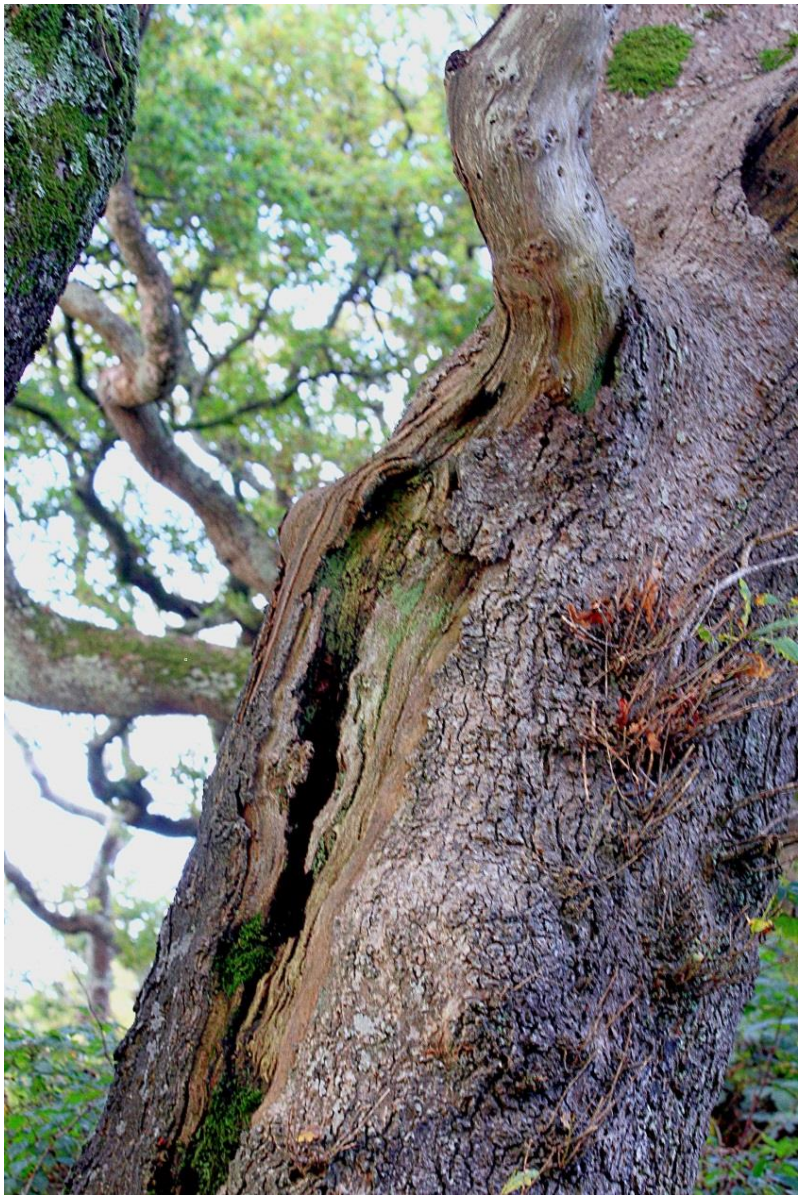
Conservation Genetics of Saproxylic Beetles

by

Max Blake

SN: 129004813

Supervisor: Prof Paul Shaw



‘They tell us that we lost our tails

Evolving up from little snails

I say it's all just wind in sails’

Devo – 1978

"Whenever I hear of the capture of rare beetles,

I feel like an old war-horse at the sound of a trumpet"

Charles Darwin

Contents

Acknowledgments

I would like to thank a number of people and organisations for their help throughout the last few years. Without them this thesis would simply not have been possible.

First and foremost, I have to thank The Leverhulme Trust for funding this work, and for deciding that both orchards and beetles are worth saving.

Huge thanks must be sent to my supervisory team, Paul Shaw, Chris Thomas, and Niall McKeown. Paul, without you taking a risk on a bright eyed, naïve, non-geneticist I wouldn't be writing this now. Thank you for having faith that one day I would eventually 'get there' and for putting up with my constant ideas for side projects. With hindsight I agree, a lot of them were total nonsense. Thanks also for 'that' car duet, the first quote is for you. Thanks Chris for your positive outlook on everything every time we met, and for helping push the SDM along. Niall, the Malcolm Tucker to Paul's PM, thank you for keeping on top of me via sound advice or a crack of the whip (usually both), and for being the fount of knowledge from which I sipped for anything genetics-related. 'Junior Supervisor' doesn't do your role in shaping both me as a scientist and a budding conservation geneticist justice. Sorry for the constant questions, and the questionable taste in music.

I'm indebted to a number of people who provided advice and help throughout the various aspects of this work. Hefin Williams, thanks for helping me with the SDM analysis and for getting me to think critically about modelling approaches. Ben Harvey, thanks for suggesting geometric morphometrics as a tool for the shape analysis, it managed to bring everything together nicely (once I'd worked out how to use it!). Mark Bushell, thanks so much for the constant reminders of life as a conservation entomologist outside of academia, and dragging me to conferences. Thanks also for encouraging the development of the spider web work, let's hope we can put it into practice. Harry Green, thanks for all the fieldwork advice for finding noble chafers, I think we managed to sift through more beetle poo in that week than any other entomologists in history. Thanks also to the staff of the Natural History Museum (Max Barclay in particular) for letting me photograph the chafers and for being forgiving about me getting in the way.

A huge thanks to everyone who has sent us samples to work on, without these the scope of the project would have been incredibly limited: Luca Bartolozzi, Adam Bates, Deborah Harvey, Nicklas Jansson, Estefanía Micó Balaguer, György Rozner, Petr Šípek, Jennifer Higgs, Dave Chesmore, Rebecca Lashley. Thanks as well to everyone else involved with the wider Leverhulme grant who's not been noted yet, Jon Sadler, Alan Gange, Lloyd Jenkins, Laura Bower, thank you all so much for being so supportive of our work. I look forward to tying everything together!

For keeping me semi-sane throughout this process, a thanks to my flatmates, Amy Healey, Lizzy Donkin, and Nathan King. Sorry for the music, 'experimental' food and grumpy mood when things weren't working. Additional thanks to everyone who has put me up for a night, talked rubbish to me on the phone, or dragged me out for a beer: Steve Briley, Louise Brown, Freddie Sutton, Ruth and Ryan Wonfor, Daniella Rabaiotti, Matt Pickering, Michèle Aubertin, Cat Baker, Jon & Corrina Downes, George Rowland, Alex Mansfield, Tom Hutton, John and Lili Dixon, Naomi Chetland, Alun Evans, Flavia Occhibove, Mathilde Bue, Norioshi Sato, Tamsin Shepherd-Waring, Hayley Watson, Mike Gwilliam, Sabrina Munshi and Rhian Rowson.

A very quick thanks to the additional things that kept me going: coffee, Scott Walker, the tegus, Frank Zappa, coffee, King Crimson, Ginger Baker, Melvins, the un-named hissing cockroaches, Pokémon, Shining, C_2H_6O, Swans, Hermione my orchid mantis, motorik, Silver Mt. Zion, photography, coffee, The Necks.

Finally and most importantly I'd like to thank my biological and extended family for their love, patience and support throughout the last few years. This could not have been written without their stoic backing and belief in me; to witter on with extended prose dilutes the impact of this fact.

Thank you from the bottom of my heart.

Declaration

I hereby declare that the work presented in this thesis was conducted by myself under the supervision of Professor Paul W. Shaw, as well as Professor Chris J. Thomas and Dr Niall J. McKeown, with the exception of those instances where the contribution of others has been specifically acknowledged. All material that has been quoted or referenced from other sources has been clearly noted, including figures.

Chapter 6 is partly based on two published peer-reviewed research papers lead by myself, both of which are printed at the end of this thesis.

The microsatellite development in chapters 2 and 6 was completed alongside Dr Niall J. McKeown.

The null model bioinformatics pipeline was developed and implemented by Dr Hefin Williams and published in Williams et al. 2015, which is referenced throughout Chapter 7.

I hereby confirm that this thesis has not been submitted for any other qualification or degree at this university or any other institution.

Max Blake

October 2016

SN 129004813

Abstract

Saproxylic beetles feed on decayed dead wood, representing some of the most ecologically important fauna in mature forests dominated by veteran trees, though this habitat is a shadow of its former self in Britain and Europe; confined to a handful of unconnected sites it remains nationally important for the diverse community assemblage it supports. Saproxylics can mediate and manipulate the fungal decay of wood, though this knowledge about the keystone status of saproxylic beetles is still poorly understood, despite some saproxylic beetles being some of the largest and most charismatic invertebrates in Europe. Habitats for saproxylics have been carved up over the course of human development, whilst forestry management practices encouraged the 'tidying up' of woodland by removing dead wood, leading to localised extinctions of saproxylics throughout Britain and Europe. Even in cases where extinction/threatened statuses are formally recognised servicing of conservation plans is constrained by limited understanding of their biodiversity.

In this research the development and application of a suite of morphological, genetic, and ecological simulation approaches to characterise various components of saproxylic beetle diversity is described. This study focusses on three saproxylics: the Stag Beetle (*Lucanus cervus*), the Noble Chafer (*Gnorimus nobilis*), and the Bee Beetles (*Trichius* spp.). For Stag Beetles, the first microsatellites were developed and utilised along with mitochondrial (mtDNA) Cytochrome Oxidase 1 sequencing to perform a preliminary genetic analysis. mtDNA supported the predominance of a single clade across the species' distribution which exhibited signatures of historical expansion/contractions linked to glacial/interglacial periods. Both mtDNA and microsatellites revealed a markedly lower level of genetic variation among UK samples compared to continental European counterparts and the implications for sustainability and remedial actions are discussed.

For *Gnorimus nobilis*, a rot-hole specialist confined in England to traditional orchards, geometric and mtDNA sequence analysis characterised and confirmed differences between it and its relative, the Variable Chafer (*G. variabilis*), as well as supporting recognition of other proposed sub-species. A salient feature of the intraspecific genetic variation was the deep divergence between the Eastern and Western clades, indicating historical vicariance and limited post-glacial overlap between clades. For *Trichius*, analysis of three distinct species revealed three reciprocally monophyletic mtDNA clades. However, despite support for three distinct morphotypes, there was a high level of mtDNA/ morphological incongruence, i.e. multiple morphotypes were associated with a single clade. Nuclear DNA sequencing supported the hypothesis that the pattern has been generated by historical vicariance and introgression upon secondary contact.

Species Distribution Modelling of multiple saproxylics (n=14) indicated that many species had similar glacial refugia: a northern refugium across France and Germany was predicted for many species. In addition, many species were predicted to find the Mediterranean region increasingly unsuitable under climate change simulations, but are also predicted to find suitable climates opening up toward the north-east of Europe. The project represents the first combined study of morphological taxonomy, phylogeography, population genetics and habitat modelling in rot-hole associated beetle species across Europe, and should help direct conservation efforts for these and other saproxylic beetles.

Chapter 1

Biodiversity conservation and saproxylic beetles

1.1 Invertebrate Conservation

The earth is in the midst of a period of increased rate of species extinctions (Lawton & May 1995; Pimm et al. 1995; IUCN 2016a). However, current knowledge about the patterns and processes of extinctions are heavily biased toward large charismatic taxa such as mammals and birds (Honeycutt et al. 2010; Cardoso et al. 2011a), with relatively little research interest on the conservation biology of invertebrates, despite estimates of millions of species being at imminent threat of extinction (McKinney 1999). Despite terrestrial invertebrates contributing to ecosystem functioning by nutrient recycling (Abbadie et al. 1992), forming soil on bedrock (Lawrence & Samways 2003), plant pollination (Kremen & Chaplin-Kremen 2007), ecosystem modification (Buse et al. 2008a & 2008b) and a myriad of other interactions (see Samways 2010), knowledge about what the future holds for invertebrates, particularly with respect to climate change, is largely lacking. Given the finite resources available to invertebrate conservation, there must be considerable focus on building knowledge bases for focal species, then extrapolating and testing the extent to which such knowledge can be applied to other related species (both taxonomically and ecologically). This allows for strong cross-species hypotheses to be built to answer such questions as 'where were populations of this species found during past climate changes', 'how well connected are populations of this species across space and time', 'to what scale and where should management strategies be employed', or 'to what extent does variation in the phenotype represent variation in the genotype'.

Though the mammal/bird taxon bias in research and funding focus is well understood, two major factors contribute to this: 1. Taxonomic inadequacies - we likely have scientific names for only 10% of invertebrates (Lawton & May 1995); 2. Public perception - people are biased against invertebrates and ignorant of the ecosystem services that they provide (Goldman et al. 2010; Cardoso et al. 2011b). These two factors work in tandem to disrupt invertebrate conservation: inaccurate taxonomy can lead to cryptic extinctions or underassessment of species' ecosystem services. If one morphological species actually represents a number of genetically and ecologically distinct species, each one avoiding competition from the other by specialising (using a different host species, different breeding season, etc.), then conservation targeted at this 'single species' is likely to oversimply the true niche of the 'species', leading to cryptic extinction.

However, these biases are being slowly overturned. Citizen science, observations and data produced by amateurs (Gardiner et al. 2012; Lindenmayer et al. 2013), is becoming increasingly important in identifying species and mapping ranges, particularly in Western Europe where people are more likely to find a terrestrial invertebrate in the wild than they are a large mammal. Records can be uploaded to databases for verification, and eventually to a freely accessible data portal (such as GBIF, www.gbif.org) for general use by scientists and the public. This is promising for invertebrate conservation, leading to large increases in the recorded extent of occurrence for some protected beetles (Zapponi et al. 2016). However, species distinctions in invertebrates often do not match with morphological identification (Sites & Marshall 2003; Bai et al. 2014; Ober & Connolly 2015) so research addressing the correct taxonomy of species, and being able to match these to identifiable criteria, is of paramount importance.

Taxonomic challenges can largely be solved by utilising science to test species boundaries. Understanding how invertebrate populations interact over time and space is a key question which relies on accurate taxonomy and an understanding of the fundamental differences between evolutionary units. Combining genetic and morphological methods to identify species is the best way to diagnose and identify species, testing hypotheses about species boundaries and identifying evolutionarily significant sub-taxa, and then conservationists can base ideas and species survival programs on strong foundations as well as being confident about utilising citizen science data if this is collected in a robust manner.

1.2 Saproxylic organisms and tree-rot

Saproxylic organisms are "species which are involved in or dependent on the process of fungal decay of wood, or on the products of that decay, and which are associated with living as well as dead trees" (Alexander 2008a.). Microorganisms, fungi and insects represent the dominant groups of saproxylics, though the insects usually feed on wood partially broken down by fungi and microorganisms. Though apparently a narrowly defined term, the word 'saproxylic' includes "wood-feeders, bark-feeders, feeders on wood-decomposing fungi, associated predators, parasitoids, detritivores feeding on their waste products, and … commensals" (Grove 2002), thus including a wide range of species that exploit these niches. Additionally, key habitat features utilised by saproxylics include "standing and fallen dead wood of various diameters and in various states and stages of decay, wood-rotting and other

dependent fungi (hyphae and sporocarps), fissures and crevices in bark, water- or humus-filled rot- holes and other tree cavities, sap-runs, … the tunnels and frass of wood-borers, charred wood, and waterlogged wood" (Grove 2002), so again encompassing a wide habitat diversity and extent. Relationships between habitat features and saproxylics themselves can be extremely complicated, due in part to the diversity of tree species and macro habitats across the world, as well as changes in saproxylic succession. One group of insects however has adapted to this diverse potential habitat with great success and diversity, the beetles (Coleoptera).

Beetles represent the dominant macroscopic saproxylic organisms throughout most of the earth's forests, including major keystone species within European deciduous temperate forests (*Cerambyx* spp., Buse et al. 2007) and Amazonian rainforests (*Acrocinus* spp., Zeh et al. 2003). Though the true number of saproxylic beetles is unknown, 56% of forest beetles in Germany are saproxylic (Köhler 2000), implying that over 500,000 of the estimated 1.5 million beetle species on earth (Stork et al. 2015) may be saproxylics. Saproxylic organisms also comprise up to one quarter of the total biodiversity within forests (Sittonen 2001; Humphrey et al. 2005), but despite this essential role in nutrient cycling there is a lack of knowledge surrounding the interactions between the fungi associated with the different habitats within a tree and other saproxylic species. However, one saproxylic habitat has been well researched, in particular with regard to saproxylic beetle – saproxylic fungi – tree host interactions: heartwood rot.

