
HEL5	127	151	205	205	180	194	395	395	238	238	386	386
HEL6	131	151	185	185	196	196	389	395	232	232	386	389
HEL7	131	151	183	183	196	196	395	398	163	163	386	389
HEL8	127	127	185	205	196	196	392	386	187	193	386	389
HEL9	127	151	183	205	194	194	395	398	232	235	386	389
HEL10	127	127	185	205	196	196	395	398	151	241	386	389
HEL11	127	151	183	257	194	194	395	398	157	238	386	389
LEM1	151	151	183	183	194	196	395	395	238	241	389	389
LEM2	127	131	183	205	196	196	389	398	229	241	386	389
LEM3	127	151	183	185	194	196	398	398	232	232	389	389
LEM4	127	151	183	205	196	200	278	386	232	238	386	386
LEM5	143	151	185	207	196	196	395	398	235	238	389	389
LEM6	143	143	183	183	180	196	380	395	145	232	386	389
LEM7	151	151	183	183	194	196	392	386	232	232	386	389
LEM8	143	143	183	183	196	196	395	398	232	235	386	389
LEM9	127	151	183	205	196	196	392	386	232	232	389	389
LEM10	139	167	183	183	194	194	278	395	226	232	389	389
LEM11	143	143	183	183	180	196	386	395	259	268	386	386
LEM12	131	151	183	183	180	194	386	395	166	226	389	389
LEM13	127	151	195	205	194	194	395	398	166	166	386	389
LEM14	139	151	183	205	194	196	386	395	229	238	389	389
LEM15	127	127	183	183	190	196	380	398	238	238	389	389
ORT11	147	147	183	205	180	194	278	278	154	163	389	389

Sample	loc16359		loc4107		loc12208		loc321		loc3012		loc1624	
ORT12	147	155	183	185	180	180	395	404	163	163	389	389
ORT17	151	151	185	205	194	196	278	395	163	163	389	389
ORT18	151	151	185	185	194	196	395	401	226	232	389	389
ORT24	151	155	185	205	180	196	395	410	238	238	416	416
ORT25	127	131	185	185	194	196	278	278	232	241	389	389
ORT26	131	155	185	205	194	196	389	395	232	235	386	389
ORT27	127	151	185	185	194	200	389	395	226	232	386	389
ORT28	127	127	183	185	196	196	386	401	229	241	383	389
ORT29	131	151	185	183	196	196	278	398	232	232	386	389
ORT30	131	143	185	205	196	196	278	389	238	238	389	389
ORT31	127	143	185	205	180	194	398	398	244	244	386	389
ORT33	143	151	183	205	196	196	401	407	232	235	389	389
ORT34	151	155	183	183	196	196	386	392	157	226	389	389
ORT35	127	155	183	183	194	200	386	392	226	226	389	389
OBJ1	127	127	205	205	180	194	386	395	238	238	389	389
OBJ2	127	127	183	183	194	194	278	278	193	211	389	389
OBJ3	151	151	195	195	194	198	386	386	232	232	389	389
OBJ4	151	151	183	183	194	196	386	395	235	238	389	389
OBJ5	143	151	185	195	194	204	386	386	226	232	389	407
OBJ6	127	127	185	205	194	204	386	395	268	271	389	389
OBJ7	131	151	183	185	178	194	386	395	238	244	386	389
OBJ8	127	127	183	195	180	194	395	395	229	238	389	389
OBJ9	139	151	185	185	194	200	386	401	232	232	386	389
OBJ10	127	127	185	185	194	200	395	398	223	235	389	389
OBJ11	127	127	195	195	180	200	278	278	229	244	386	389
OBJ12	139	151	183	183	180	194	278	278	232	238	386	389
OBJ13	131	151	183	195	194	200	386	395	226	241	386	389
OBJ14	131	131	183	207	192	196	386	395	232	238	386	386
OBJ15	127	151	195	205	194	194	278	278	232	235	389	389
RUG1	127	131	203	205	194	196	386	401	229	241	389	389
RUG2	143	151	195	205	180	198	386	386	172	181	389	389
RUG3	127	127	183	183	180	194	395	395	193	193	389	389
RUG4	127	151	183	195	180	194	386	395	232	238	389	389
RUG5	127	155	185	185	192	192	395	398	238	241	389	395
RUG6	127	131	183	183	180	180	392	395	226	238	389	389
RUG7	131	155	183	195	194	194	275	275	193	193	386	389
RUG8	127	127	185	185	180	196	275	275	229	235	389	389
RUG9	127	127	183	195	200	200	395	398	232	238	389	389
RUG10	155	155	183	183	194	196	395	395	175	184	386	389
RUG11	127	143	183	183	194	200	395	395	169	175	389	389
RUG12	127	135	183	205	194	194	398	398	193	193	389	389
RUG13	127	127	183	205	180	194	443	443	172	178	389	389
RUG14	143	143	183	205	194	196	395	398	193	193	389	389
RUG15	127	131	183	185	194	200	275	275	193	202	386	389
OST2	127	155	185	187	194	194	386	386	223	226	389	389

Sample	loc16359		loc4107		loc12208		loc321		loc3012		loc1624	
OST3	151	159	183	209	194	194	386	386	226	232	389	389
OST4	159	159	185	185	186	186	386	395	223	226	386	386
OST5	159	159	185	209	194	194	275	275	229	232	386	386
OST6	151	151	185	197	186	194	275	386	226	226	386	389
OST7	163	167	183	209	194	194	383	395	232	235	386	389
OST8	147	167	185	185	186	194	386	386	226	226	386	389
OST9	131	131	185	185	194	198	386	386	226	226	389	389
OST10	159	163	197	197	188	198	386	389	226	235	386	386
OST11	127	139	183	183	192	192	383	386	226	226	389	389
OST12	127	139	183	205	130	130	275	275	226	226	386	389
OST13	127	127	183	183	186	186	275	386	226	226	386	389
OST14	131	163	183	183	194	194	386	395	226	226	389	389
OST15	147	151	183	183	192	192	383	386	226	226	389	389
OST16	147	151	197	205	186	194	275	386	226	232	389	389
VAX1	127	131	177	177	180	180	395	395	238	238	386	389
VAX2	151	159	183	183	180	180	395	395	238	238	386	389
VAX3	151	159	183	183	180	180	392	392	160	232	386	389
VAX4	143	151	177	183	180	192	275	275	235	244	386	389
VAX5	151	159	183	185	180	192	275	395	226	244	386	389
VAX6	135	151	177	177	192	204	275	395	244	244	386	389
UTO1	131	151	185	185	180	192	392	395	232	238	386	389
UTO2	127	151	183	183	192	204	395	395	232	232	389	389
UTO3	131	151	183	185	194	204	275	395	229	229	386	389
UTO4	143	151	205	205	184	192	395	395	226	238	389	389
UTO5	151	151	205	205	192	192	275	395	229	229	386	389
VOR1	139	139	183	197	194	194	275	395	229	238	389	389
VOR2	151	151	183	205	192	194	275	392	226	232	389	389
VOR3	139	139	183	183	194	200	275	383	229	235	380	383
VOR4	131	147	183	205	192	192	275	275	226	229	386	389
VOR5	139	139	183	197	184	186	275	395	160	226	386	389
VOR6	131	139	177	205	194	194	275	392	226	226	389	389
VOR7	131	151	183	205	194	194	275	386	205	241	389	389
VOR8	131	131	183	205	194	194	275	275	226	226	389	389
DRA1	131	139	185	207	192	192	275	275	226	235	389	389
DRA2	127	139	177	185	194	194	275	275	205	235	386	386
DRA3	135	139	185	197	182	196	275	275	226	244	386	407
DRA4	147	147	183	185	180	194	275	275	229	235	389	389
DRA5	143	143	183	205	194	194	275	275	226	226	389	389
KIR1	139	139	197	205	194	196	275	404	229	229	386	386
KIR2	139	147	183	197	192	194	275	275	226	229	386	386
KIR3	127	155	183	205	186	192	275	275	229	235	389	389
KIR4	139	139	205	205	194	194	275	275	229	241	386	386
KIR5	151	151	197	197	186	186	275	392	223	241	386	386