Mechanical damage to a tree can allow parasitic fungi or bacteria to enter through the protective bark and sapwood, which will begin to degrade and rot the heartwood of the tree (Wagener & Davidson 1954; Whitehead 2003). Over time, and depending on the fungal and/or bacterial species present, the temperature of the tree trunk, water availability and a host of other factors, a dry rot-hole may develop (Tyler 2008). Progression speed of the fungi through the wood is extremely variable, taking longer in deciduous than coniferous trees (Wagener & Davidson 1954). Though the process whereby a rot-hole is created is still poorly understood (Alexander 2003; Müller et al. 2014), in a single mature tree the variety of factors involved allow for saproxylic fungi and beetles to become highly specialised on certain tree species or rot types.

The two main rot types produced by fungi can be partitioned into those attacking lignin (leaving behind cellulose) and those attacking cellulose (leaving lignin). These cause a

crumbly ‘red rot’, and a pale, soft ‘white rot’ respectively (Alexander 2003), creating a void in the centre of the tree filled with rotten wood and fungal mycelia, known as pabula. Trees can survive with a rot-hole for many years; this may be a symbiotic relationship as it encourages the breakdown of the heavy, inflexible heartwood (Whitehead 2003). Large, slow growing trees like oaks and beech have the potential to hold rot-holes for hundreds of years. The fate of most large trees may be to senesce and become a rot-hole holder, tree age and diameter being strongly correlated with rot-hole presence in oaks (Ranius et al. 2009a). Typical examples of bracket fungi involved in the heartwood rot of deciduous trees include: *Fomes fomentarius*, a widespread species found throughout northern Africa, Europe and eastern North America which regularly attacks oak, beech and birch, as well as cherry, willow and lime (Wood 2006); *Piptoporus quercinus*, another widespread European species, but specialising on oak (Crockatt et al. 2010); *P. betulinus*, a specialist on birches; and *Fistulina hepatica*, a common species across Europe specialising in oak and chestnut.

Rot-holes are the rarest of all saproxylic habitats in Europe (Müller et al. 2014), and the fact that three of the four Biodiversity Action Plan listed beetles in the UK are exclusively associated with rot holes is not a coincidence (Alexander 2003; Barnard 2011). Additionally, most of the beetles included in Annexes II and IV of the EU Habitats Directive are saproxylics (Campanaro et al. 2016), a direct consequence of forest management leading to a decrease in the volume, diversity and quality of dead wood, and subsequent regional extinctions of saproxylics (Grove 2002; Carpaneto et al. 2015).. Though only 11% of European saproxylic beetles (Coleoptera) are considered threatened with extinction, the poor knowledge surrounding the group means 28% remain Data Deficient, whilst 57% have an unknown population trend (Nieto & Alexander 2010).

In addition to ancient, unmanaged forests, man-made habitats can also be important refuges of biodiversity. Orchards are important habitats for saproxylic beetles throughout Europe, with many species of conservation concern utilising them as breeding habitat (Horák 2014). Though large veteran fruit trees on large stock roots are considered to be too small to sustain breeding populations of rot-hole breeding saproxylics in the long term (Horák 2014), with enough trees in an orchard and/or within dispersal distance of a particular species there is no reason why traditionally managed orchards cannot act as at least medium term (30-50 years) conservation habitats for rot-hole breeders. Large species of rot-hole breeders such as *Osmoderma* and *Propomacrus* have been found breeding in orchards (Tezcan & Pahlivan

2001; Dubois *et al.* 2009). Both genera hold extremely rare species (see S7.1), and in certain areas of their range orchards are likely to be of crucial importance to their continued survival.

Saproxylics face a number of threats at present, and have seen radical changes in their environment since the historical expansion of human populations (Horák et al. 2012). Considering the extensive loss of saproxylic old-wood habitat (Alexander 2008a; Nieto & Alexander 2010; Horák et al. 2012), a number of important questions relevant to conservation of saproxylic beetle species can be formulated. What is the demographic and genetic status of current beetle population compared to past populations, and can we predict future challenges facing them? At what sort of geographical scales do saproxylic beetle populations interact with each other? Has there been a loss of genetic diversity corresponding to a loss of habitat quantity and fragmentation? Can we make predictions about the molecular ecology (genetic diversity and distribution) of different saproxylic beetles based on what we know about their ecology?

1.3 Objectives and methods of Conservation Genetics

1.3.1 Genetic Variation

The fields of evolutionary and conservation genetics are both focused on the ability of populations to evolve in response to environmental change (Reed & Frankham 2001). A central principle to both fields is that genetic variation provides the basis for selection, adaptation and speciation (Amos & Harwood 2008). Therefore, as genetic variation decreases so does adaptive potential (Frankham 2005). Understanding the partitioning and persistence of genetic variation in natural systems is a fundamental objective for evolutionary biologists. Conservation genetics involves the application of genetic tools to the conservation of biodiversity (Frankham 2010). Following from the landmark paper by Frankel (1974), both areas have been increasingly amalgamated in the field of evolutionary conservation focused on conserving 'evolvability' of populations/species (Crandall et al. 2000; Allendorf & Luikart 2007).

The relationship between population size and genetic variation is well established (Young et al. 1996), and endangered species tend to have lower levels of genetic variation in comparison to related, non-threatened species (Frankham 1995). Alternatively, low levels of

genetic variation in seemingly healthy species/populations have been attributed to historical events such as founder effects, bottlenecks and metapopulation dynamics (Hoelzel et al. 1993; Hedrick 1996). In this context the effective population size is arguably the most important parameter in conservation genetics.

Effective population size (N_e) is an estimate of the number of individuals in an idealised population undergoing the same amount of random genetic drift as the actual population in question (Fisher 1930; Wright 1931; Lande & Barroclough 1987). This differs from census population size (the actual amount of individuals in a population, N_c), which can be extremely hard to estimate with rare or secretive species. Populations with low N_e often suffer from a disproportionate loss of genetic diversity over generations (Frankham 2005), potentially leading to inbreeding depression, whereby inbred individuals have lower reproductive success than non-inbred individuals (Crnokrak & Roff 1999; Reed & Frankham 2003), and potentially sending small populations into an extinction spiral. Differences between N_c and N_e (N_e is typically the smaller value) can be attributed to a number of situations: population crashes in species with disproportionate individual-level fecundity (Hauser et al. 2002); fluctuations in population size being more likely in large populations, causing N_e/N_c ratios to be smaller in large populations (Frankham 1995; Pray et al. 1996); N_e/N_c ratios increasing as N_c decreases due to reduced variance in male reproductive success (Ficetola et al. 2010). The N_e/N_c ratio is likely to be more predictable in species with low fecundity and low variance in reproductive success, but N_e and allele frequencies are usually stochastic in populations with overlapping generations, like some saproxylic beetles (Jorde & Ryman 1995; Luikart et al. 2010).

A well-established approach to estimating population parameters such as effective population size is the application of inherited (genetic) markers. Technological advances in genetic (i.e. DNA) marker development have resulted in a number of tools from allozymes, restriction-fragment length polymorphisms (RFLP) and microsatellites (Sunnucks 2000), through to single nucleotide polymorphisms (SNPs) (Moran et al. 2004) and the recent explosion in next generation sequencing and genomics (Hawkins et al. 2010). Genetic variation can be broadly considered as neutral, i.e. not subject to selection, or adaptive and subject to selection. Neutral genetic markers may be poor indicators of quantitative genetic variation and local adaptation (McKay & Latta 2002), however there is growing evidence of a correlation between neutral genetic variation and fitness (Ellegren 1999; Hedrick &Kalinowski 2000).Neutral genetic variation is, according to the neutral theory, genetic variation shaped

by the processes of drift, gene flow and mutation (Kimura 1983; Holderegger et al. 2006). Therefore, at the population level analysis of neutral variation can provide insights into patterns of population structure (Wright 1931), demography (e.g. population size changes and bottlenecks) (Garza & Williamson 2001; Palstra & Ruzzante 2008), speciation (Barluenga et al. 2006) and hybridisation (Roy et al. 1994).

Microsatellites, short simple sequences composed of tandemly repeated motifs (e.g. GTGTGTGT) of non-coding DNA are among the most popular neutral genetic markers (though there are numerous cases of selection effects detected at loci once presumed to be neutral), owing to their high mutation rates and widespread occurrence in genomes (Sunnucks 2000). Such high mutation rates can result in homoplasy and misinterpretation of long term genetic data (Hendrick 1999; Estoup et al. 2002; Brito & Edwards 2009) but also confers considerable power to elucidate fine scale processes, such as individual-level data about dispersals between populations (Paetkau et al. 1995; Keller et al. 2005). In addition to their utility in standard F_{ST} (a measure of the proportion of the total heterozygosity due to differences in allele frequencies among subpopulations within a population; Wright 1931) based analyses a major strength of microsatellites is there applicability to individual based assignment and clustering analyses. Such analyses offer the power to elucidate non-equilibrium processes (Slatkin 1993; Charlesworth et al. 1997).

Mitochondrial DNA (mtDNA) is the classic genetic marker for phylogeographic studies (Avise 1995). MtDNA has the special features that it is haploid, (usually) maternally inherited, (mainly) selectively neutral (Rand et al. 1994) and (usually) non-recombining. Also, the mutation rate is higher (at least for some regions of the mtDNA genome) than many regions of the nuclear DNA genome. The effective population size of mtDNA is ¼ that of diploid nuclear loci, assuming an equal sex ratio (Birky et al. 1983) meaning that the mtDNA locus may experience more genetic drift, and by extension greater differentiation, than nuclear loci (Birky et al. 1989). The lack of recombination makes it reasonably easy to reconstruct the phylogeny of mtDNA haplotypes. However, the assumption that such a phylogeny reflects population evolutionary history must be subjected to critical evaluation as mtDNA essentially represents just one, maternally inherited locus. Therefore, a prudent approach is to combine mtDNA and nuclear markers in assessments of evolutionary history (Rand & Harrison 1989; Knowles & Maddison 2002; Shaw 2002; Flanders et al. 2009). Typical non-genomic studies utilise a mixture of mitochondrial and neutral nuclear markers to provide a more detailed picture of the pattern of phylogeny, phylogeography and post-

glacial expansion in the study species (e.g. Holderegger et al. 2006; Vila et al. 2006; Říčan et al. 2008).

High levels of F_{ST} among populations usually denotes low levels of effective dispersal (dispersal including breeding success), whilst low levels indicate populations with high breeding mobility (such as long-distance reproduction or mobile populations) (Halliburton 2004). Gene flow between populations can happen in a number of ways, depending on the organism concerned and its habitat. Species restricted to patchy habitats (habitat 'islands'), such as saproxylic beetles, usually show gene flow consistent with island models, such as 'continent-island' where one-way gene flow occurs from a large population to a smaller one, 'island-island' where multiple isolated populations exchange genes, along with the Isolation by Distance model (Wright 1943) whereby the probability of two individuals breeding decreases as their distance apart increases. These three models can, in certain circumstances, influence population structuring via metapopulation models, whereby numerous local populations exist of varying sizes, with gene flow between sub-populations, and long-term patch extinction and recolonisation (Ranius 2000, S3.2). Combining such models can produce more complex multi-faceted models, such as the Extinction Ratchet (Templeton et al. 1990). Here, dispersal between populations is below a crucial threshold whereby suitable habitat patches (or islands) in which the species has become extinct (or were never inhabited) remain unexploited by the species, and thus each localised extinction (a click on the ratchet) renders the species more and more threatened, eventually leading to its extinction. Though usually studied in areas where there have been events that cause rapid population declines and extinctions, but with short-term regeneration of the habitat (such as after fires; Templeton et al. 2001), the applicability of this model to populations in the long term is poorly understood.

Population dynamics in small isolated populations can also result in the extinction of populations with high levels of mortality due to unsuccessful emigration from the patch (Thomas 2000), so if emigrating individuals are selected against, genes favouring dispersal will be lost from the population, and it may evolve to become more sedentary. This is a major issue in specialist species in 'simple' discrete habitats (Tscharnte et al. 2002), including mature trees with rot-holes. This was indicated by work on heathland ground beetles (Carabidae) which suggests that low dispersal ability correlates with vulnerability to decline in a fragmented landscape (de Vries et al. 1996). It may not be the ability to disperse long distances that endangers fragmented populations, but instead the proportion of reproductive adults making those dangerous emigrations.

Habitat fragmentation can reduce population connectivity over spatial scales by reducing the number of inhabitable patches from which an individual can leap-frog into a new, unoccupied habitat, thus reducing the rate or likelihood of recolonisation of patches predicted from the source-sink metapopulation model (Schmuki et al. 2006). Source populations tend to be larger on average for most animals, and work suggests that greater volumes of wood mould in trees supports more beetles, both at the tree level and for a fragmented forest (Ranius 2000; Rukke 2000). Both founder effects and genetic bottlenecking can take place in metapopulations in fragmented habitats, and these usually enhance genetic drift and reduce population fitness (Wade et al. 1996; Reed et al. 2003), unless selection is high enough to purge deleterious recessive alleles which otherwise could reach high frequencies (Reed & Frankham 2003). These events, plus inbreeding and strong directional selection, can all affect fitness negatively in a number of ways (Wade et al. 1996). Fitness is defined in a number of ways, but they all involve an assumption of higher fitness correlating with reproductive success and adaptations of lineages to changing environments (Reed et al. 2003; Fitzpatrick & Evans 2009).

Genetic variation is often spatially structured, whereby gene flow within and between populations (due to the movement of gametes, individuals or groups of individuals) can either constrain evolution by breaking up locally adapted gene complexes (Slatkin 1987; Huff *et al*. 2011), or it can promote evolution by spreading advantageous genes throughout a species (Wright 1931). Restricted gene flow between geographically separated populations should lead to the independent evolution of said populations via the allopatric speciation model (Mayr 1963). However, gene flow must exceed a certain level between lineages/populations to prevent substantial genetic differentiation (Wright 1931), lest speciation take place. Adaptive landscapes (synthesised from Fisher 1930 and Wright 1931) are theoretical models whereby populations exist in a 3D surface, with the X and Z axis controlled by measurements of phenotypic characters and Y representing mean fitness. Under selection populations should evolve along the phenotypic character axes toward the closest high point of mean fitness, but without declining in fitness. The landscape may have different areas for fitness peaks, which may end up causing speciation as different populations separate along the phenotypic axes toward distinct peaks of mean fitness (Slatkin 1987).

Effective population size can be measured with the equation $N_e = \pi/(4\mu)$, where π is nucleotide site diversity, and μ is the mutation rate per nucleotide site (Crow & Morton 1955; Wang 2005). Estimating N_e of structured populations (including metapopulations) can involve

subdividing metapopulations into local, interconnected populations (demes; Wakeley 2008). If migration among demes leaves relative deme size unchanged, then mean allele frequencies across demes also remains unchanged, mean within-deme nucleotide diversity therefore being an appropriate measure to compare across multiple species living within metapopulations, though the subtleties of their exact metapopulation behaviour (regularity of patch recolonisation and extinction, generation time, fecundity and population size, etc.) can complicate this measure (Nagylaki 1980; Wakeley & Aliacar 2001; Charlesworth 2009).

Adaptive genetic variation can be defined as heritable genetic variation that is directly subject to natural selection, and thus impacts an individual's fitness (Garcia de Leaniz et al. 2007). Certain alleles are favoured at the expense of others, and therefore genotypes are of adaptive or selective significance (Holderegger et al. 2006). However, these are generally genes that are monomorphic within populations because selection removes less fit variants. Adaptive variation can be assessed by a whole genome approach, which compares the genomes of multiple individuals, either within or between species, and identifies genes which harbour different alleles between the individuals or species (most commonly single nucleotide polymorphisms, SNPs; Hohenlohe et al. 2010). However, this is difficult to achieve for non-model taxa where reference genome sequence scaffolds are often separated by tens of millions of years of evolution, and as such bottom up approaches (where candidate genes are screened for allelic richness) can often prove to be more useful. Genes whose function is conserved across taxa, such as the Heat Shock Protein family (Li & Srivastava 2004) or the olfactory gene *Orco* (Jones et al. 2005), are sometimes investigated in phylogeography due to their ability to be reliably amplified across species using conserved primers, and their presumed functional importance.

Assessments of migration rates, patch extinction risks and short term population dynamics are short term goals able to be explored by high-resolution microsatellite markers, whereas analysis of historical processes shaping species distributions, the evolutionary history of the taxon in question, and the placing said taxon in a robust taxonomic framework, requires additional genetic markers, such as mtDNA or sequenced nuclear genes. However, analysis of both these fine and wide-scale processes can only take place on the back of a good sampling regime, with samples collected specifically to answer questions. But with rare, cryptic or endangered non-model species, sampling efforts often have to look beyond the usual methods of fieldwork and museum specimens.

1.3.2 Molecular Taxonomy

The IUCN Red List, the most important list of species at risk of extinction for its objective criteria (IUCN 2016), employs its primary unit of conservation as the species. However, the list currently incorporates a taxonomic bias away from invertebrates (Cardoso et al. 2011), hampering conservation efforts (such as preventing access to funding specifically targeting species with an IUCN Red List status). The Red List has been slow to catch on to modern developments in genetics, which suggest that morphologically defined subspecies may not represent true genetic diversity (Zink 2004); the word root "geneti" only appears nine times in the Red List Guidelines document (IUCN 2016), despite the IUCN designating genetic diversity to be one of the three forms of biodiversity deserving of conservation (Reed & Frankham 2003). As a list with the precise aim of moving beyond subjective opinion on species conservation status' (Rodrigues et al. 2006), incorporating genetics into conservation and taxonomy (allowing cross-clade analysis and comparison) should help push Red List-based conservation further along its current trajectory of being a resource free of bias with criteria assessable and comparable across taxa (Bowen 1999; Nybom 2004; Pauls et al. 2013).

There are also issues with assessing invertebrates on a species-by-species basis thanks to cryptic speciation which can't be detected by standard morphological analyses. For instance, Vila et al. (2006) assessed genetic diversity using four microsatellite loci alongside mtDNA sequences in six Iberian populations of the butterfly *Erebia triaria*, currently split into two subspecies. The combination of these neutral markers showed that alongside the previously identified subspecies, there were an additional two distinct genetic units which had not been identified morphologically, bringing the number of subspecies in Iberia to four. Many similar studies have reported disjunct divergent populations not identified by morphology (Williams 2002 on fritillary butterflies), misidentifications to species level of widespread endangered beetles (Audisio et al. 2009), elevation of subspecies to full species status (Morgan et al. 2000 on tiger beetles), and even genus-level divergences not previously identified by purely morphological methods (Hill et al. 2015 on American cicadas). Though still controversial, the continued development of DNA taxonomy and the reduced costs associated with it allows for hypothesis testing and evolutionary justifications for species descriptions, rather than relying on opinion (Vogler & Monaghan 2007).

Because of the problems with the species concept (De Queiroz 2011; Kunz 2013), the Evolutionarily Significant Unit (ESU) has been proposed as an alternative to the terms species, subspecies, geographic race, population or stock in certain circumstances (Crandall et al. 2000). By labelling conservation units as ESUs, we avoid misconceptions in weighting certain 'levels' of populations over and above others, to the potential exclusion of levels with less apparent 'worth' (Luck et al. 2003; Gompert et al. 2005). However, if ESUs are to be defined via genetic methods, then multiple loci need to be used to avoid oversimplifying evolutionary relationships (Rosenberg & Nordborg 2002), with both mtDNA and nuclear markers used. This also helps detect hybridisation, which is almost impossible to assess from maternally inherited mtDNA alone. MtDNA is usually more effective than microsatellites for investigating longer term processes in evolutionary history, and is the most commonly used marker. But a combination of mtDNA alongside multiple microsatellite loci is often the most effective way to answer most phylogenetic questions (Balloux & Lugon-Moulin 2002; Pamilo et al. 2005).

Correct species identifications, and accurate assessments of the distribution of species within genera, are extremely important for analysis of how individual taxa have responded to large scale temporal and spatial events such as glacial cycles, and how they will respond to future (and current) climatic changes (Habel et al. 2011; Pauls et al. 2013). Robust analysis of future and past species distribution changes can only be accurate if there is a robust taxonomy underlying the system, and that the species distribution data have been accurately reassessed in light of population genetics and cryptic taxa.

1.3.3 Historical processes shaping species genetic diversity

Populations of most temperate species (including saproxylic beetles) currently distributed across northern Europe spread to these areas in relatively modern times when the climate began to warm after the last glacial maximum (the LGM, between 23,000 and 18,000 years ago (Hewitt 2004; Shennan et al. 2006)) when permanent ice sheets and areas of permafrost receded northwards. Mild cooling-warming cycles have been occurring regularly (on a 41 thousand year (KY) cycle) since the start of the Quaternary 2.4 MYA (millions of years ago), but since 0.9 MYA the cycles have slowed to an increasingly dramatic 100KY cycle (Hewitt 2000), greatly affecting distributions and abundances of organisms in northern Europe (Hewitt 1999). These glacial cycles have received much research interest in Europe (Birks

1989, Hewitt 2011), particularly following early work on insects relating to the most recent post-glacial stage (the Holocene) by Lindroth (1969), Coope (1977) and Hewitt (1988). Approaching the coldest parts of the cycles when much of northern Europe was glaciated or under permafrost, many species would have become extinct throughout much of northern Europe as the climate shifted. This 'range contraction' (more accurately northern population extinctions) resulted in multiple southern refugia (localities with similar climates 20KYA to northern Europe in modern times) where much of Europe's fauna and flora survived in greatly reduced areas and abundances (Provan & Bennett 2008). Whilst physically isolated within particular refugia, populations evolve along their own trajectories, eventually becoming genetically distinct, a phenomenon usually analysed using barcoding genes (Hewitt 1999), allowing source populations to be identified for populations which spread following glacial retreat.

A traditional view from early work suggested three southern European (Mediterranean) primary refugia from which populations expanded following glacial retreat after the LGM: in Iberia, Italy and the Balkans (Taberlet et al. 1998), which were later expanded to include two further refugia in northern Africa and the Caucasus (Fig. 1.1; Schmitt 2007; Husemann et al. 2014). However this view, despite strong evidence supporting it, has been challenged and certainly does not hold for all species. Instead, more northerly refugia have been proposed for dozens of species (Schmitt 2007; Schmitt & Varga 2012; Tzedakis et al. 2013; also see Svenning et al. 2008), for example of trees (Wielstra et al. 2013), newts (Juřičková et al. 2014) and snails; Fig. 1.1). Postglacial colonisation routes from the three major southern European refugia have been well reviewed by Hewitt (1999; 2000; 2004), who proposed four generic patterns illustrated by key species types:

- The Grasshopper, from work on *Corthippus parallelus*, whereby populations expanded largely from the Balkans, meeting the trapped Iberian and Italian populations in the Pyrenees and Alps respectively and forming hybrid zones;
- The Hedgehog, from work on the sister species *Erinaceus europeus* and *E. concolor*, whereby populations expanded equally from all three primary refugia, producing hybrid zones throughout western and central Europe;
- The Bear, from work on *Ursus arctos*, whereby populations expanded from Iberia toward the North East, meeting populations from the Balkans expanding North West,

and forming a hybrid zone with trapped Italian populations in the Alps, plus a separate expansion from the Caucasus;

- The Chub, from work on *Leuciscus cephalus*, whereby populations expanded throughout Europe and into Russia from multiple minor refugia within the Balkans.

Habel et al. (2005) added a fifth pattern:

- The Butterfly, from work on *Melanargia galathea*, whereby populations expanded from the Italian and Balkan refuges to meet the trapped Iberian populations in the Pyrenees.

Similar approaches and concepts can be applied to plant species, including the tree species upon which saproxylic beetles depend. Despite complications arising from low species differentiation and haplotype sharing between widespread species within genera (Petit et al. 2002; Maliouchenko et al. 2007), patterns of post-glacial recolonisation are well understood for most European trees. The two common white oak species in northern Europe (*Quercus robur*, the pedunculate oak, and *Q. petraea*, the sessile oak) are wind pollinated across great distances, but the acorns produced are dispersed locally by animals. The two species also share mtDNA haplotypes (a group of mutations on a locus within mitochondrial DNA that is transmitted down the maternal line), but despite this complicating factor the post-glacial recolonisation pathways for white oaks are well understood (Hewitt 1999). Most haplotypes spread from Iberia and the eastern Pyrenees, through France and into Germany, the UK and Denmark (Petit et al. 2002), along with additional re-colonisations north east from Italy and the Balkans (Taberlet et al. 1998), a rough hybrid between the Bear and the Hedgehog patterns. Though the spread of *Quercus* is estimated at 150-500m yr^{-1} (Birks 1989), habitats suitable for most saproxylics will lag 100-300 years behind this, allowing for the trees to grow, senesce and for dead wood to accumulate.

Another giant tree important for saproxylics, the beech (*Fagus sylvatica*, Müller et al. 2014), has experienced a radically different pattern of post-glacial recolonisation. Combining genetic and fossil pollen data, Magri et al. (2006) demonstrated a spread throughout northern Europe from a refugium over modern day Slovenia (in a manner similar to the Chub), but that the speed of habitat colonisation was considerably slower than that of oaks; *Quercus* reached Ireland, Wales and southern England 9KYA, *Fagus* only reached southern England 3KYA (Birks 1989, Taberlet et al. 1998; Magri et al. 2006). Other beech populations spread north (southern Italy to northern Italy, the northern Alps in France to western France), but these did not greatly contribute to the overall post-glacial expansion of the species.

Relationships between the two main European tree-forming *Betula* (*B. pendula* and *B. pubescens*) birches are complicated, especially due to haplotype sharing between the two species and low species differentiation (Maliouchenko et al. 2007). In comparison to the oaks and beech discussed above, southern populations of tree-*Betula* did not extend northwards to a large extent, but instead current northern populations have been re-populated from northern refugia (Palmé et al. 2003), with genetic boundaries showing a distinct east-west split. These 'cryptic northern glacial refugia' are poorly understood, likely being small pockets of suitable habitat where favourable microclimates persisted (Provan & Bennett 2008; Tzedakis et al. 2013). Subfossil, pollen, genetic and radiocarbon data all suggest that tree-birches were found in a number of northern refugia, and were some of the first tree inhabitants of a post-glacial British Isles (Coope 1998).

The literature on other symbiotic relationships from a variety of systems strongly indicates that the post-glacial recolonisation pattern of symbionts will be similar (Nieberding et al. 2008, see Criscione et al. 2005 and Hoberg et al. 2012 for reviews). However, some parasite taxa with vector mediated dispersal do not show phylogeographic patterns consistent with their hosts. In contrast, research on fungi shows that they usually show population genetic structuring in concordance with their hosts, overturning older hypotheses that fungi populations were largely unstructured (Lumbsch et al. 2008; Gladieux et al. 2015; Sheedy et al. 2015), though there is much work to do in this area: no fungi involved in heartwood decay have yet been researched in Europe. Stag beetle females use yeasts grown within their mycangium to break down wood for their larvae to feed on, creating a symbiotic relationship which is reflected in the phylogenies of each group (Tanahashi et al. 2010; Tanahashi et al. 2016). This indicates that if other beetles are similarly dependent on symbiotic relationships with fungi, then not only their phylogenies but also their phylogeographic patterns should be similar between host trees, fungi and beetles.

As the trees and fungi form the habitat for saproxylic beetles, beetle distributions would be expected to closely track suitable habitat as it spreads (de Bruyn et al. 2011), though lagging behind the leading edge of host species until suitable wood decay habitats have had time to form (Hewitt 2000). As with the southern richness – northern purity hypotheses for trees (Hewitt 1996), similar patterns have been reported for saproxylic scarabs (Cox et al. 2013; Ahrens et al 2013). However, southern populations of trees are at risk of extirpation in Europe (Aitken et al. 2008) if they can't adapt to the changing climate. The 'southern edge' is usually the area of a species' range (for European terrestrial taxa) that harbours the most

genetic diversity, and contributes disproportionately more to the survival and evolution of taxa (Hampe & Petit 2005), and thus should be awarded additional conservation attention (Feliner 2011; Provan & Maggs 2011).

1.3.4 Morphological modelling and morphometrics

Accurate species delimitation is fundamental to research in macroevolution, biogeography, ecology and conservation biology, but the advent of DNA based taxonomy has in many cases contradicted traditional views on systematics and species identification based on morphological traits (Sites & Marshall 2003). Studies comparing traditional taxonomic approaches focussing on morphology with DNA-based inferences are needed (Wiens & Penkrot 2002; Raupach et al. 2016) for robust analysis of phenotypic and genetic divergence; a central topic in evolutionary biology (Slatkin 1987).

As body size can significantly influence the shape of arthropods (Chown & Gaston 2010), geometric morphometrics can be used to describe the shape of organisms using a series of landmarks to define discrete anatomical loci homologous across the study specimens (Zelditch et al. 2012). Morphological landmarks (biologically definable points on a shape) are the most commonly employed coordinate in geometric morphometrics, being more responsive to changes in 'shape angle' than outline methods (Adams et al. 2004). A Procrustes superimposition is often employed to find landmark configurations which are then fed into a canonical variate analysis (CVA) to maximise the differences between pre-defined taxa relative to the variation within taxa (Campbell & Atchley 1981; Rohlf & Slice 1990; Klingenberg et al. 2012), thus being an efficient method to detect differences between units, including species and sexes. Detection of significant pairwise differences in mean shape can then be tested using Mahalanobis distance (the distance in standard deviations of point P from the mean of D) (Klingenberg et al. 2012).

This technique is often applied to studies on Coleoptera body shape to answer a range of biological questions, in particular looking at variation within and between species and sexes (Bai et al. 2014; Eldred et al. 2016). This becomes increasingly powerful when combined with molecular approaches to identify subspecies and between-species relationships (Garnier et al. 2005; Zinetti et al. 2013; Ober & Connolly 2015), and analysing potential hybridisation between sister species (Pizzo et al. 2006). Work on *Carabus* ground beetles has demonstrated significant variation in shape between the sexes and between populations at scales under

200km (Alibert et al. 2001), whilst the technique has also allowed for the identification of cryptic, previously unidentified ground beetle species (Roggero et al. 2013).

1.3.5 Species Distribution Modelling

Understanding the interactions between genetic, phenotypic and environmental variables is a major step forward for conservation programs. Modern advances in genetics and computing power allow for the interactions between species and their habitats to be analysed in greater detail than ever before, and its ever reducing cost opens the field to analyses that would have been impossible until recently. For European terrestrial biota, key hotspots of genetic diversity are typically in the south in proposed glacial period refugia (Taberlet et al. 1998), which also matches with hotspots of species and subspecies richness (Myers et al. 2000; Schuldt & Assmann 2010). However, these southern areas are predicted to undergo drastic environmental changes due to effects of climate change, resulting in the potential extinction of endemic species in southern areas, and the extirpation of genetically diverse and distinct southern populations of other species which have spread to the north (Provan & Maggs 2011). Analysis of these distinct subunits, and hotspots for genetic diversity, can then be combined with species distribution modelling (SDM) with forecasting to account for climate change. This can show areas most at risk of extinction, allowing for conservation priorities to be drawn up, especially where these areas match with those with the highest genetic diversity.

As well as being able to make predictions for the future, Species Distribution Modelling (SDM) approaches can also help make inferences about species distributions in the past, in particular predicting likelihood and locations of glacial refugia (Provan & Bennett 2008; Svenning et al. 2008). This latter application of SDMs becomes particularly powerful when combined with genetic methods, allowing cross-discipline hypothesis testing (Knowles et al. 2007; Wielstra et al. 2013). In one of the most commonly applied and accessible approaches to developing SDMs, climatic data (such as the Bioclim bioclimatic variables dataset; Hijmans et al. 2005) are used alongside species presence data (such as Maxent, a machine-learning method which minimises relative entropy between the probability density of the species presence points and the wider study landscape; Phillips et al. 2004; Philips et al. 2006; Elith et al. 2011; Williams et al. 2015) to produce a robust model of species distribution with a statistical basis. Difficulties can arise, however, where the species used in the modelling is a habitat specialist, but the habitat itself is fragmented or locally patchy. In

the case of saproxylic beetles feeding on decayed trees for at least part of their life cycle (Alexander 2008a) their habitat is usually specific, such as heartwood rots in certain tree species caused by certain fungi. Models predicting changes in areas of suitable climate for saproxylic beetle species under climatic change scenarios will therefore lead to an over simplification of the species' niche: regardless of the suitability of the climate, if the habitat is not suitable the species will be unable to colonise an area. Good modelling approaches in these circumstances would involve not using high resolution climate grids to avoid model overfitting, multi-species comparisons from large datasets, and modelling the habitat as well as the dependent species.