Appendix B - Observed and Expected Heterozygosities

Table 24. Observed and Expected Heterozygosity per Population and Microsatellite Locus

Expected (H_E) and observed (H_O) heterozygosity for 21 *Macroplea mutica* populations (Pop.) and six microsatellite loci. Sample abbreviations are listed in Table 2.

	loc16359		loc4107		loc12208		loc321		loc3012		loc1624	
Pop.	H_E	H_O	H_E	H_O	H_E	H_O	H_E	H_O	H_E	H_O	H_E	H_O
WAL	0.07	0.07	0.29	0.20	0.54	0.53	0.43	0.27	0.51	0.40	0.33	0.40
YOR	0.70	0.73	0.46	0.13	0.51	0.53	0.86	1.00	0.69	0.67	0.56	0.53
CAM	0.40	0.47	0.37	0.20	0.67	0.60	0.88	0.73	0.55	0.53	0.51	0.53
ESU	0.45	0.27	0.25	0.27	0.73	0.87	0.79	0.53	0.66	0.67	0.50	0.53
HOL	0.72	0.63	0.43	0.25	0.64	0.75	0.80	0.50	0.43	0.13	0.67	0.50
NFL	0.80	0.54	0.28	0.23	0.77	0.62	0.58	0.38	0.82	0.77	0.54	0.54
NFS	0.65	0.60	0.13	0.13	0.66	0.73	0.82	0.67	0.84	0.40	0.48	0.47
PLW	0.75	0.67	0.19	0.20	0.66	0.53	0.81	0.80	0.78	0.67	0.40	0.40
JUN	0.60	0.53	0.71	0.60	0.75	0.53	0.72	0.60	0.78	0.73	0.48	0.40
JUW	0.56	0.47	0.25	0.27	0.80	0.60	0.79	0.80	0.85	0.80	0.50	0.47
HEL	0.60	0.64	0.72	0.64	0.61	0.45	0.71	0.82	0.81	0.64	0.52	0.73
LEM	0.77	0.60	0.49	0.47	0.63	0.60	0.81	0.87	0.85	0.67	0.43	0.33
ORT	0.83	0.73	0.66	0.67	0.67	0.60	0.89	0.80	0.89	0.53	0.40	0.33
OBJ	0.70	0.40	0.77	0.47	0.73	0.87	0.73	0.53	0.87	0.80	0.42	0.40
RUG	0.71	0.60	0.70	0.60	0.77	0.67	0.80	0.40	0.89	0.73	0.25	0.27
OST	0.89	0.67	0.77	0.47	0.72	0.33	0.66	0.60	0.50	0.47	0.48	0.33
VAX	0.80	1.00	0.62	0.33	0.53	0.50	0.67	0.33	0.82	0.50	0.55	1.00
UTO	0.64	0.80	0.73	0.20	0.76	0.80	0.51	0.60	0.78	0.40	0.47	0.60
VOR	0.72	0.38	0.68	0.88	0.60	0.38	0.61	0.75	0.80	0.75	0.44	0.38
DRA	0.89	0.60	0.84	1.00	0.76	0.40	0.00	0.00	0.80	0.80	0.60	0.20
KIR	0.76	0.40	0.71	0.60	0.78	0.60	0.38	0.40	0.76	0.80	0.36	0.00

Appendix C - Pairwise Genetic, Geographical and Ecological Distances

Table 25. Pairwise Genetic, Geographical and Ecological Distances

Genetic distances (F_{ST}, R_{ST}, residual R_{ST} calculated for the whole data set (Res.R_{ST}), and Nei's DA), geographic distances in km (Dist.), swan traffic (Traffic) and average swan abundances from April to September (Sum.), October to March (Win.) and corrected annual average (Ann.) for 210 pairs of *M. mutica* populations. Sample abbreviations are listed in Table 2.