Using multiple species under a single SDM framework allows for robust hypotheses to be drawn up from cross-species comparisons (Svenning et al. 2008). In the present study modelling was focussed on a range of species from Scarabaeoidea, including species specialising in certain habitats (climatic specialists, such as those associated with Mediterranean or boreal climates) or food sources (food specialists in rot-holes versus generalists feeding on a range of rotten wood, plus those species flexible enough to feed on other decayed biological material). It was predicted that there would be little difference in the glacial refugia utilised by food-source specialists and generalists, but much stronger responses to hind and forecasting by climatic specialists. Whereas climate change may prove to result in much of the Mediterranean region becoming unsuitable for some generalists, it may lead to expansions of Mediterranean specialists as suitable climates are found further north (Williams et al. 2015), replacing the generalist species in environments they are currently found in. Comparing the glacial refugia of this mixture of species will also allow for refugial hotspots to be identified and compared to insights from population genetics and phylogeography (Forester et al. 2013).

1.4 Study Species

European saproxylic invertebrates have received limited research attention, despite often being large, easy to find (with the correct trapping method) and indicators of healthy climax forests (Speight 1989). As saproxylics are usually specialists on breaking down woody tissue (Harmon et al. 1986), they perform unique roles in nutrient recycling: species specialise in different parts of the heterogeneous habitat, with nutrients travelling through saproxylics ending up spread throughout and between forests. Swift (1977) estimated the quantity of

nutrients recycled by saproxylics to be approximately 50% of the energy recycled by the annual leaf fall in deciduous woodland, but that the nutrients are recycled and spread throughout the year, not just in a single Autumnal event. Despite the interesting link between specialist saproxylics and their hosts, very little work has been focussed specifically on assessing patterns of post-glacial recolonisation and population structuring within species in forests, with small sample sizes or restricted sample distribution generally hampering any conclusions that can be drawn from general phylogeographic studies.

The work that has been completed to a degree sufficient to analyse refugia has shown interesting patterns: Ahrens et al. (2013) indicated an expansion from Balkan and Alpine refugia in the common *Cetonia aurata* (Scarabaeidae), with genetically distinct groups in Southern Italy, Lebanon, Iberia and Sardinia. Other more geographically restricted papers have indicated high levels of Italian and Balkan diversity in *Morimus* (Cerambycidae; Solano et al. 2013), complicated patterns of refugial diversity and sub-speciation with expansion from joint Iberian-Italian refugia in *Lucanus cervus* (Lucanidae; Cox et al. 2013, Solano et al. 2016, McKeown *pers. comm.*), a split between the north and south Balkans in *Rosalia alpina* (Cerambycidae; Drag et al. 2015), and expansion from one (unidentified, either Iberia or Italy) refugium throughout its western European range in *Osmoderma eremita* and likely expansion from a Balkan refugium in *O. barnabita* (Scarabaeidae; Audiso et al. 2008). Thus, unsurprisingly, there is not one pattern specific to saproxylics, but a range of different patterns which may, or may not, match with their host trees.

Research focussing on fine-scale interactions between saproxylic beetle populations has shown that species with greater levels of habitat specialisation show greater levels of population substructuring than generalists in a variety of systems. Schmuki et al. (2006) using allozyme markers showed log-dwelling saproxylic Tenebrionids in Australia to have clear (though shallow) genetic structuring over 10km areas of continuous native forest, but reduced gene flow between subpopulations in areas of fragmented native habitat percolated by alien plantations. These effects were obvious 21-36 years (approximately 10-15 generations) after the forests became fragmented. Oleksa et al. (2013) used Amplified Fragment Length Polymorphism (AFLP) to demonstrate that populations of the generalist rot-hole breeder *Protaetia marmorata* were more genetically diverse than populations of the specialist *Osmoderma barnabita* within a restricted habitat, and had substantially lower kinship coefficients (less within-population relatedness) below distances of 1000m. *Osmoderma barnabita* displayed high relatedness at local scales (Kinship = 0.3 within trees), with

kinship values falling within 95% confidence intervals from the null hypothesis of random genotype distribution only above 10,000m, whereas in *P. marmorata* kinship values are within random genotype expectations above 200m scales and kinship coefficients within trees = 0.05 (Fig. 1.2). The results clearly show that Isolation By Distance (IBD, where populations geographically close to each other are more genetically similar than each is to populations further away, due purely to distance affecting gene flow; Wright 1943) has a much greater effect on the genetic diversity of *O. barnabita* than of *P. marmorata*.

The Noble Chafer (Fig. 1.3, *Gnorimus nobilis* L. 1758) is a Trichiini (Cetoniinae) chafer belonging to the diverse beetle family Scarabaeidae. This study follows the traditional scarab taxonomy, where Scarabaeidae is a family within the superfamily Scarabaeoidea (Browne & Scholtz 1999). *Gnorimus* itself belongs to the subfamily Trichiinae (Trichiini), and is united here with the related European *Trichius (*Fig. 1.3). Trichiinae is sometimes listed as a tribe of the better known and more diverse group Cetoniinae, though other work has suggested that Trichiinae is a separate clade to Cetoniinae (Kalinina & Shabalin 2008) and is polyphyletic (Šipek et al. 2016; Fig. 1.4). In general, studies of scarab phylogeny are hampered by low sample sizes relative to the number of species in the family (30,000+ species), which explains the current state of flux (Ratcliffe & Jameson 2004). Here, *Gnorimus* is regarded as belonging to the subfamily Trichiinae within Scarabaeidae (Tauzin 2004).

The Noble Chafer is the focal study species throughout this work, acting as a model for saproxylic beetle conservation within the scope of the project. Other ecologically and taxonomically related species are also studied, including *Trichius* bee beetles, and the stag beetle *Lucanus cervus* (Fig. 1.3). The *Osmoderma* hermit beetles have also received much study throughout their range, so they act as a useful comparison to the other species, having similar life histories to *Gnorimus*. Other species are also discussed and analysed: *Gnorimus variabilis* (L.) and *G. subopacus* (Motschulsky, 1860) are used in the morphological study to compare to *G. nobilis*, in addition to a number of other saproxylic scarabs introduced with species distribution modelling. By cross-comparing species, taking into account their taxonomic and ecological relatedness, strong multi-species models can be tested and analysed in the context of knowledge regarding other model taxa, *Lucanus cervus* and *Osmoderma* (Fig. 1.3) in particular.

In Britain the Noble Chafer is currently restricted to orchards containing mature *Prunus*, *Malus* and *Pyrus* species (Philp 2006), with primitive *Prunus* species and cultivars being

preferred (Schenke 2012), feeding as a larva on 'pabula', a mulch of rotten wood, fungi and debris in rotten cavities within large trees (Smith 2003). Before the establishment of man-made orchard habitat, the Noble Chafer most likely lived in oak (*Quercus* spp.), beech (*Fagus sylvatica*) and willow (*Salix* spp.) dominated forests before they were largely destroyed and fragmented in Britain (Jessop 1986; Whitehead 2003). In the UK, the Noble Chafer is found only in England, with Worcestershire, Gloucestershire and Herefordshire being the stronghold regions, plus additional isolated populations in Kent, Oxfordshire and the New Forest (Schenke 2012). The New Forest population is enigmatic, with no records for years (Whitehead 2003; Schenke 2012), but as the species was still living in oak hosts when adults were last found (Smith *pers. comm.* 2012) there is no reason to assume that the population is extinct as the oaks are still standing. A recent push to better map the species' distribution in the UK has not led to any additional records of populations not in orchard trees (Bates *pers. comm.*). The orchards *G. nobilis* inhabits in the UK tend to be below 300m above sea level, whereas in Italy the species is semi-montane (Trizzino et al. 2013). This suggests that the ecology of the species in the UK will be significantly different to that of *G. nobilis* found further south in its range.

Within the UK, identifying how different the various Noble Chafer populations are from each other, particularly the isolated population in Kent and the oak-feeding population in the New Forest, will be extremely important in advancing our knowledge of how best to conserve the species. For instance, is there a chance that translocating some individuals from Herefordshire to Kent would break up adaptive gene complexes via outbreeding depression (Frankham et al. 2011; Huff et al. 2011), and is this a risk worth taking to bolster a flagging natural population? This analysis should also help if the decision is made to transfer some chafers to a new location that they may take centuries to colonise naturally (e.g., an orchard in Somerset); which populations should be sampled to ensure an adaptive mix of genes for the new population to survive over the long term (Hedrick & Fredrickson 2010)? That range shifts in saproxylic Coleoptera are taking place due to climate change is well understood (Hickling et al. 2006; de la Giroday et al. 2012; Lenoir & Svenning 2015), but the extent to which this will affect Noble Chafer populations is unknown. It is also unknown where suitable areas of climate will be found in the future, if Noble Chafer can disperse to these areas naturally, and if said areas have suitable host trees.

Gnorimus is found naturally on wild species of *Malus* and *Prunus*, though these are rare host species (Whitehead 2003). It is likely that during the Holocene, when *Gnorimus* moved into

Britain, it used the wild *Malus* and *Prunus* species present, such as *M. sylvestris* and *P. spinosa*, as hosts (Alexander 2008 b.). Though oaks may be considered as in 'early middle age' at 200 years old, and beech trees may be veterans at this age, *Malus*, *Prunus* and *Salix* would all likely be veterans at 70 years (Alexander 2008 b.; Hall & Bunce 2011). As *Salix* rots at a similar age to both common fruit hosts, and is considered to be a major *Gnorimus* host species, then it is possible that most British populations of *G. nobilis* were already feeding on such short-lived trees before the apparent switch toward orchard fruit trees recorded since 1900 (Whitehead 2003). The availability of wood mould in fruit trees peaks at 50-80 years of age, and thus they should be most suitable for *G. nobilis* then (Whitehead 2003), a considerably younger age than in oaks (Tyler 2008). Data on *Gnorimus nobilis* in orchards on continental Europe is lacking, but there is no evidence to suggest that it is in any way as tied to orchard trees there as in Britain. The possible switch in major hosts may in part represent a lack of sampling effort directed toward this elusive species, both in Europe and Britain: if only 15% of individuals are dispersing from their natal tree (assumed from *Osmoderma*, Hedin et al. 2008), and a dispersive movement takes 2 days (the mode in Svensson et al. 2011, with over 54.2% of dispersals above this time frame), then the chances of actually finding an adult will be rare without breaking open a tree. However, due to the interest surrounding the species, it is likely that the shift toward orchard trees does represent a genuine host shift, and is not due to a lack of sampling effort.

At present, traditional orchards are in a fragmented state in Britain, with gaps of many kilometres separating orchards. Traditional orchards are a Biodiversity Action Plan listed habitat (UK BAP 2008), and are defined "as groups of fruit and nut trees planted on vigorous rootstocks at low densities in permanent grassland; and managed in a low intensity way" (UK BAP 2008). With an estimated 25,350ha of traditional orchards in the UK (with 24,600ha (97.04%) in England), the habitat is more rare than upland oak woods (85,000ha), though due to the veteran characteristics common in traditional orchards, these will be more suitable for *G. nobilis* until the oaks have grown into veterans (UK BAP 2008). However, there are clusters of densely packed orchards, particularly in regions traditionally associated with cider manufacture (Somerset, Gloucestershire, Worcestershire and Herefordshire in particular), and it is no coincidence that within these regions the major populations of *G. nobilis* are found, with the notable exception of Somerset (PTES 2013). In England, loss of traditional orchards has reached 87% since the 1950s (PTES unpublished data). Habitat generation requires trees over 50 years old, so it will be many years before potential new habitat can be re-established

from scratch for *G. nobilis*. Studies on other British insects have demonstrated that after a long-term decline in population size, morphological traits may evolve in response to selection. The Garden Tiger moth, *Arctia caja*, has declined by 85% in 30 years, whilst the wings of the species evolved to be longer and narrower (Anderson *et al*. 2008), indicating that during and after the decline, selection favoured more mobile adults. Studies like this are difficult to complete for *G. nobilis* due to a lack of historical sampling as the species has always been considered rare in the UK (Whitehead 2003), though there is no reason that selection could not have favoured longer wings or further dispersive flights in the species.

Data for dispersal distances and regularity is lacking for *G. nobilis*. However, Whitehead (2003) estimated a maximum of 700m in a single flight, and Bates et al. (2014, unpublished) reached a similar conclusion, though they noted that flight patterns in the species were strongly controlled by ambient temperature and time after sunrise, the species rarely flying in the afternoon or on cooler days. As larvae are confined to rot-holes, with no clear method of dispersal, the only period in which the species can disperse is when they are adults. This stage usually lasts for six weeks (Whitehead 2003), though adults may only disperse for a fraction of that time. Eliasson (2000) also suggested that *Gnorimus* (*G. variabilis*) preferred sun-exposed canopies when flying, and were highly unlikely to fly at temperatures below 20°C. This suggests that the species can disperse throughout a medium sized orchard, but dispersals to orchards over 1km away are highly unlikely, and may only be possible if suitable food plants are in bloom between the two orchards for the disperser to replenish their fuel.

The Noble Chafer itself has some named subspecies and varieties (Tauzin 2004), though their validity is questionable as they may just represent intraspecific variation. Indeed, the primary criterion for splitting the nominate subspecies and the Russian *G. nobilis bolshakovi* is a green, rather than black, metatarsal protrusion, whilst over 40 colour forms have been 'officially' named. The final subspecies *G. nobilis macedonicus* is extremely similar to *G. n. bolshakovi*, but is restricted to Macedonia (Baraud 1992). In addition, other *Gnorimus* species will be incorporated into the analysis where possible, most notably *G. subopacus* (possibly the closest relative to *G. nobilis* which is found in the eastern Palaearctic) and *G. variabilis* (allopatric to *G. nobilis*, but a less thermally tolerant species specialising in oak and chestnut rot holes) (Tauzin 2004a; b).

A close relative of *Gnorimus*, the bee beetles (*Trichius*; Trichiini) will also be investigated alongside *Gnorimus*. The three common species across Europe (*T. fasciatus* L., *T. sexualis*

Bedel, 1906 and *T. gallicus* Dejean, 1821; Krell 2010; Krell 2012) are also heartwood-rot feeders, though they typically prefer smaller trees than *Gnorimus*, such as birch (*Betula* spp.). *Trichius fasciatus* is a very widely distributed species, from Iberia, northern Fennoscandia, and throughout Russia to Kyrgyzstan, which lives further north than any other rot-hole specialist scarab. This suggests that the species is tolerant of a wide range of environmental conditions: indeed, the species is tolerant to sub -10°C temperatures throughout its life span (Vernon & Vannier 2001). *Gnorimus* display similar levels of tolerance (though not to the same lowest lethal temperature; Renault et al. 2004), whereas *Osmoderma* shows freezing tolerance in the winter, but not in the summer. This suggests that in areas where frosts outside of winter are common, only *Trichius* and *Gnorimus* will survive thanks to their year-round tolerance to low temperatures. Birch, being cold tolerant, were found close to permafrost zones in the LGM, and thus didn't move northwards after glacial retreat as drastically as temperate tree species did (Palmé et al. 2003). If *Trichius* were utilising these hosts in the LGM, then their post-glacial recolonisation pattern should be different, spreading from northern, not southern refugia (Tzedakis et al. 2013). Additionally, as *Gnorimus* are more tolerant to freezing than *Osmoderma*, they may also have made use of northern refugia, providing a contrast between all three genera.

Some of the work on *Gnorimus* can be compared to its ecological and taxonomic relative, *Osmoderma*. Though less closely related to *Gnorimus* than previously assumed (Šipek et al. 2016; Fig. 3), the genus has a very similar life history to *Gnorimus*, feeding on accumulated pabula in rot-holes in *Quercus* and other deciduous trees in Europe (Landvik et al. 2016). They are also important ecosystem engineers, opening up habitat for other, smaller invertebrates within tree cavities (Chiari et al. 2014). Five species within the genus are listed on the IUCN Red List (IUCN 2010), with three listed as Endangered. Ranius (2006) found *O. eremitia* to have a mean dispersal distance of 60m, with no dispersals measured past 190m. The dispersal function predicted 1.6% of dispersing movements could be >250m, however, only 15% of adult beetles actually dispersed within their lifetimes. Therefore, in a population of 1000 adult individuals, dispersals above 250m may only be by 2 or 3 individuals each year. Similar results were found by Svensson *et al*. (2011), who estimated a mean dispersal distance of 82±18m, with 1% of dispersals above 1km. Estimating dispersal in small animals is difficult (due to effects of small sample sizes and rare events, weight of radio transmitters – approximately 20% of adult body weight, etc. (Svensson *et al*. 2011)), but the two separate studies arrive at similar conclusions. *Gnorimus* is likely to behave in a very similar manner,

with a ‘long-tail’ distribution of dispersals. Anecdotal evidence suggests that it can fly further than 1km in a dispersive movement: an adult was found on the steps of an abbey in Worcestershire 1.5km from the nearest known population (Green *pers. comm*. 2012), and *G. variabilis* breeds on the Estonian Island of Ruhnu (Voolma & Randveer 2003), a minimum of 40km from the nearest mainland population.