PopA	PopB	Traffic	Sum.	Win.	Ann.	Dist.	R_{ST}	F_{ST}	Res.R_{ST}	Nei's D_A
WAL	YOR	0	1.7	1.2	2.8	478.4	0.050	0.217	-0.075	0.290
WAL	CAM	1	2.2	4.8	4.1	419.8	0.245	0.303	0.131	0.325
WAL	ESU	0	3.7	5.0	8.3	510.7	0.071	0.248	-0.061	0.190
WAL	HOL	0	7.7	2.0	13.1	995.6	-0.034	0.343	-0.262	0.396
WAL	NFL	0	1.1	2.2	1.9	1436.8	0.145	0.317	-0.172	0.282
WAL	NFS	0	1.2	5.8	2.0	1423.8	0.027	0.266	-0.287	0.268
WAL	PLW	0	2.2	5.0	8.5	1606.4	0.078	0.293	-0.273	0.304
WAL	JUN	0	0.8	5.2	1.2	1440.2	0.385	0.367	0.067	0.483
WAL	JUW	0	0.9	8.8	1.2	1413.1	0.475	0.366	0.162	0.358
WAL	HEL	0	1.3	6.0	3.2	1586.2	0.337	0.294	-0.010	0.356
WAL	LEM	0	8.9	1.1	16.3	1685.6	0.376	0.350	0.009	0.449
WAL	ORT	0	8.9	1.3	16.4	1683.0	0.216	0.283	-0.150	0.389
WAL	OBJ	0	1.4	4.9	8.3	1779.5	0.442	0.347	0.056	0.488
WAL	RUG	0	10.6	2.1	50.1	1927.1	0.132	0.406	-0.283	0.587
WAL	OST	0	0.8	5.1	1.2	3040.9	0.399	0.370	-0.239	0.547
WAL	VAX	0	6.2	1.1	22.9	2610.0	0.462	0.405	-0.089	0.579
WAL	UTO	0	6.2	1.3	20.7	2585.0	0.450	0.377	-0.097	0.486
WAL	VOR	0	0.9	4.9	1.4	3120.9	0.618	0.470	-0.036	0.701
WAL	DRA	0	1.2	2.1	1.7	3072.2	0.887	0.460	0.243	0.673
WAL	KIR	0	0.9	5.1	1.2	3271.2	0.767	0.444	0.083	0.629
YOR	CAM	5	2.1	1.2	4.5	177.8	0.015	0.177	-0.050	0.233
YOR	ESU	0	3.7	7.1	8.8	346.9	0.006	0.221	-0.093	0.292
YOR	HOL	0	7.6	7.3	13.5	617.9	0.027	0.164	-0.127	0.288
YOR	NFL	0	1.1	11.0	2.4	996.2	-0.009	0.212	-0.238	0.322
YOR	NFS	0	1.1	8.2	2.4	989.5	-0.007	0.203	-0.234	0.323
YOR	PLW	0	2.1	11.2	9.0	1179.4	-0.002	0.198	-0.268	0.305
YOR	JUN	0	0.8	7.3	1.7	973.4	0.031	0.235	-0.193	0.496
YOR	JUW	0	0.8	7.3	1.7	954.3	0.064	0.243	-0.156	0.386
YOR	HEL	1	1.3	0.5	3.6	1142.7	-0.025	0.130	-0.283	0.326
YOR	LEM	0	8.8	0.7	16.7	1255.3	0.041	0.191	-0.239	0.397
YOR	ORT	0	8.9	4.4	16.8	1252.6	0.030	0.130	-0.250	0.322
YOR	OBJ	0	1.3	1.5	8.7	1340.5	0.134	0.202	-0.164	0.420
YOR	RUG	0	10.5	4.6	50.5	1498.0	0.000	0.238	-0.329	0.465
YOR	OST	0	0.8	0.6	1.6	2566.1	0.113	0.226	-0.430	0.477
YOR	VAX	0	6.2	0.7	23.3	2147.8	0.122	0.231	-0.337	0.483
YOR	UTO	0	6.2	6.7	21.2	2125.4	0.010	0.220	-0.445	0.460
YOR	VOR	0	0.9	0.5	1.8	2669.1	0.309	0.294	-0.255	0.550
YOR	KIR	0	0.9	4.3	1.7	2815.0	0.408	0.256	-0.185	0.531

PopA	PopB	Traffic	Sum.	Win.	Ann.	Dist.	R_{ST}	F_{ST}	Res.R_{ST}	Nei's D_A
CAM	ESU	2	4.2	1.5	10.1	179.9	0.023	0.107	-0.042	0.217
CAM	HOL	0	8.1	4.5	14.8	580.9	0.256	0.111	0.110	0.206
CAM	NFL	0	1.6	0.6	3.7	1017.7	-0.018	0.104	-0.251	0.218
CAM	NFS	0	1.6	0.6	3.7	1004.1	0.030	0.094	-0.201	0.208
CAM	PLW	0	2.6	6.7	10.3	1186.7	0.048	0.136	-0.219	0.247
CAM	JUN	0	1.3	0.0	3.0	1031.8	0.115	0.165	-0.121	0.346
CAM	JUW	0	1.3	2.1	3.0	999.3	0.165	0.140	-0.064	0.260
CAM	HEL	0	1.8	2.3	4.9	1167.2	0.092	0.145	-0.171	0.320
CAM	LEM	0	9.3	5.9	18.0	1265.9	0.090	0.106	-0.193	0.288
CAM	ORT	0	9.4	3.1	18.1	1263.3	0.081	0.127	-0.201	0.299
CAM	OBJ	0	1.8	6.1	10.0	1359.9	0.268	0.161	-0.033	0.350
CAM	RUG	0	11.0	2.2	51.8	1507.5	0.075	0.194	-0.256	0.422
CAM	OST	0	1.3	2.2	2.9	2638.0	0.226	0.123	-0.331	0.306
CAM	VAX	0	6.7	8.3	24.7	2195.1	0.310	0.200	-0.159	0.434
CAM	UTO	0	6.7	1.7	22.5	2169.0	0.224	0.190	-0.239	0.411
CAM	VOR	0	1.4	1.6	3.1	2702.2	0.591	0.189	0.021	0.412
CAM	DRA	0	1.7	6.0	3.4	2655.4	0.898	0.246	0.337	0.498
CAM	KIR	0	1.4	6.2	3.0	2853.6	0.761	0.279	0.160	0.514
ESU	HOL	1	9.7	9.8	19.1	509.3	0.114	0.137	-0.017	0.277
ESU	NFL	0	3.1	7.0	8.0	982.4	-0.014	0.049	-0.240	0.147
ESU	NFS	0	3.2	10.0	8.0	962.3	-0.029	0.046	-0.251	0.164
ESU	PLW	0	4.2	6.1	14.5	1132.8	-0.015	0.078	-0.271	0.169
ESU	JUN	0	2.8	6.1	7.2	1033.7	0.176	0.186	-0.061	0.386
ESU	JUW	0	2.9	12.2	7.2	988.3	0.255	0.101	0.028	0.213
ESU	HEL	0	3.3	5.5	9.2	1130.8	0.143	0.153	-0.113	0.279
ESU	LEM	1	10.9	5.5	22.3	1213.5	0.187	0.107	-0.085	0.288
ESU	ORT	1	10.9	7.1	22.4	1211.0	0.122	0.138	-0.150	0.280
ESU	OBJ	0	3.4	6.0	14.3	1314.5	0.330	0.166	0.037	0.377
ESU	RUG	0	12.6	6.2	56.1	1450.1	0.072	0.184	-0.247	0.420
ESU	OST	0	2.8	9.9	7.2	2635.7	0.288	0.212	-0.269	0.424
ESU	VAX	0	8.2	7.0	28.9	2169.2	0.358	0.209	-0.105	0.505
ESU	UTO	0	8.2	10.1	26.7	2139.3	0.297	0.161	-0.161	0.341
ESU	VOR	0	2.9	6.1	7.4	2659.2	0.586	0.219	0.024	0.443
ESU	DRA	0	3.2	6.2	7.7	2621.2	0.872	0.285	0.318	0.568
ESU	KIR	0	2.9	12.2	7.3	2815.3	0.742	0.272	0.149	0.490
HOL	NFL	0	7.1	5.6	12.7	483.7	0.156	0.139	0.030	0.295
HOL	NFS	0	7.1	5.5	12.7	459.2	0.031	0.099	-0.091	0.263
HOL	PLW	1	8.1	7.2	19.3	623.7	0.100	0.123	-0.055	0.287
HOL	JUN	0	6.8	11.0	12.0	582.0	0.337	0.186	0.191	0.409
HOL	JUW	0	6.8	6.7	12.0	518.6	0.414	0.120	0.281	0.280
HOL	HEL	1	7.3	6.9	13.9	628.7	0.276	0.177	0.120	0.415
HOL	LEM	0	14.8	10.5	27.0	704.5	0.325	0.164	0.155	0.416
HOL	ORT	0	14.9	7.7	27.1	702.1	0.145	0.180	-0.025	0.403
HOL	OBJ	4	7.3	10.7	19.0	806.8	0.343	0.175	0.152	0.446
HOL	RUG	1	16.5	6.8	60.8	940.9	0.049	0.186	-0.169	0.447

PopA	PopB	Traffic	Sum.	Win.	Ann.	Dist.	R_{ST}	F_{ST}	Res.R_{ST}	Nei's D_A
HOL	OST	0	6.8	6.8	11.9	2149.7	0.309	0.187	-0.151	0.402
HOL	VAX	2	12.2	12.8	33.7	1667.1	0.302	0.200	-0.062	0.489
HOL	UTO	1	12.2	6.2	31.5	1635.4	0.289	0.203	-0.068	0.412
HOL	VOR	0	6.9	6.2	12.1	2150.5	0.453	0.200	-0.007	0.407
HOL	DRA	0	7.2	7.8	12.4	2114.7	0.764	0.228	0.311	0.467
HOL	KIR	0	6.9	11.7	12.0	2307.4	0.597	0.221	0.106	0.417
NFL	NFS	4	0.6	11.7	1.6	40.1	-0.017	0.053	-0.054	0.160
NFL	PLW	2	1.6	9.0	8.1	197.2	-0.020	0.062	-0.089	0.139
NFL	JUN	0	0.2	9.2	0.8	214.1	0.060	0.069	-0.012	0.219
NFL	JUW	0	0.3	12.8	0.8	136.3	0.127	0.054	0.070	0.129
NFL	HEL	6	0.7	10.0	2.8	149.7	0.042	0.091	-0.018	0.210
NFL	LEM	1	8.3	13.0	15.9	263.9	0.095	0.046	0.013	0.202
NFL	ORT	1	8.3	9.1	16.0	261.1	0.087	0.087	0.005	0.199
NFL	OBJ	2	0.8	9.1	7.9	344.5	0.266	0.109	0.167	0.272
NFL	RUG	0	10.0	15.2	49.7	504.1	0.047	0.088	-0.083	0.273
NFL	OST	0	0.2	8.6	0.8	1666.6	0.234	0.165	-0.129	0.378
NFL	VAX	1	5.6	8.5	22.5	1186.9	0.291	0.114	0.023	0.376
NFL	UTO	1	5.6	10.1	20.3	1157.2	0.201	0.100	-0.060	0.304
NFL	VOR	0	0.3	14.0	1.0	1684.7	0.552	0.175	0.186	0.428
NFL	DRA	0	0.6	14.1	1.3	1640.7	0.851	0.230	0.493	0.482
NFL	KIR	0	0.3	14.7	0.8	1837.1	0.708	0.255	0.311	0.505
NFS	PLW	4	1.6	0.5	8.2	192.2	-0.035	0.021	-0.103	0.119
NFS	JUN	0	0.3	0.7	0.8	251.6	0.124	0.163	0.044	0.349
NFS	JUW	0	0.4	4.3	0.8	171.5	0.187	0.071	0.122	0.199
NFS	HEL	6	0.8	1.5	2.8	169.6	0.091	0.135	0.027	0.295
NFS	LEM	2	8.4	4.5	15.9	265.9	0.147	0.108	0.064	0.319
NFS	ORT	2	8.4	0.6	16.0	263.2	0.084	0.158	0.002	0.326
NFS	OBJ	2	0.9	0.6	7.9	355.8	0.274	0.149	0.173	0.365
NFS	RUG	3	10.0	6.7	49.7	508.5	0.017	0.177	-0.114	0.432
NFS	OST	0	0.3	0.0	0.8	1694.1	0.244	0.170	-0.125	0.397
NFS	VAX	1	5.7	0.0	22.5	1208.3	0.287	0.190	0.016	0.511
NFS	UTO	1	5.7	1.6	20.4	1177.3	0.207	0.202	-0.058	0.463
NFS	VOR	0	0.4	5.5	1.0	1700.0	0.513	0.223	0.143	0.483
NFS	DRA	0	0.7	5.5	1.3	1659.0	0.809	0.268	0.448	0.569
NFS	KIR	0	0.4	6.2	0.9	1854.1	0.668	0.246	0.267	0.501
PLW	JUN	0	1.3	8.5	7.4	366.5	0.120	0.172	0.017	0.369
PLW	JUW	0	1.3	14.1	7.4	309.4	0.202	0.097	0.111	0.214
PLW	HEL	14	1.8	14.3	9.4	116.8	0.094	0.121	0.041	0.279
PLW	LEM	171	9.3	18.0	22.5	80.9	0.171	0.111	0.126	0.303
PLW	ORT	171	9.4	15.2	22.6	78.5	0.129	0.157	0.083	0.295
PLW	OBJ	16	1.8	18.2	14.5	188.4	0.323	0.169	0.256	0.384
PLW	RUG	15	11.0	14.3	56.3	320.8	0.058	0.168	-0.036	0.378
PLW	OST	0	1.3	14.3	7.4	1564.1	0.296	0.205	-0.047	0.451
PLW	VAX	3	6.7	20.3	29.1	1055.0	0.357	0.193	0.116	0.495
PLW	UTO	3	6.7	13.7	26.9	1019.8	0.279	0.210	0.045	0.428