In a study of *Osmoderma eremitia* tree preference, Dubois *et al*. (2009) found 59 mature trees (total mature trees=7279) in the 16km^2 study site to have a rot-hole containing wood mould inhabited by *O. eremitia*. Two frequently inhabited trees were *Quercus robur* (43 trees with wood mould, 23 inhabited (53.49%)) and *Malus domestica* (94 trees with wood mould, 19 inhabited (20.21%)). *Osmoderma* is considered an oak specialist, largely avoiding *Acer* maples and *Fraxinus* ashes even when these have wood mould present in a rot-hole (Dubois *et al*. 2009). In Britain, much of the mature oak-dominated woodland has been felled (Hopkins & Kirby 2007), leaving only younger oaks which lack the rot-holes *Gnorimus* requires. Due to the use of orchard trees by *Osmoderma* and *Gnorimus*, these are vital to the survival of these two threatened genera.

Osmoderma has been demonstrated to exhibit metapopulation dynamics, with each rot-hole holding a group of inbreeding individuals (Ranius 2000; Ranius 2007). As approximately 85% of the adults don’t disperse but presumably breed in the rot-hole, each rot-hole should hold a group of inbreeding individuals, with a yearly migration of 15% (mean dispersal distance is 70m with a long tail dispersal distance reaching 1km (1% of individuals)) of the adults into and out of the rot-hole (Ranius 2006; Hedin et al. 2008). In a grove of oaks, each with a rot-hole inhabited by *Osmoderma*, following Levins’ (1969) original description of a metapopulation as “a population of populations”, both *Osmoderma* and *Gnorimus* should exist in a natural state in metapopulations. The Source and Sink model of Pulliam (1988) may well be present as well, though due to the slow life cycles of both *Osmoderma* and *Gnorimus*, patch extinction may only rarely be witnessed in the field (Ranius 2000; Whitehead 2003). Wood mould volume has a positive effect on the size of *Osmoderma* populations found within the tree (Ranius et al. 2009b), and as fruit trees have smaller volumes of wood mould than oaks, *G. nobilis* populations could be small per tree.

There are some issues with using *Osmoderma* as a model for *Gnorimus*. *Osmoderma* possesses large phenotypic differences between males and females, and males produce a pheromone used in mate attraction (Larsson et al. 2003), indicating female-biased dispersal.

Traps set up by Svensson et al. (2011) showed females dispersed over significantly greater distances than males (1331m±114m for females, and 48m±30m for males; $p<0.01$). However, this may have been due to the low abundance of short female dispersals, which stands in opposition to other work (e.g. Hedin et al. 2008). Genetic data also supports female-biased dispersal in *Osmoderma* (Oleksa et al. 2013). The only non-genital phenotypic difference between male and female *Gnorimus* is in the shape of the meso and metathoracic legs, which are spoon-shaped in males, and there is no knowledge of any male-produced pheromone. Males may have a larger lip at the front of the head, but there exists no data to show whether this is a reliable characteristic to differentiate the sexes or not.

The work conducted on *Gnorimus nobilis* will be compared and contrasted to results from the Stag beetle, *Lucanus cervus* (Lucaninae, Lucanidae, Scarabaeoidea (Smith et al. 2006)). The Stag beetle is another saproxylic (feeding mainly on fallen trees and branches of oak (Mann 2006)) on the BAP list, with partial legal protection in the UK (Smith 2003). Though the species has a wide range, it is threatened with extinction in many regions (Harvey et al. 2011). *Lucanus* have very different dispersal strategies to *Osmoderma* (and by inference *G. nobilis*): females disperse via flight within four days of emerging as an adult, and then rarely fly, preferring to walk (Rink & Sinsch 2007), whereas males fly consistently for 9 days after emerging, before they stop flying. Males also rarely walk, leading to male-biased dispersal differences (males- median 369.6m, range 0-2065m, n=18; females - median 51.1m, range 0-762.6m, n=38; Mann-Whitney *U*-test: $P=0.0051$ (Rink & Sinsch 2007)). Because of the effects of sexual selection, *L. cervus* is expected to possess different genetic population structuring to *Gnorimus*, including having more variable patterns in mtDNA than nuclear microsatellites due to male-biased dispersal (Section 4.3). Differences are also expected due to differences in the habitat structure of the two species: orchards are typically well defined, small 'clumped' habitats, whereas the forests *Lucanus* inhabit are larger on a geographical scale, with suitable larval food sources of a range of sizes (from tree trunks to small sections of branches).

Due to the small, localised populations of *Gnorimus nobilis*, combined with dispersal abilities unlikely to be great enough to link populations over 2-3km, it is expected that outlying populations are genetically distinct from the main population centre (both at a European and inter-UK scale). Outlying populations within the UK are probably relicts from when the species was more widespread (Smith 2003; Whitehead 2003) separated for a minimum of 100 years following the destruction of the native old-growth forests and a subsequent population

crash. Being cut off from other populations, the outlying populations are also likely to be suffering from inbreeding effects. Reduced genetic diversity inferred from mtDNA haplotypes (a group of mutations on a locus within mitochondrial DNA that is transmitted down the maternal line) after a population crash can indicate increased inbreeding effects, and a reduced long-term evolutionary potential (Frankham 2005). However, despite the population crash in *Arctia caja*, genetic diversity and mutation rates were still high enough to cause an evolutionary shift toward possessing longer wings (Anderson et al. 2008), due to the selective pressure on the moths resulting in increased dispersal ability.

1.5 Aims and Objectives

This project will apply a suite of genetic markers, morphological taxonomy and Species Distribution Modelling to provide complementary information on the historical processes acting on the genomes of European saproxylic beetles (focussing on *Gnorimus nobilis*, but also using *Lucanus cervus* and *Trichius* spp. for comparison), including glacial period contractions and recolonisations, host shifts, population connectivity, and historical and contemporary habitat fragmentation. The primary questions for the study are:

- Which standard patterns of post-glacial recolonisation do *Gnorimus nobilis*, *Lucanus cervus* and *Trichius* spp. adhere to, and can any differences be attributed to their biology?
- What is the status of the three proposed subspecies of *G. nobilis*, and are they reliably identifiable by genetic and morphological means?
- How closely related are the three *Trichius* species, and are their morphological differences robust enough to allow for identification in field conditions?
- Do markers under selection show similar patterns to neutral markers in analysing *Gnorimus* phylogeography? Does the "southern richness, northern purity" paradigm still hold for genes under selection?
- Can species distribution modelling be used in tandem with multi-locus phylogeography to provide the best prediction of glacial refugia possible, or are there too many dataset biases? Is there a general "saproxylic post-glacial recolonisation pattern", or are different species predicted to have utilised different refugia?
- How strongly do previous results from other researchers match with SDM predictions about glacial refugia?

- Can areas predicted to become unsuitable for rare saproxylic scarabs be identified and can new climatically suitable areas be identified with climate change predictions?
- How robustly do the three main methods utilised (SDMs, morphological modelling, and phylogeography) support each other, and where are the weakest areas of support?

It is predicted that both *Gnorimus nobilis* and *Lucanus cervus* will show reduced genetic diversity in the UK compared to populations from mainland Europe, particularly those populations in areas predicted to have acted as glacial refugia. These may be in contrast to *Trichius*, which might show higher genetic diversity in norther regions thanks to possible use of northern glacial refugia. These methods will be used to generate phylogenies of both species on European and UK scales, showing patterns of post-glacial recolonisation and the genetic effects of habitat loss. It is expected that habitat loss for both *Gnorimus* and *Lucanus* will have impacted their genomes by increasing localised inbreeding, encouraging local adaptation and discouraging potentially fatal unsuccessful dispersals. Microsatellite analysis can identify kinship patterns within a tree and an orchard, and may show how regularly individuals emigrate from one tree to another in an orchard, or potentially further. Using sample sites across the UK distribution of *G. nobilis*, along with multiple frass samples per tree and orchard, will show how homogenous UK *G. nobilis* populations are, plus if there is any way to rescue isolated populations genetically via reintroductions (Hedrick & Fredrickson 2010).

The work aims to produce a comprehensive genetic map of UK *Gnorimus nobilis*, concentrating on its current distribution, but also using museum and European specimens to provide a wider temporal and spatial analysis of the genetics of this species. By using a variety of techniques, including next-generation sequencing, the genetic effects of rapid habitat loss on the species will be quantified.

1.6 General Abbreviations

ANOVA – Analysis of Variance, a group of statistical models used to analyse differences among group means

ATCG – Adenine, Thymine, Cytosine and Guanine, the fundamental base codes which make up DNA

BAP- Biodiversity Action Plan, an international program to address the threats facing specific species

COI – Cytochrome Oxidase 1, a mitochondrial gene commonly used in barcoding

CR – Control Region, a mitochondrial gene used in barcoding

CTAB - Cetyltrimethyl ammonium bromide, a buffer solution

CVA – Canonical Variate Analysis, an analysis method for identifying pre-defined taxa within a 'population' of variable taxa

DNA – Deoxyribonucleic Acid, the molecule carrying the genetic instructions for living organisms

ESU – Evolutionarily Significant Unit, a monophyletic unit of evolution

F_{ST} - the most important of Wright's *F*-statistics for assessing subpopulation structuring

IUCN – the International Union for Conservation of Nature, the global authority on the status and conservation of the natural world

KYA – thousands of years ago from present

mtDNA – mitochondrial DNA, found within eukaryotic cell mitochondria

MYA – millions of years ago from present

n – sample size

N_e – effective population size, the number of individuals in an idealised population undergoing the same level of genetic drift as an actual population

P – the p-value, the probability that the statistical result would be the same as or more extreme than the actual result

PCA – Principle Component Analysis, a method of converting a set of possibly correlated observations into a smaller set of uncorrelated principle component values

PCR – Polymerase Chain Reaction, a technique for amplifying a copy of a piece of DNA over several orders of magnitude

RAD-seq – Restriction Site Associated DNA Sequencing, a fractional genome sequencing method

SDM – Species Distribution Modelling, a suite of methods for predicting species presence in mathematical enviro-space

Figure 1.1. Major and minor refugia during the last glacial maximum. R refugia represent southern Mediterranean refugia, M refugia represent northern refugia. The light blue line shows the approximate extent of the northern ice sheets, whilst the dark red line shows the approximate northern extent of woody vegetation. R1: the Maghred, R2: northern Iberia, R3: eastern Iberia, R4: southern Italy/the Adriatic, R5: the southern Balkans, R6: Turkey, R7: the Caucasus. M1: southern France, M2: eastern France, southern Germany and Switzerland, M3: eastern Alps, Slovenia and Austria, M4: Dinaric Alps, M5: the western Carpathians, M6: the eastern Carpathians, M7: the Balkan mountains. Additional minor refugia occur on Mediterranean islands and likely through Russia. Data from Schmitt 2007; Habel et al. 2010; Tzedakis et al. 2013, Juřičková et al. 2014.

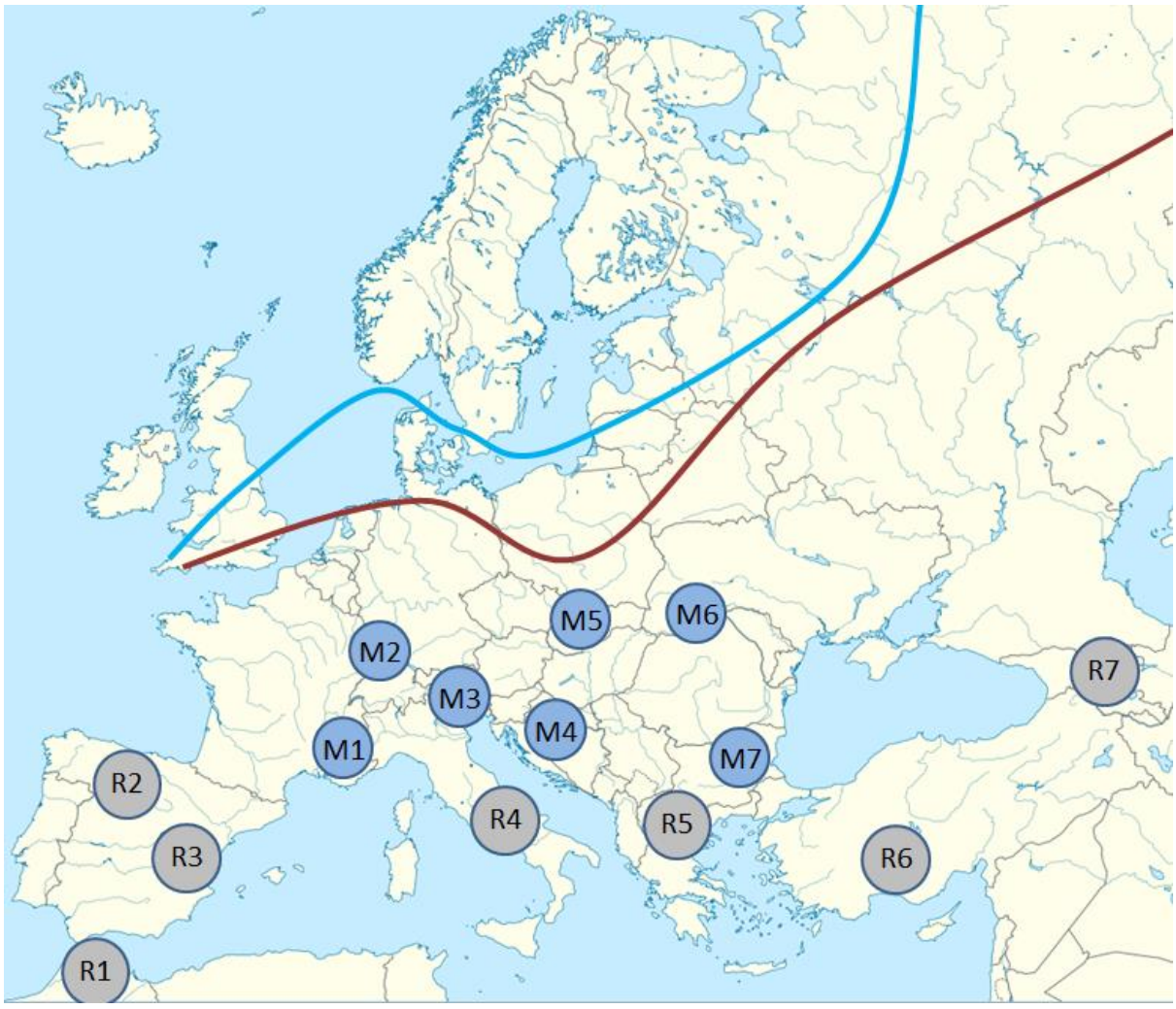


Figure 1.2: Kinship coefficients between pairs of a habitat specialist (*Osmoderma barnabita*, top) and a habitat generalist (*Protaetia marmorata*, bottom) against logarithmic geographic distance in the whole population, with a 95% confidence interval from randomly distributed genotypes (dashed line) and standard error from jack-knifing over loci (error bars) (modified from Oleksa et al. 2013).

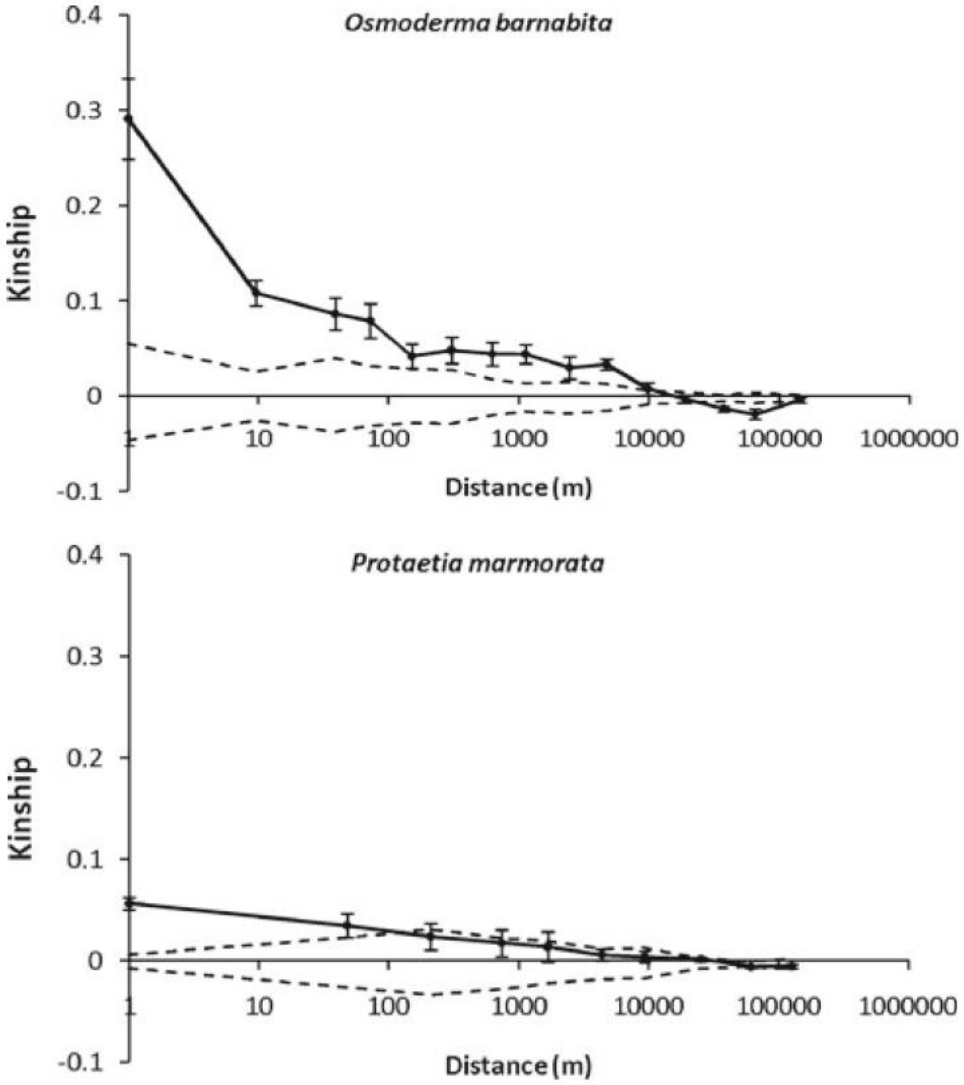
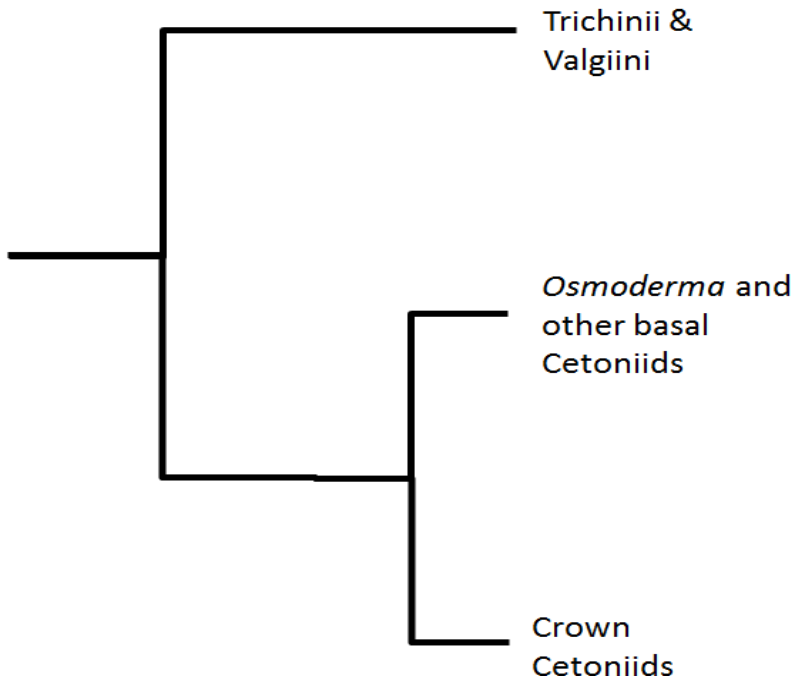


Figure 1.3. The main study species. Clockwise from top left, *Gnorimus nobilis*, Max Blake; *Trichius gallicus* (by Wikipedia user GabrielBuissart, public domain); *Lucanus cervus* (by Wikipedia user Orchi, public domain); *Osmoderma eremita* (by Wikipedia user Magnefl, public domain)

Figure 1.4: relationships between groups within Cetoniinae (data from Šipek et al. 2016)



Chapter 2

Development of microsatellite markers and a preliminary study of nuclear and mitochondrial DNA diversity in the Stag Beetle (*Lucanus cervus* L.)

2.1 Introduction

Analysis of genetic population structure provides information as to the geographical scale at which evolutionary processes can occur for a species, and by extension the spatial scales on which conservation management may need to focus (Waples 1995). Analysis of genetic diversity can also be used to identify populations with unusual genetic characteristics and to identify populations with low levels of genetic variability that might have reduced probability of persistence (Avise 2004).

Genetic structure and diversity within and among populations are influenced by both historical and recurrent processes. Genetic structure inferred from analyses of mitochondrial DNA (mtDNA) often reflects long term demographic processes associated with historic geological events, such as glaciation (Tabarlet et al. 1998; Hewitt 2000) and can be used to elucidate evolutionary heritage among populations, to identify evolutionary significant units (ESUs; Moritz 1994) and clarify taxonomic uncertainties, for the conservation of evolutionary diversity in cryptic species complexes, subspecies and ecologically isolated populations (e.g. Hebert et al. 2004; Segelbacher & Piertney 2007). Genetic analysis of more rapidly evolving markers, such as nuclear microsatellites (Ellegren 2004) can provide information as to recurrent population dynamics (Balloux & Lugon-Moulin 2002). In addition, microsatellites are useful for studies of kinship, parentage and inbreeding (Queller & Goodnight 1989; Hadfield et al. 2006; Wange 2011), which are amongst the most important parameters in conservation genetics (Allendorf & Luikart 2007).

Saproxylic invertebrates have been identified as one of the most threatened components of European fauna (Nieto & Alexander 2010) with their declines largely attributed to intensification of forestry and agricultural management practises. The European stag beetle (*Lucanus cervus* L.) is undoubtedly the most charismatic and popular saproxylic beetle in Europe and is regarded as the most emblematic flagship species for biological conservation in Europe, as well as habitat conservation for saproxylic beetles (Thomaes et al. 2008; Campanaro et al. 2016). It occurs throughout Europe but evidence from several countries suggests that its range is decreasing and it has become extinct in Denmark (Tochterman 1987; van Helsdingen et al. 1995). As a consequence it has been listed in the IUCN Red list of Threatened Species as "near-threatened", in the EU Habitats Directive and as a priority species of community interest (Appendix II) (Nieto & Alexander 2010). In the UK, it has been classified as 'Nationally Scarce, category B' (Percy et al. 2000).

Despite the abundant formal recognition of the threatened status of the stag beetle, conservation plans are fundamentally limited by a lack of population data. For holometabolous insects, where larvae and adults exhibit different ecological requirements it has been suggested that studying adults represents the most effective approach for monitoring populations (Tikkamaki & Komonen 2011). However, for stag beetle, attempts to estimate abundance using various approaches such as mark-recapture and lures have often yielded contradictory results. As such there are considerable uncertainties regarding the reliability of ecological parameters (e.g. abundance, dispersal) inferred using such methods. As genetic markers offer a powerful approach to understand such processes the objective of this study was to (i) develop microsatellite markers as a resource for stag beetle conservation and (ii) to perform a preliminary analysis of microsatellite and mitochondrial DNA diversity between British and continental European samples.

2.2 Materials and methods

2.2.1 Sample collection and DNA extraction

For microsatellite and mitochondrial DNA analysis, leg samples were obtained from across the UK region (n=48, collected by Dr Dave Chesmore (University of York)) and from Spain (n=52, donated by Dr Deborah Harvey (Royal Holloway University of London)). No finer geographical information or sex data was available but the UK samples were obtained from a much wider geographical area than the Spanish samples which were predominantly collected around Madrid. For mtDNA analysis sequences were also obtained for individuals collected from Italy (n=1), France (n=15), Romania (n=2) and Germany (n=2). DNA was extracted using a standard CTAB-phenol-chloroform-isoamyl alcohol method unless DNA was to be used for RAD-seq (see below), in which case it was extracted using a Qiagen DNeasy Blood and Tissue kit.

2.2.2 Identification of microsatellites

Microsatellite arrays were identified by two methods (1) construction of a microsatellite enriched genomic library and (2) exploration of sequences obtained from a Restriction

enzyme Associated DNA sequence (RAD-seq) library generated as part of a European wide collaboration investigating stag beetle conservation genetics.

1. Microsatellite-enriched genomic library construction: An enriched library was constructed following methods outlined in McKeown & Shaw (2008). Genomic DNA was digested with the restriction enzyme *Rsa*I (New England Biolabs) and the blunt ended fragments ligated to double-stranded SuperSNX linkers. Enrichment was then performed by selective hybridization of biotin-labelled repeat motif oligonucleotide probes [$(TG)_{12}$, $(GA)_{12}$, $(AAAT)_8$, $(AACT)_8$, $(AAGT)_8$, $(ACAT)_8$, $(AGAT)_8$] with hybridised complexes captured using streptavidin-coated magnetic beads (DYNAL) and unbound DNA removed by a series of washes. DNA fragments were then eluted from the magnetic beads and amplified by polymerase chain reaction (PCR) using the SuperSNX24F oligonucleotide. The PCR products were cloned using the TOPO-TA cloning kit (Invitrogen) and recombinant colonies identified by disruption of β-galactosidase activity. Recombinants were individually transferred into 50 µl of water and incubated at 95^{o}C for 10 min to promote plasmid DNA release. 1µl of each plasmid extract was subjected to PCR using M13 forward and reverse primers. The PCR mixture contained 1X buffer, 1.5mM MgCL2, 0.2mM dNTPs, 0.2U of *Taq* DNA polymerase (Bioline, UK), 1pmol of each primer and the thermoprofile consisted of 30 cycles of [95^{o}C for 30 sec, 50^{o}C for 30 s and 72^{o}C for 30 s]. PCR products were then sequenced directly using the internal T7 vector primer.

2. RAD-seq: Total genomic DNA was extracted using a Qiagen DNeasy Blood and Tissue kit according to manufacturer's protocols. DNA was quantified and 2.5µg per individual used to generate and sequence RAD tags following the methods outlined by Baird et al (2008), Hohenlohe et al. (2010) and Emerson et al. (2010). In brief, sequencing adaptors and individual barcodes were ligated to *Sbf* I-digested total genomic DNA, and the resulting fragments were sequenced from the restriction sites. RAD samples were jointly sequenced on a MiSeq (Illumina) and demultiplexing, quality scoring and trimming (to 120 base pairs) performed using the software CLC Workbench. Sequences were then surveyed for the presence of microsatellite arrays using MICROSAT commander.

2.2.3 Microsatellite genotyping and statistical analysis

PCR primers were designed from sequences flanking microsatellite arrays using PRIMER 3.0 (Koressaar & Remm 2007). For each locus the respective forward primer was labelled with a fluorescent dye at the 5'-end (Life Technologies). For genotyping the UK and Spanish samples each locus was individually PCR amplified in a 10µl reaction containing 100-200ng of DNA, 5µl Biomix (Bioline, UK) and 0.2pmol of each primer. PCR thermoprofiles included an initial denaturation step (95^{o}C for 3 min) followed by 35 X [95^{o}C for 30 sec, 55^{o}C for 30 sec and 72^{o}C for 30 sec]. Amplicons were separated using an AB3500 (Applied Biosystems), and alleles subsequently designated using the GENEMAPPER software (version 4.1, Applied Biosystems).

Numbers of alleles (N_A), allelic richness (A_R; El Mousadik & Petit 1996), observed heterozygosity (H_O) and expected heterozygosity (H_E), were calculated using FSTAT 2.9.3.2 (Goudet 1995). Genotype frequency conformance at individual loci to Hardy-Weinberg equilibrium (HWE) expectations and genotypic disequilibrium between pairs of loci were tested using exact tests with default parameters in GENEPOP 3.3 (Raymond & Rousset 1995). Genetic differentiation among samples was quantified by the unbiased F_{ST} estimator, θ (Weir & Cockerham 1984) calculated in FSTAT, with significance assessed by 1000 permutations of genotypes among samples. Genetic heterogeneity was also assessed using the Bayesian clustering analysis implemented in the programme STRUCTURE (Pritchard et al. 2000), which was used to identify the most likely number of genetic clusters (*K*) in the data (from comparison of models of *K* =1-4). Each MCMC run consisted of a burn in of 10^6 steps followed by 5 X 10^6 steps. Three replicates were conducted for each *K* to assess consistency. The *K* value best fitting the data set was estimated by the log probability of data [Pr(X/K)].

2.2.4 mtDNA sequencing and analysis

GenBank sequences for *Lucanus cervus* were used to design the species specific primers [F: 5'-ATGGCAATTGGCCTTCTTG-3'; R: 5'-CGTAATGAAGAGAATGCCTCTCAG-3'] permitting PCR amplification of a 750bp stretch of the Cytochrome Oxidase I (COI) gene. PCRs were performed in a total volume of 20ul, containing ~100ng of template DNA, 1 uM of each primer, 10ul Biomix and using a thermoprofile of 3min at 95^{o}C, followed by 35 X [30s at 95^{o}C, 30s at 52^{o}C and 45 s at 72^{o}C] followed by a final 3 min extension at 72^{o}C. PCR

products were purified using ExoSAP IT and sequenced from both ends with the PCR primers on an ABI 3130 DNA sequencer. Sequence chromatograms were examined and edited in CHROMAS. Sequence alignment was performed using the CLUSTAL W (Thompson et al., 1994) program executed in BIOEDIT (Hall 1999) with adjustments made by eye where necessary. Sequences collected by Cox et al. (2013) were included in the sequence alignment.

All analysis was performed using ARLEQUIN 3.1 (Excoffier et al. 2005) unless stated otherwise. Genetic variation was described using indices of haplotype and nucleotide diversity (*h* and π respectively; Nei & Tajima 1981; Nei 1987) and their variances. A minimum spanning network was constructed in NETWORK (www.fluxus-engineering.com/sharenet.htm). Fu's *Fs* (Fu 1997) and Tajima's *D* (Tajima 1989) tests were used to test for deviations from mutation-drift equilibrium that could be attributed to selection and/or population size changes. Mismatch distributions (Harpending 1994), the frequency distribution of numbers of pairwise differences between haplotypes within a sample, and simulated distributions under a model of demographic expansion were compared with the sum of squared deviations (SSD) between observed and expected distributions (significance assessed after 10 000 bootstrap replicates) used as a test statistic.

2.3 Results

From the enriched genomic library, screening of 48 clones yielded 2 microsatellite arrays with suitable flanking regions for PCR primer design. One locus (Lc-1) was chosen for genotyping samples. From the RAD-seq library over 300 unique sequences containing microsatellite arrays (>10 repeats of a dinucleotide repeat) with sufficient flanking sequence for PCR primer design were identified. Four loci (Lc2-5) were randomly selected for genotyping of samples. Primer sequences are given in Table 2.1.

Overall each of the five assayed loci were polymorphic [total Na per locus: Lc-1 (4), Lc-2 (8), Lc-3 (7), Lc-4 (4), Lc-5 (10)] with an average of 7 alleles per locus. The UK sample exhibited a markedly lower level of genetic variation than the Spanish sample (Table 2.2), with Lc-2 and Lc-3 being fixed for a single allele among UK samples. No significant deviations from random associations of genotypes between loci were detected (global test with UK and Spanish samples pooled). For the Spanish sample each locus exhibited a significant heterozygote deficit (Table 2.2). Due to the fixation for single alleles, tests of

HWE could not be performed for Lc-2 and Lc-3 for the UK sample, but no significant deviation from HWE was detected for Lc-1 and Lc-5 while a significant heterozygote deficit was reported for Lc-4 (Table 2.2).

STRUCTURE analysis unanimously supported a model of K=2 (posterior probability = 100% and nearly zero for other K values tested) wherein the UK and Spanish samples were partitioned (Fig. 2.1). The corresponding F_{ST} was 0.287 and highly significant (P<0.0001).

For the *de novo* COI sequenced samples [UK (n=48); Spain (n=52); France (n=15); Italy (n=1); Germany (n=2); Romania (n=2)] 725bp of sequence could be aligned across all individuals. However, to facilitate alignment with data from Cox et al. (2013) this was trimmed to 325bp. This did not affect the number of haplotypes resolved.