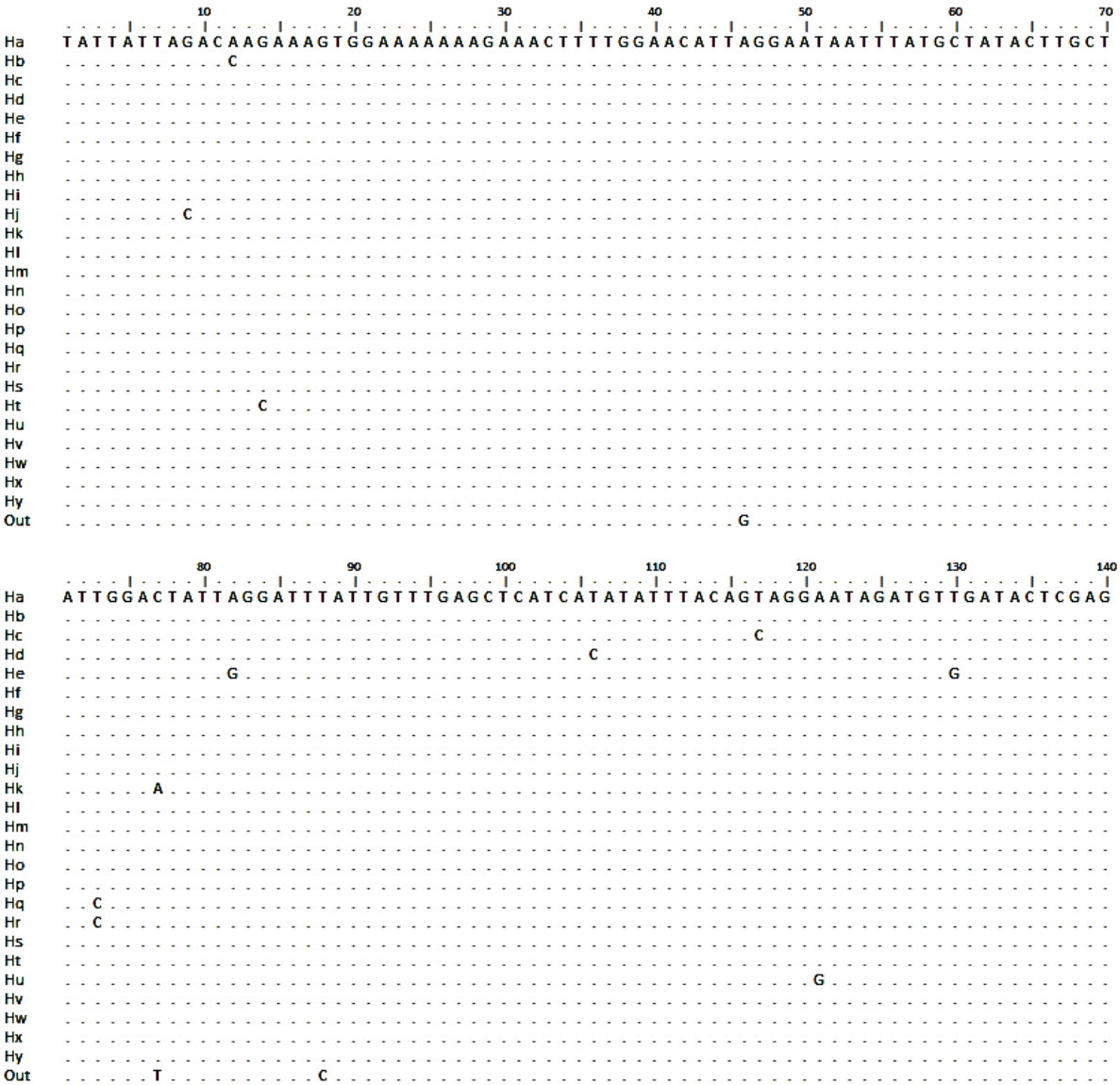
PopA	PopB	Traffic	Sum.	Win.	Ann.	Dist.	R_{ST}	F_{ST}	Res.R_{ST}	Nei's D_A
PLW	VOR	0	1.4	13.7	7.6	1526.9	0.579	0.229	0.244	0.488
PLW	DRA	0	1.7	15.3	7.9	1493.7	0.856	0.263	0.527	0.541
PLW	KIR	1	1.4	19.2	7.4	1684.6	0.728	0.227	0.361	0.446
JUN	JUW	1	0.0	19.2	0.1	82.7	0.001	0.063	-0.045	0.176
JUN	HEL	1	0.5	19.8	2.0	260.8	-0.011	0.105	-0.093	0.282
JUN	LEM	0	8.0	22.2	15.2	402.9	0.024	0.095	-0.087	0.275
JUN	ORT	0	8.0	13.7	15.2	400.4	0.088	0.079	-0.022	0.263
JUN	OBJ	0	0.5	13.6	7.2	436.0	0.210	0.056	0.093	0.230
JUN	RUG	1	9.7	13.8	49.0	608.2	0.064	0.043	-0.087	0.245
JUN	OST	0	0.0	17.4	0.0	1606.9	0.213	0.105	-0.138	0.321
JUN	VAX	0	5.4	14.6	21.8	1174.9	0.286	0.102	0.021	0.355
JUN	UTO	0	5.4	17.7	19.6	1153.7	0.159	0.080	-0.101	0.298
JUN	VOR	0	0.0	13.7	0.3	1702.6	0.576	0.152	0.206	0.403
JUN	DRA	0	0.4	13.8	0.6	1641.9	0.878	0.176	0.519	0.461
JUN	KIR	0	0.1	19.8	0.1	1844.6	0.739	0.240	0.341	0.506
JUW	HEL	0	0.5	13.2	2.0	217.9	0.025	0.113	-0.048	0.227
JUW	LEM	0	8.1	13.1	15.2	358.1	-0.006	0.062	-0.107	0.194
JUW	ORT	0	8.1	14.8	15.2	355.4	0.088	0.115	-0.012	0.251
JUW	OBJ	0	0.6	18.6	7.2	410.3	0.174	0.078	0.063	0.231
JUW	RUG	0	9.7	18.7	49.0	581.2	0.094	0.064	-0.052	0.244
JUW	OST	0	0.0	19.3	0.0	1647.5	0.207	0.139	-0.152	0.359
JUW	VAX	0	5.4	21.7	21.8	1196.9	0.276	0.165	0.007	0.382
JUW	UTO	0	5.4	13.1	19.6	1172.5	0.130	0.158	-0.134	0.318
JUW	VOR	0	0.1	26.8	0.3	1714.8	0.584	0.166	0.211	0.397
JUW	DRA	0	0.4	0.5	0.6	1660.3	0.896	0.225	0.535	0.495
JUW	KIR	0	0.1	0.7	0.1	1861.0	0.756	0.232	0.354	0.432
HEL	LEM	15	8.5	4.4	17.1	142.3	0.014	0.027	-0.044	0.132
HEL	ORT	15	8.5	1.5	17.2	139.8	0.060	0.044	0.002	0.161
HEL	OBJ	2	1.0	4.6	9.1	199.0	0.178	0.058	0.108	0.196
HEL	RUG	1	10.2	0.6	50.9	365.9	0.038	0.091	-0.065	0.225
HEL	OST	0	0.5	0.7	2.0	1526.4	0.184	0.159	-0.151	0.399
HEL	VAX	0	5.9	6.7	23.7	1039.0	0.255	0.131	0.018	0.383
HEL	UTO	0	5.9	0.1	21.5	1008.5	0.109	0.083	-0.123	0.281
HEL	VOR	0	0.5	0.0	2.2	1535.0	0.531	0.196	0.194	0.448
HEL	DRA	0	0.9	1.7	2.5	1491.4	0.859	0.217	0.531	0.523
HEL	KIR	0	0.6	5.5	2.1	1687.5	0.697	0.208	0.329	0.511
LEM	ORT	910	16.1	5.6	30.3	2.8	0.015	0.034	-0.015	0.156
LEM	OBJ	31	8.6	6.2	22.2	111.0	0.074	0.049	0.022	0.140
LEM	RUG	29	17.7	8.6	64.0	242.8	0.048	0.056	-0.030	0.203
LEM	OST	0	8.0	0.0	15.1	1489.0	0.082	0.135	-0.246	0.359
LEM	VAX	6	13.4	13.7	36.9	975.8	0.127	0.127	-0.097	0.378
LEM	UTO	6	13.4	13.2	34.7	940.1	-0.014	0.094	-0.232	0.276
LEM	VOR	0	8.1	0.5	15.3	1446.1	0.455	0.140	0.136	0.355
LEM	DRA	0	8.4	0.7	15.6	1413.0	0.797	0.193	0.484	0.439
LEM	KIR	1	8.1	4.3	15.2	1603.7	0.626	0.222	0.275	0.463