Across all samples a total of 26 haplotypes were identified (Table 2.3). Haplotype 1 was the most abundant (present in 76% of sequences), with the remaining haplotypes occurring in one or two individuals (h = 0.424). The haplotype network revealed haplotype 1 to be at the centre of a star shaped genealogy wherein adjacent haplotype were separated by one site difference in all but two cases (Fig. 2.2; π=0.002). Significantly negative values were obtained for Tajima's D (-2.2504; P<0.0001) and Fu's Fs (-31; P<0.0001) across all samples and mismatch distribution was compatible with a model of rapid population expansion (PSSD=0.86). Despite the similar sample sizes mtDNA diversity was notably lower among UK samples ($n$hap=4; h=0.12 (SD=0.064)) than the Spanish samples ($nhap$=13, h=0.5561 (SD = 0.084)).

2.4 Discussion

This study reports on the isolation and characterisation of the first microsatellite loci for *Lucanus cervus* and by extension represents the first assessment of presumably neutral nuclear genetic diversity for the species. The assayed microsatellite markers were all polymorphic and produced clear PCR products. As the markers obtained from the RAD-seq library were randomly chosen from a large (>300) number of unique microsatellite containing sequence reads with sufficient flanking sequences to develop PCR primers, this indicates that such untested microsatellites are likely to be useful markers for future studies

For the Spanish sample each of the microsatellite loci exhibited strong heterozygote deficiencies. Strong heterozygote deficiencies at microsatellite loci have been widely

reported for invertebrates (e.g. Zouros & Foltz 1984; Huang et al. 2000) and can be attributed to technical artefacts (e.g. null alleles), selection, inbreeding and/or Wahlund effects. While null alleles are common among microsatellites, a recent paternity study within our research group comparing parent and offspring genotypes from mating experiments has indicated no evidence of null alleles at the assayed loci (McKeown *pers. comm.*). Selection can also be discounted as while there is increasing evidence of selection effects at 'outlier' microsatellite loci (Nielsen et al. 2009), for selection to be occurring at all five loci assayed here must be considered extremely unlikely. Therefore it seems that the patterns have likely been generated by inbreeding and/or a Wahlund effect. Inbreeding could occur as a consequence of small population size and would be is consistent with the overall low levels of genetic diversity. However, for the Spanish samples haplotype diversity was moderately high, contrary to the prediction that mtDNA diversity would be lost more rapidly in a system with prevalent inbreeding suggesting that the heterozygote deficits may not reflect inbreeding *per se*. Wahlund effects occur when two or more genetically different groups are sampled as a single population. Although STRUCTURE analysis failed to detect any substructuring within the Spanish sample this may be due to the resolution threshold of the analysis (Latch et al. 2006). As the sample consisted of pooled individuals collected across Madrid and surrounding areas it is possible that multiple populations may have been sampled and thus, the heterozygote deficits reflect cryptic population structuring.

A second aim of this study to compare genetic diversity (nuclear and mitochondrial) among UK samples with sampled continental European populations. This was envisaged as a preliminary study rather than the compilation of a full European phylogeography and therefore, *de novo* sampling was restricted to a few locations. While including relatively few individuals per location is not unusual (e.g. Taberlet et al. 1998; Questia et al. 1999) future analyses should sample a greater geographical range. A striking feature of the microsatellite variation was the low levels of variability among UK samples compared to the Spanish samples. This pattern of lower diversity among UK samples was also evident in mtDNA diversity, for which some additional geographic samples were included.

The mtDNA phylogeny revealed a shallow level of divergence, conforming to a typical star-shaped phylogeny wherein a central, most abundant haplotype was surrounded by a number of closely related low frequency haplotypes. The lack of any mtDNA diversification suggests the predominance of a single clade across western Europe, and although based on a small portion of the mtDNA genome is in agreement with the more extensive data from Cox et al.

(unpublished) suggesting that western European stag beetles can be regarded/conserved as a single ESU.

The star shaped phylogeny is expected in populations that have undergone a historical decrease in size followed by rapid expansion (Grant & Bowen 1998) because population bottlenecks decrease molecular diversity and rapid population growth enhances retention of new mutations (Slatkin & Hudson 1991; Rogers & Harpending 1992). The demographic tests (Tajima's *D*, Fu's F_S, mismatch distribution) were also compatible with a population expansion. Such population contractions/expansions can almost certainly be linked to Pleistocene glacial/intergalcials. The higher diversity among Spanish samples is also consistent with a classic northward pattern of postglacial recolonization from refugia in Southern Europe (Tabarlet et al. 1998), and matches a similar pattern in a primary stag beetle larval food source, oak trees (*Quercus* spp., Hewitt 2000; Petit et al. 2002). While the specific phylogeography of the British Isles was poorly resolved due to low levels of mtDNA variation, there was no evidence of colonisation from more than one lineage (e.g. the Celtic fringe scenario; Searle et al. 2009).

Overall levels of microsatellite diversity were low but comparable to values reported in other saproxylic beetles (Drag et al. 2013). Rather than reflecting low genome wide mutation rates such low levels of variation are more likely to be a consequence of stochastic loss of alleles in small isolated populations (Drag et al. 2015). Despite the low overall level of microsatellite variation, the UK samples still exhibited a significantly lower level of nuclear and mtDNA variation comparted to their continental counterparts. The low genetic diversity among UK samples might be a consequence of founder effects during post-glacial south-north colonisation events. Furthermore, as there are no historic nuclear genetic diversity data available to compare to contemporary diversity it is not possible to ascertain whether the observed genetic patterns reflect more recent genetic erosion. British stag beetle populations are at the most north-west of the species distribution, and confined to the milder southern areas. British populations, isolated and constrained by historical and contemporary climatic factors have probably undergone higher levels of genetic drift relative to others in Europe. While some studies have reported increases in local distributions, successive national surveys have indicated that its range in Britain has declined (Bowdrey, 1997; Pratt 2001). Analysis of ancient DNA samples (e.g. Brace et al 2015) would permit an assessment of historical versus recent genetic loss. The microsatellites developed from the NGS library are extremely short

(~110 base pairs) and may be especially suitable to analysis of low quality ancient DNA templates.

Notwithstanding the underlying causes, genetic diversity in UK populations was extremely low. There is considerable evidence that reduced genetic diversity can increase long-term extinction risk (Frankham 2005, 2010) even when reduced fitness is not immediately apparent (Johnson et al. 2009). The low level of genetic diversity among UK samples suggests that concern over evolutionary potential and population persistence may be warranted. Neutral diversity may poorly reflect adaptive diversity, and for loci like microsatellites with high mutation rates, neutral diversity may overestimate genomic adaptive diversity (Moss et al. 2003; Vali et al. 2008). If adaptive diversity in UK stag beetle is low concern over long term adaptability and consideration of translocations to increase genetic diversity in particularly depauperate populations may be warranted. Translocations can help rehabilitate populations but are notoriously difficult (Tallmon et al. 2004; Frankham 2005). To minimise outbreeding depression efforts are typically made to ensure that stocked individuals are taken from genetically similar populations, although Frankham et al. (2011) argue that concerns over outbreeding depression may be less severe where populations have been recently fragmented. For stag beetle, there are substantial ecological differences between UK and continental European populations (Harvey et al. 2011). British populations are typically associated with more urban areas whereas European populations are more prevalent in rural areas (Harvey et al. 2011). There is also considerable variation in larval duration (100% variation in the number of instars across Europe) and adult size (beetles from Spain, Germany and the Netherlands are larger than those from the UK) (Harvey et al. 2011). While such differences may be largely due to phenotypic plasticity, the potential for underlying adaptive divergences must be considered in potential stocking plans to ensure that locally adapted gene complexes are not disrupted.

The primary goals of conservation genetics are to quantify genetic connectivity among, and diversity within, populations, consider the consequences for population viability and apply appropriate conservation actions (Frankham 2010). Though based on limited sampling this study indicates that while a single beetle clade extends over Europe, British populations must be monitored and managed separately from their continental counterparts. This readily aligns with current geopolitically delineated conservation strategies (Harvey et al. 2011; Campanaro et al. 2016). An essential step now is to perform more extensive population genetic analysis to understand patterns of connectivity/isolation, with the potential Wahlund effect for the

Spanish sample pointing to structuring occurring on local scales. Such studies will require analysis of more precisely georeferenced individuals and a greater number of microsatellite loci and must be performed before any translocations are attempted, and alongside habitat restoration / management schemes. Improved habitat quality might facilitate greater connectivity and expansion of populations (e.g. Johnstone et al. 2011). The RAD-seq library developed here represents a considerable resource for the development of a large number of microsatellite markers suitable for fine-scale genetic monitoring of stag beetle populations.

Table 2.1.Stag beetle microsatellite primers used in the study

Locus name	Forward Primer	Reverse Primer
locus lc-1	5'-TGCAGATTATGAACACGTG-3'	5'-TGTGAAAGAGCCAAGATACACG-3'
locus lc-2	5'-TGCAGTTTCATTTATAAATGTG-3'	5'-AACCACCGTCGTGCAGTTAG-3'
locus lc-3	5'-TGCAGCTTTTTATTATTTCTTGC-3'	5'-TGTCGCCTGAAAATAACTTGTC-3'
locus lc-4	5'-TGCAGTCTAATCTGAATTGAG-3'	5'-TGGTCCCATTCGAACCAC-3'
locus lc-5	5'- TGCAGGGATGCAAAAACG -3'	5'- GGGTTGTTCTCCACCGTACC-3'

Table 2.2 Molecular diversity for UK (n=48) and Spanish (n=52) stag beetles as assessed by microsatellites. Significant deviations from Hardy-Weinberg equilibrium are marked in bold.

Locus		UK	Spain
LC-1	Number of alleles	3	4
	Allelic richness	2.87	4
	Observed Heterozygosity	0.09	0.20
	Expected Heterozygosity	0.08	0.43
	Probability of being in Hardy-Weinberg equilibrium	1	**< 0.0001**
LC-2	Number of alleles	1	8
	Allelic richness	1	7.72
	Observed Heterozygosity	0.00	0.53
	Expected Heterozygosity	0.00	0.58
	Probability of being in Hardy-Weinberg equilibrium	-	**< 0.01**
LC-3	Number of alleles	1	7
	Allelic richness	1	6.83
	Observed Heterozygosity	0.00	0.36
	Expected Heterozygosity	0.00	0.46
	Probability of being in Hardy-Weinberg equilibrium	-	**<0 .01**
LC-4	Number of alleles	2	4
	Allelic richness	1.8	4
	Observed Heterozygosity	0.12	0.33
	Expected Heterozygosity	0.20	0.56
	Probability of being in Hardy-Weinberg equilibrium	**0.04**	**< 0.0001**
LC-5	Number of alleles	2	10
	Allelic richness	1.99	9.67
	Observed Heterozygosity	0.06	0.33
	Expected Heterozygosity	0.06	0.57
	Probability of being in Hardy-Weinberg equilibrium	1	**< 0.0001**

Table 2.3 Distribution of COI haplotypes across Europe in stag beetle. 'N' represents the total sample size per country.

	N	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17	H18	H19	H20	H21	H22	H23	H24	H25	H26
Britain	48	45	1	1	1																						
Spain	52	34									2	1	2	1	2	1	1	2	1	1	1	2					1
Portugal*	1	1																									
France*	20	15				1	1	1															1	1			
Belgium*	2	1		1																							
Germany	2	2																									
Switzerland*	1	1																									
Italy*	3	2							1																		
Romania*	3									2																1	
Hungary*	1								1																		
Greece*	3	2																							1		
Ukraine*	1	1																									

* includes sequences from Cox et al. 2013

Figure 2.1 STRUCTURE output (K=2) for UK and Spanish stag beetles

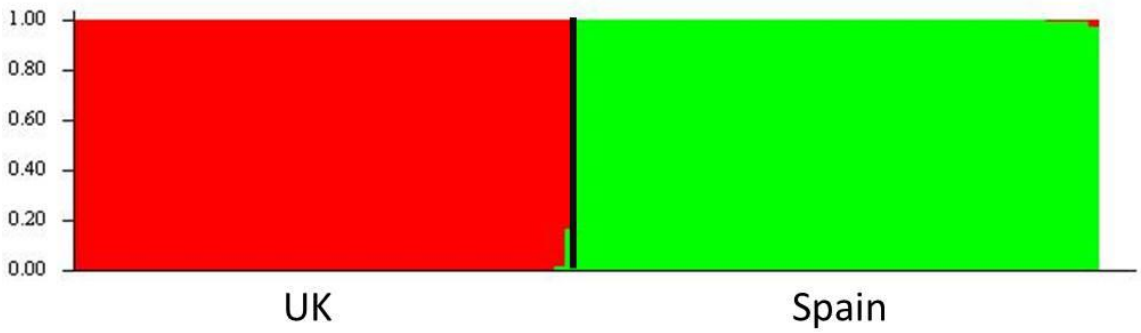
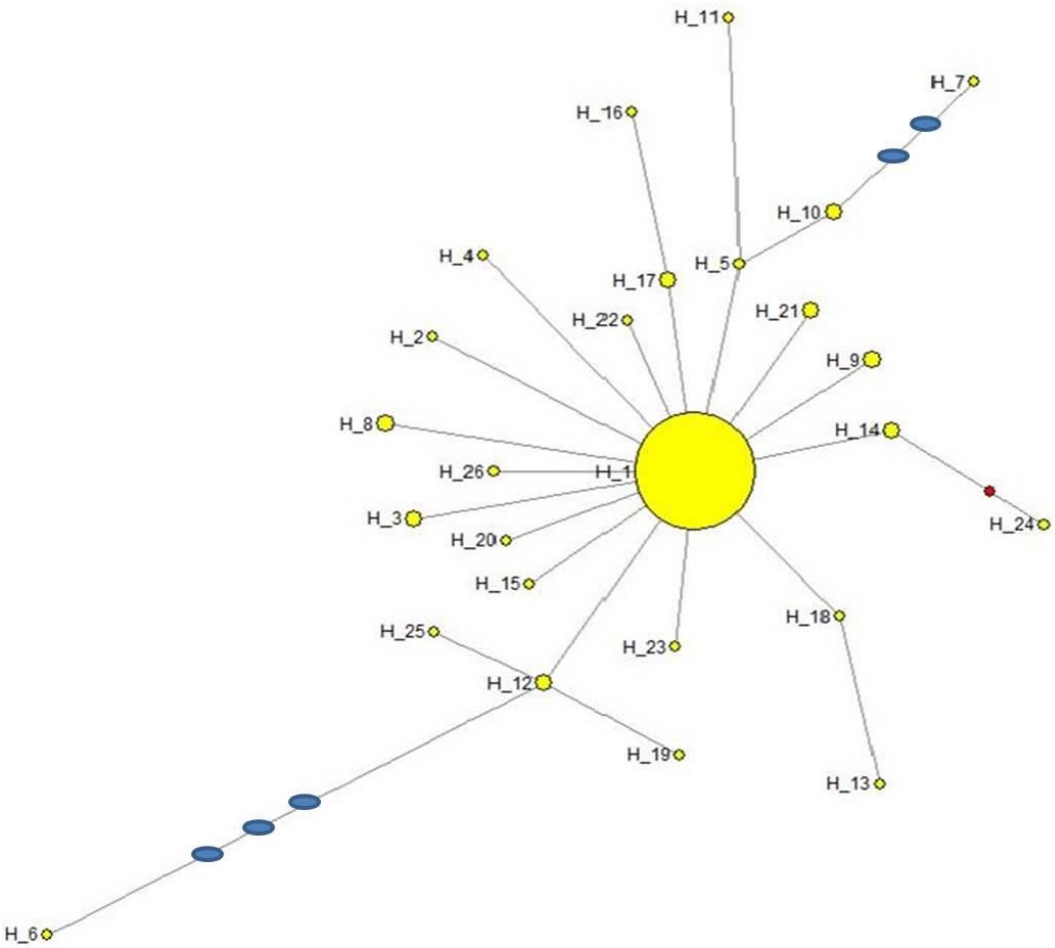


Figure 2.2. Haplotype network for stag beetle COI. All branch lengths represent one mutation, but the blue ellipses represent one mutation each



Chapter 3

Trichius phylogeography and taxonomy

3.1 Introduction

The advent of DNA based taxonomy has in many cases contradicted traditional views based on morphological traits (Sites & Marshall 2003; Audisio et al. 2009; Hill et al. 2015). Studies comparing traditional taxonomies with DNA-based inferences are needed as accurate species delimitation is fundamental to research in biogeography, ecology, macroevolution and conservation biology (Wiens & Penkrot 2002; Raupach et al. 2016). Despite a push toward 'turbo-taxonomy', where new species are described based on genetic barcodes and morphology (Butcher et al. 2012; Summers et al. 2014), species described many years ago may be being left behind in the rush to close the gap on the number of species left to describe (Mora et al. 2011).