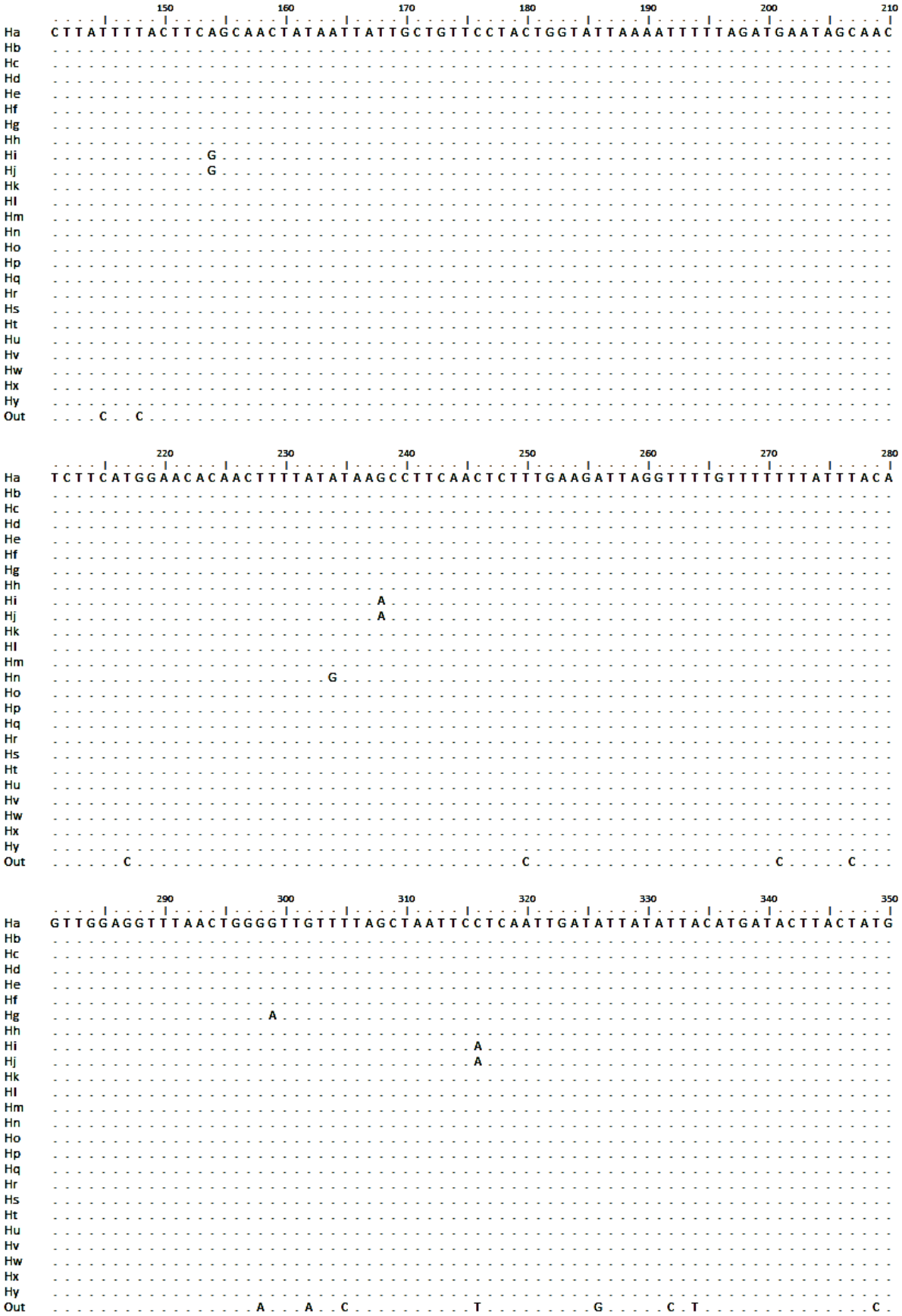
PopA	PopB	Traffic	Sum.	Win.	Ann.	Dist.	R_{ST}	F_{ST}	Res.R_{ST}	Nei's D_A
ORT	OBJ	24	8.6	1.5	22.3	112.6	0.000	0.039	-0.052	0.181
ORT	RUG	26	17.8	4.5	64.1	245.5	-0.030	0.055	-0.109	0.227
ORT	OST	0	8.0	0.6	15.2	1490.6	-0.019	0.110	-0.347	0.359
ORT	VAX	6	13.4	0.6	36.9	977.8	-0.016	0.135	-0.242	0.402
ORT	UTO	5	13.4	6.7	34.8	942.1	-0.068	0.080	-0.286	0.306
ORT	VOR	0	8.1	0.0	15.4	1448.5	0.211	0.131	-0.108	0.400
ORT	DRA	0	8.4	0.0	15.7	1415.2	0.566	0.137	0.254	0.416
ORT	KIR	1	8.1	1.6	15.3	1606.1	0.364	0.214	0.013	0.495
OBJ	RUG	30	10.2	5.5	56.0	172.3	0.077	0.028	0.013	0.179
OBJ	OST	0	0.5	5.5	7.1	1378.0	-0.025	0.085	-0.330	0.306
OBJ	VAX	105	5.9	6.2	28.8	866.7	-0.057	0.126	-0.260	0.365
OBJ	UTO	98	5.9	8.5	26.7	831.9	-0.095	0.075	-0.292	0.258
OBJ	VOR	1	0.6	0.0	7.3	1345.2	0.180	0.116	-0.119	0.353
OBJ	DRA	3	0.9	13.7	7.6	1308.0	0.574	0.153	0.283	0.409
OBJ	KIR	0	0.6	13.1	7.2	1500.8	0.336	0.206	0.006	0.489
RUG	OST	0	9.7	0.0	48.9	1311.5	0.061	0.120	-0.231	0.358
RUG	VAX	23	15.1	0.5	70.7	774.4	0.064	0.122	-0.120	0.365
RUG	UTO	19	15.1	0.7	68.5	733.0	-0.007	0.093	-0.184	0.275
RUG	VOR	0	9.8	4.3	49.2	1214.3	0.278	0.102	0.005	0.352
RUG	DRA	0	10.1	1.5	49.5	1190.1	0.617	0.130	0.349	0.378
RUG	KIR	0	9.8	4.5	49.0	1375.4	0.434	0.212	0.129	0.462
OST	VAX	0	5.4	0.6	21.7	554.6	-0.065	0.173	-0.206	0.437
OST	UTO	0	5.4	0.6	19.6	605.2	-0.068	0.146	-0.218	0.348
OST	VOR	0	0.0	6.7	0.2	609.0	0.196	0.075	0.045	0.238
OST	DRA	0	0.4	0.0	0.5	469.9	0.600	0.088	0.477	0.311
OST	KIR	0	0.1	0.0	0.1	585.2	0.362	0.164	0.216	0.323
VAX	UTO	321	10.8	1.6	41.3	53.2	-0.045	0.074	-0.085	0.239
VAX	VOR	0	5.4	5.5	22.0	550.9	0.122	0.175	-0.018	0.403
VAX	DRA	0	5.7	5.5	22.3	469.6	0.540	0.165	0.416	0.396
VAX	KIR	1	5.5	6.2	21.8	675.2	0.261	0.207	0.096	0.551
UTO	VOR	0	5.4	8.5	19.8	557.5	0.268	0.123	0.127	0.271
UTO	DRA	0	5.8	0.0	20.1	488.2	0.741	0.168	0.614	0.398
UTO	KIR	0	5.5	13.7	19.6	691.3	0.448	0.179	0.280	0.414
VOR	DRA	2	0.4	13.1	0.8	141.0	0.258	0.019	0.200	0.248
VOR	KIR	1	0.1	0.0	0.3	171.0	0.017	0.093	-0.047	0.218
DRA	KIR	0	0.4	0.0	0.6	206.4	0.120	0.059	0.049	0.283

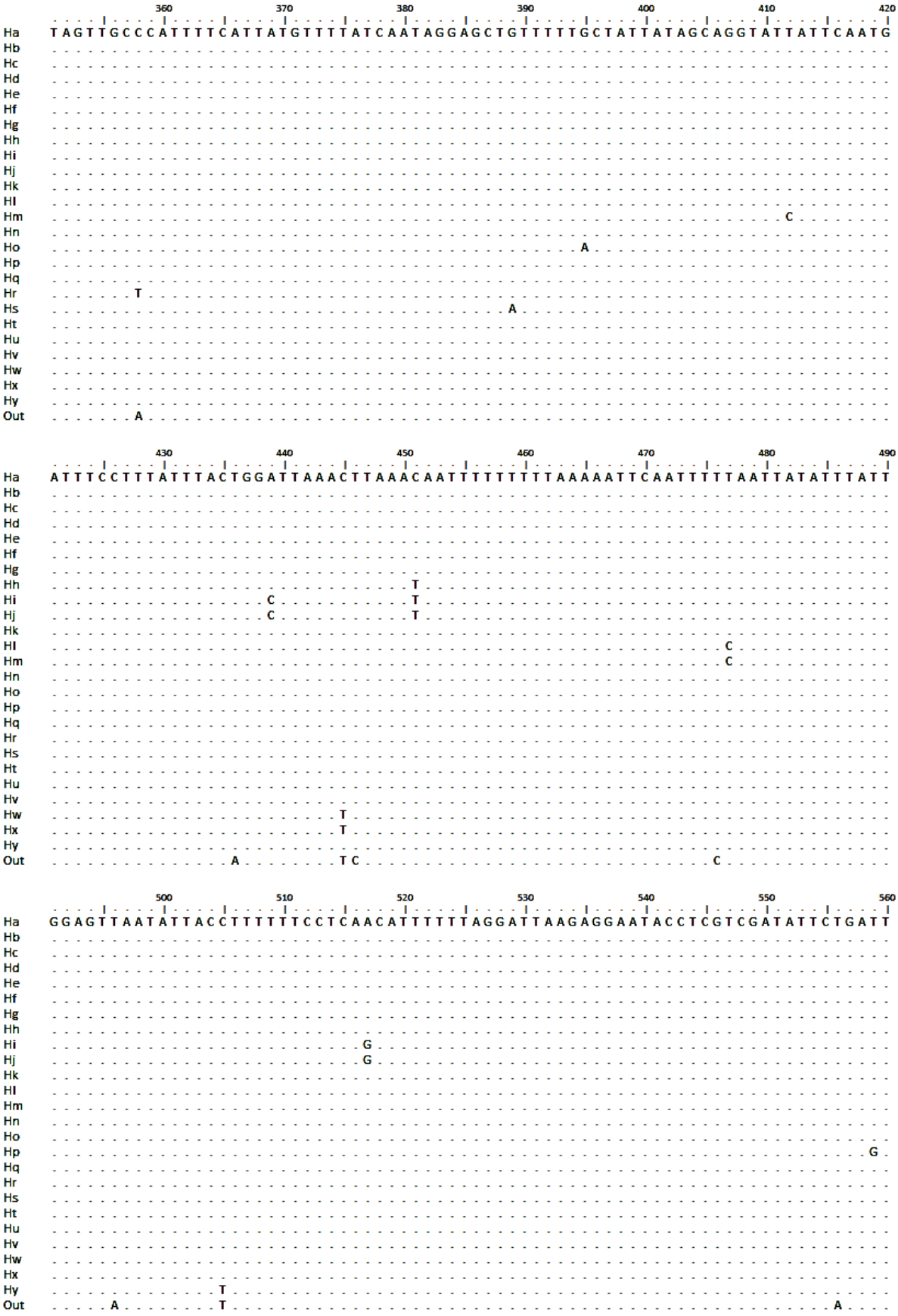
Appendix D - Partial COI Sequences

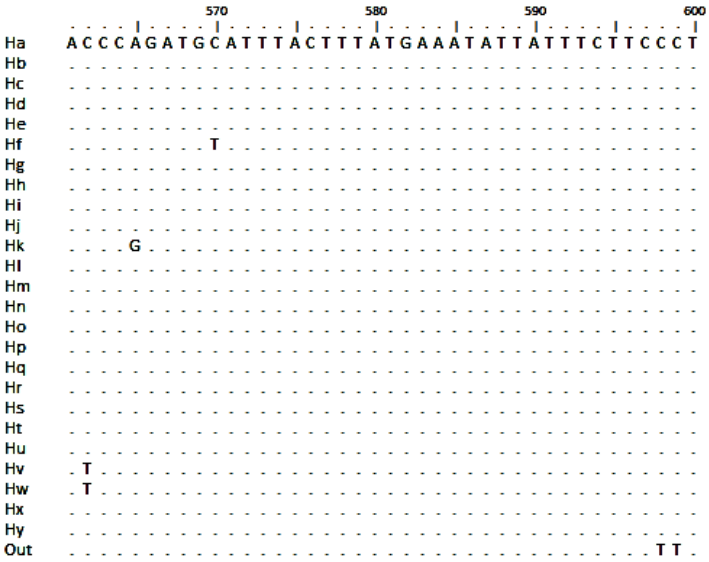
Table 26. COI Sequence Alignment

a) Sequence alignment for 26 detected haplotypes based on a 600 bp long section of the mitochondrial *cytochrome oxidase I* gene (COI). The full sequence is shown for haplotype Ha. For haplotypes Hb to Hy and the outgroup haplotype (Out), nucleotide substitutions with respect to haplotype Ha are shown. Numbers signify the base pair position within the analysed COI fragment.




150 160 170 180 190 200 210
Ha CTTATTTTACTTCAGCAACTATAATTATTGCTGTTCCTACTGGTATTAAAATTTTTAGATGAATAGCAAC
Hi G
Hj G
Out C C
220 230 240 250 260 270 280
Ha TCTTCATGGAACACAACTTTTATATAAGCCTTCAACTCTTTGAAGATTAGGTTTTGTTTTTTTATTTACA
Hi A
Hj A
Hn G
Out C C C C
290 300 310 320 330 340 350
Ha GTTGGAGGTTTAACTGGGGTTGTTTTAGCTAATTCCTCAATTGATATTATATTACATGATACTTACTATG
Hg A
Hi A
Hj A
Out A A C T G C T


360 370 380 390 400 410 420
Ha TAGTTGCCCATTTTCATTATGTTTTATCAATAGGAGCTGTTTTTGCTATTATAGCAGGTATTATTCAATG
Hb
Hc
Hd
He
Hf
Hg
Hh
Hi
Hj
Hk
Hl
Hm C
Hn
Ho A
Hp
Hq
Hr T
Hs A
Ht
Hu
Hv
Hw
Hx
Hy
Out A
430 440 450 460 470 480 490
Ha ATTTCCTTTATTTACTGGATTAAACTTAAACAATTTTTTTTTAAAAATTCAATTTTTAATTATATTTATT
Hb
Hc
Hd
He
Hf
Hg
Hh T
Hi C T
Hj C T
Hk
Hl C
Hm C
Hn
Ho
Hp
Hq
Hr
Hs
Ht
Hu
Hv
Hw T
Hx T
Hy
Out A T C C
500 510 520 530 540 550 560
Ha GGAGTTAATATTACCTTTTTTCCTCAACATTTTTTAGGATTAAGAGGAATACCTCGTCGATATTCTGATT
Hb
Hc
Hd
He
Hf
Hg
Hh
Hi G
Hj G
Hk
Hl
Hm
Hn
Ho
Hp G
Hq
Hr
Hs
Ht
Hu
Hv
Hw
Hx
Hy T
Out A T A


570
580
590
600
Ha ACCCAGATGCATTTACTTTATGAAATATTATTTCTTCCCT
Hb
Hc
Hd
He
Hf T
Hg
Hh
Hi
Hj
Hk G
Hl
Hm
Hn
Ho
Hp
Hq
Hr
Hs
Ht
Hu
Hv T
Hw T
Hx
Hy
Out TT

b) Individual samples within the haplotypes Ha to Hy (for 186 individuals of *Macroplea mutica* from 21 sampling sites) and the outgroup haplotype (5 individuals of *Macroplea appendiculata* from two sampling sites). Sample abbreviations are listed in Table 2.

COI haplotype	Sample(s) within haplotype
Ha	BRA01, BRA02, BRA03, BRA04, BRA05, BRA06, BRA07, BRA08, BRA09, BRA10, WAL05, WAL06, WAL07, WAL08, WAL10, WAL16, WAL18, WAL19, WAL20, YOR1, YOR, ESU02, ESU03, ESU04, ESU05, ESU06, ESU07, ESU08, ESU09, ESU10, ESU15, HOL2, HOL22, HOL4, HOL21, HOL31, HOL32, NFS2, NFS5, NFS7, NFS8, JUN6, JUN7, JUN8, JUN9, JUN11, JUN12, JUN13, JUN14, JUN7e, JUN9e, JUN11e, JUN12e, JUW2, JUW5, JUW6, JUW9, JUW10, JUW12, JUW13, JUW14, JUW15, JUW16, JUW17, PLW1, PLW2, PLW3, PLW4, PLW5, PLW6, PLW7, PLW9, PLW10, ORT28, ORT29, ORT34, ORT36, ORT46, SEE14, SEE16, SEE17, SEE19, SEE21, SEE22, SEE24, RUG6, RUG11, RUG4, HEL7, HEL1, HEL11, HEL3, HEL5, HEL4, HEL8, HEL10, HEL6, HEL9, UTO45, UTO46, UTO47, DRA50, DRA51, DRA53, DRA54, OST60, OST02, OST05, ORT61, ORT62, ORT63, ORT64, ORT65, ORT66, KIR70, KIR71, KIR72, KIR73, VAX74, VAX76, VAX78, VAX79, CHI95, CHI84, CHI86, CHI88, CHI89, CHI91, CHI93
Hb	SEE18
Hc	UTO48
Hd	ORT27, ORT31
He	BALB80, BALS81, BALB82, BALB83
Hf	OST56, OST57, OST58, OST01, OST03, OST04, OST06
Hg	OST55
Hh	ORT33, DRA52
Hi	SAR01, SAR02, SAR03, SAR04, SAR05, SAR06, SAR07, SAR09, SAR10
Hj	SAR08
Hk	VAX75
Hl	JUN10, JUN8e, JUN15, PLW8
Hm	NFS1, NFS10, NFS3, NFS4, NFS6, NFS9
Hn	UTO49
Ho	VAX77
Hp	HOL13
Hq	RUG1, RUG3, RUG9, RUG2, RUG7, RUG12
Hr	RUG14
Hs	WAL09
Ht	ORT30
Hu	ORT35
Hv	OST59, KIR69
Hw	JUN10e
Hx	HOL18
Hy	CHI85
Outgroup (*Macroplea appendiculata*)	MAS1, MAN1, MAN 4, MAN 5, MAN 6