Combined analysis of phenotypic and genetic divergence is also a central topic in evolutionary biology (Slatkin 1987). Patterns of morphological and genetic diversity reflect varying influences of historical and contemporary processes. The influence of historical climatic events has been investigated through numerous phylogeographic studies, which have permitted identification of glacial refugia and recolonization dynamics (Taberlet et al. 1998). A number of studies have reported congruent patterns of morphological and genetic variation reflecting postglacial history, and by extension point to the putative influences of selection and stochastic events (Huang & Lin 2010; Ahrens et al. 2013). Studies reporting discordant patterns of morphological and genetic variation suggest a more prominent role for selection and/or short term plasticity in shaping phenotypic variation (Babik et al. 2005; Meraner et al. 2008; Toews & Brelsford 2012). Therefore, an integrative approach can provide insight into the eco-evolutionary processes shaping biodiversity.

The bee beetles (*Trichius*; Scarabaeoidea, Cetoniinae) are a small genus found throughout the Palaearctic. Following the taxonomic clarifications by Krell (2010; 2012) three species are described throughout Europe west of the Caucasus, *T. fasciatus, T. gallicus gallicus* and *T. sexualis*. The three species are generally identified based on the pattern of white stripes on the male sternites (Baraud 1992; Krell 2012). There is a general consensus that the three species differ slightly in distribution (Fig. 1a, b, c), but aren't known to inhabit significantly different environments or host tree preferences (Alexander 2002), i.e. there are areas of sympatry. *T. fasciatus* has been identified throughout Russia, Kazakhstan, Kyrgyzstan and Scandinavia, therefore inhabiting the greatest range extent, but all three species overlap in eastern France,

Germany, Italy, Poland, Romania and the Balkans (Fig. 1a) A primary objective of this research was to combine phenotypic and genetic markers to jointly assess both the genetic validity of the three species and their facility to be identified using phenotypic traits. Genetic variation was assessed using a standard DNA barcoding approach based on sequencing a portion of the COI gene. Based on limitations of this approach mtDNA genotyping was complemented by sequencing a portion of the nuclear *Wingless* gene. Phenotypic variation was then overlaid onto genetic phylogenies to provide a holistic appraisal of these taxa in both a taxonomic and evolutionary framework.

Due to the potential influence of climate change on beetle biodiversity, a secondary objective was to interpret genetic patterns in a phylogeographic context to assess the role of historical climate change. *Trichius* are breeders in heartwood-rot, specialising in utilising tree-birches (*Betula* spp.; Jessop 1986, Koch 1989, Alexander 2002, Mannerkoski et al. 2010c) as hosts. Other species occasionally used include aspen (*Populus tremula*), beech (*Fagus sylvatica*), and oak (*Quercus* spp.). This reliance on discrete habitats allows for predictions to be made regarding the phylogeographic patterns exhibited by the genus. As the most widespread species (Fig. 1a), appearing to inhabit more northern regions than the other species, *T. fasciatus* would be expected to have closely followed the distribution of birch heartwood-rot throughout history. In contrast to other northern European tree species, which spread from southern glacial refugia as the climate warmed after the last glacial maximum (LGM; Hewitt 1997; Provan & Bennett 2008), southern populations of tree-*Betula* did not colonise northwards to a large extent, northern areas of the present distribution being instead re-populated from existing northern refugia (Palmé *et al*. 2003), with genetic boundaries showing a strong east-west split. Therefore, *T. fasciatus* might not conform to a standard "southern richness, northern purity" model of genetic diversity (Hewitt 1996), but instead some pre-existing elements of species diversity may be found in the genome following form the effects of living through the LGM in widespread northern refugia alongside the cold-tolerant *Betula* (Coope 1998; Svenning et al. 2008; Tzedakis et al. 2013). *Trichius fasciatus* might therefore not show genetic signals of population crashes and expansions, instead tightly tracking the habitat which itself has not shown evidence of large population crashes/expansions consistent with glacial cycles. The other two *Trichius* species (Fig. 1b & c), with more southerly ranges, might not have inhabited northern refugia, instead conforming to standard post-glacial recolonisation patterns (Taberlet et al. 1998; Hewitt 2000).

3.2 Methods

3.2.1 Sample Collection

Specimens identified as belonging to all three European *Trichius* species were obtained from private beetle collectors. These specimens were used in the genetic and morphological analyses (n=47). A further six specimens (one male and female from each species) were photographed from the Natural History Museum (London, UK) and used for morphometric analysis (Supplementary Table 1), but were not used in the genetic analysis. As these individuals were identified by experts at the NHM these specimens were assumed to represent standard 'phenotype models' of the three European *Trichius* species.

3.2.2 DNA extraction

DNA was extracted from all samples using a modified phenol-chloroform-isoamyl alcohol protocol (Winnepenninckx et al. 1993). Tissue was initially taken from a meta- or mesothoracic leg from preserved adults up to the trochanter. For specimens which yielded very little DNA via this method, a separate leg was used up to and including the coxa, or by cutting into the thorax and using preserved flight muscles instead. The tissue was washed in distilled water, dried by blotting on lab roll, and placed into a 1.5ml Eppendorf tube. 350µl of CTAB buffer and 10µl of 5µg/ml Proteinase K was added, followed by 10s on a vortex. This was incubated overnight (18-22 hours) at 37°C. 350µl of equilibrated 6.7/8.0pH Phenol-Chloroform-Isoamylalcohol was added into each tube before being shaken by hand for 10 minutes. The tubes were then centrifuged for 10 minutes at 13,000RPM, after which the top layer was pipetted into a new tube, and the waste discarded. 990µl of 100% Ethanol was added to the new tube, and the solution was incubated at -20°C for at least two hours. This was then centrifuged at 13,000RPM for 10 minutes; the liquid solution poured off, with small remaining volumes removed using a pipette. DNA pellets were then dried in open tubes in a fume cupboard for 20 minutes, after which 50µl of distilled water was added. DNA was then left overnight at 4°C to go into solution before assessment by running 5µl DNA stock on 1% agarose gels.

3.2.3 PCR & Sequencing Protocols

Depending on the concentration of DNA visible on a gel, stock DNA was diluted to between 1/10 and 1/100 for mitochondrial PCRs. Most samples were diluted to either 1/20 or 1/40. Genus-specific primers (Table 3.1) were designed to amplify a region of the COI gene (approximately 520bp) from an alignment of sequences from all three *Trichius* species on GenBank. These primers were then used for both PCR and sequencing. PCRs were performed in 20µl volumes consisting of 10µl of Biomix (Bioline, London, UK), 1µl of each primer at 10µM, 5µl of ddH_2O, and 3µl of diluted DNA. The standard PCR thermoprofile was 95°C/3 minutes, 55X (95°C/30s, 55°C/45s, 72°C/45s), 72°C/3 minutes. For specimens that didn't work first time the annealing temperature was reduced to 52°C.

Genus-specific primers for *Wingless* were developed (Table 3.1) from sequences initially generated from using the LepWg1a [5'-GARTGYAARTGYCAYGGYATGTCTGG-3'] and LepWg2a [5'-ACTICGCARCACCARTGGAATGTRCA-3'] conserved primers designed by Brower & De Salle (1998). Genus-specific primers designed to amplify a 200bp fragment spanning the most variable regions among the two initial sequences were developed using Primer3 (Koressaar & Remm 2007) to permit more consistent amplification, as initial tests with primers designed to amplify 450bp struggled to amplify *Wingless* in the poorer quality samples (regardless of PCR conditions). Genotyping PCRs were performed in 20µl columns consisting of 10µl of Biomix (Bioline, London, UK), 1µl of each primer at 10µM, 5µl of ddH_2O, and 3µl of diluted DNA. PCR conditions were 95°C/3 minutes, 55X (95°C/30s, 55°C/45s, 72°C/45s), 72°C/3 minutes.

Amplicons for both genes were checked on a 2% agarose gel, then cleaned with SureClean Plus (Bioline) following the manufacturer's protocol (but increasing the initial centrifugation step to 20 minutes), and sequenced with AB BigDye technology in both directions for *Wingless*, and in a single direction (forward) for COI.

Sequences for both genes were then checked and edited in Chromas Lite (Version 2.1; 2012; Technelysium Pty Ltd). Additionally, for *Wingless*, mixed peaks were edited to include degenerate base codes if these were present. Mixed peaks are base positions in the sequence showing two clear peaks representing different nucleotides, presumed to result from a heterozygous base position. To be classed as a mixed peak, points of inflection in both peaks had to match exactly, whilst the smaller peak could be no less than one third of the size of the

larger (to account for PCR and sequencing biases; most mixed peaks were near equal in size, Fig. 3.2), also, the mixed peaks had to be present in two replicated amplicons for an individual. Both genes were checked for identity using BLAST (Altschul et al. 1990). PHASE was used to construct *Wingless* alleles not assuming recombination due to their short length (Stephens et al. 2001; Stephens & Scheet 2005), and was implemented in DNASP. To test for differences between haplotype confidence probability thresholds (following recommendations from Garrick et al. (2010)), the data were run in separate analyses at 0.6 and 0.95 thresholds (1000 iterations, 1000 burn in).

3.2.4 Statistical analysis of COI sequence data

Sequences were downloaded from GenBank from other *Trichius* species for COI. Recent pushes to barcode European beetles have resulted in a large number of sequences being uploaded to GenBank (Hendrick et al. 2015; Rougerie et al. 2015) for this genus. In addition, three other Trichiini species were used as phylogenetic outgroups: *Trichiotinus assimilis* (KR491060.1), *Gnorimus nobilis* (KM286279.1) and *G. variabilis* (KM285777.1). The 74 sequences were then aligned in BioEdit (Version 7.1.11; 2013; Hall 1999) using the CLUSTAL W algorithm with default settings (Thompson et al. 1994).

The Phylogenetic relationships among sequences were constructed using Maximum Likelihood (built using MEGA6 (6.06; 2013; Tamura et al. 2013)), Neighbour Joining (in MEGA6) and Bayesian Inference (using MrBayes (v3.2.6; Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003)). The most appropriate model for the Maximum Likelihood analysis was chosen using jModelTest (V. 2.1.10; Guindon & Gascuel 2003; Darriba et al. 2012) using the model with the lowest value for the Akaike Information Criterion (AIC) statistic. The General Time Reversible substitution model with a gamma distributed with invariant sites among-site rate (GTR+I+G) was shown to be the most appropriate method (P=80, AIC=3697.05) and was used in the phylogeny with an extensive subtree-pruning-regrafting (SPR) tree inference method, and 1000 bootstrap replications. For Neighbour Joining, the next lowest AIC value model executable in MEGA6 (Tamura 3-parameter) was chosen, (with 1000 bootstrap replicates) and assuming an inverse gamma distribution of site variations, 100,000 generation burn in and 1,000,000 total generations for the Bayesian approach. Additionally a haplotype network was built for the *Trichius* sequences using DNASP (version 5.10, Librado & Rozas 2009) to produce an .rdf file for use in Network

(fluxus-engineering.com), with both Median Joining networks (Bandelt et al. 1999) and Maximum Parsimony networks (Polzin & Daneshmand 2003).

For COI haplotype and allele diversity statistics were calculated in ARLEQUIN (3.5.2.2; 2015; Excoffier & Lischer 2010); number of haplotypes (*H*), haplotype diversity (h), and nucleotide diversity (π) and number of polymorphic sites (*P*) (Nei & Tajima 1981; Nei 1987). The between group variation in haplotypes (corrected for within group variation) was calculated using net nucleotide divergence between groups (Da –Kimura 2 parameter model) and the mean pairwise divergence within groups (uncorrected P distances) were calculated in MEGA.

Fu's *Fs* (Fu & Li 1993; Fu, 1997) and Tajima's *D* (Tajima, 1989) tests were used to test for deviations from mutation-drift equilibrium that could be attributed to selection and/or population size changes. Mismatch distributions (Rogers & Harpending 1992; Harpending, 1994), the frequency distribution of numbers of pairwise differences between haplotypes within a sample, and simulated distributions under a model of demographic expansion, were compared with the sum of squared deviations (SSD) between observed and expected distributions (significance assessed after 10 000 bootstrap replicates) used as a test statistic, and the expansion parameter τ estimated. All statistics were computed using ARLEQUIN (3.5.2.2). COI mutation rates are likely to be between 3.54%My^{-1} (Papadopoulou et al. 2010) and 2.34%My^{-1} (Brower 1994).

3.2.5 Statistical analysis of *Wingless*

For *Wingless*, allele diversity statistics were calculated in ARLEQUIN: number of alleles (*H*), heterozygosity (h), and nucleotide diversity (π) and number of polymorphic sites (*P*). In addition, the number of segregating sites (S), mean number of pairwise differences (*k*), and estimates of nucleotide polymorphism (π and θ) were calculated in DNASP. As a test for selection the number of synonymous substitutions per synonymous sites (d_S) and the number of nonsynonymous per nonsynonymous site (d_N) were calculated in MEGA. The variances of Ds and d_N were computed by bootstrap analysis (1000 replicates) and with this information the null hypotheses of neutral evolution ($d_N = d_S$), positive selection ($d_N > d_S$) and purifying selection ($d_N < d_S$) were tested using z-tests. Additionally nucleotide sequence-based estimates of genetic differentiation (K_{ST} and G_{ST}) for Wingless were calculated in DNASP

(5.10.01; 2010; Librado & Rozas 2009), assessing statistical significance by permutating sequences among samples (with 10000 permutations) for K_{ST}.

3.2.6 Morphological analysis

The aim of this analysis was to assess whether a morphometric technique can recognise and recover three distinct morphological groups of individuals that correspond to their initial species ID (based on supposedly discriminant characters, and comparison to "model" types based on the original species identifications from the Natural History Museum collections) and /or mtDNA haplogroup clade determined by COI sequencing (see above and Results). In effect, whether haplogroup or morphologically distinct clustering can reliably assign individuals back to traditional species descriptions.

Each specimen was photographed with a Canon EOS 7D (Canon Inc., Tokyo, Japan) mounted to a Sigma 105mm f/2.8 DG EX Macro (Sigma Corporation, Kanagawa, Japan) lens.

From the literature (Baraud 1992, Krell 2010; Krell 2012), five characteristics were chosen for their apparent reliability to differentiate between species regardless of preservation method or specimen age (Table 3.2). One sex-specific character was chosen for each sex. These were then checked against photographs of the specimens to check for their reliability and ease of use, and used alongside the NHM "models" to confirm specimen species ID. The identification cues were then sorted into a points-based system (Table 3.3) which was used to score every individual. The characters were:

1. The presence and extent of white bands on the male sternites
2. The presence and size of the mesotibial tooth, which varied from a total lack of a tooth, to a small bump on the mesotibia, to a tooth which comes to a distinct point
3. The extent of a central black band which stretches from the posterior elytra toward the scutellum
4. The extent of the black spot over the elytral 'shoulders', from either total absence or a single small spot, to a thick black band reaching the scutellum
5. The level of indentation on the female pygidium, from a convex pygidium to one with two clear rounded indentations

In addition, basic measurements (using a pair of digital callipers) of length and width were used to infer body shape: width across the eyes, the pronotum and the elytra at the widest points for each, along with the length of the left elytra as a proxy for body size which isn't affected by the position in which the beetle has been mounted/killed. Width measurements were then recalculated as a ratio to elytral length to provide measurements of shape regardless of absolute size.

Different tests were then completed on the dataset:

- Initially, all males and females were grouped together for the analysis without including any sex-specific characteristics. The data were tested for normality using the Shapiro-Wilk test. As the data were shown to be a mixture of normal and non-normal data, Principle Components Analysis was used as the most appropriate way to reduce the complexity of the dataset. The three principle components explaining the greatest amount of variation within the dataset were taken forward for the rest of the analysis. A discriminant analysis was then completed on the principle components, inputting all individuals with their COI haplotype.
- Next, the male dataset alone was used (including male-specific external sexual characteristics) with Principle Components recalculated, and the three components explaining the greatest variation within the dataset taken forward for the rest of the analysis. Two discriminant analyses were then run, one identifying the individuals by their haplogroup, and the other by their morphogroup (original species ID)
- Next, the male analysis was repeated exactly as above, but without including male-specific characteristics in the dataset.
- Next the female dataset was used (including sex-specific characteristics), and the same methodology repeated as for the males analysis.
- Finally, the female analysis was repeated, but without including female-specific characteristics.

Thus in total, nine discriminant analyses were run, with paired tests on the different sexes and with or without sex-specific characteristics to see which grouping (morphological ID or haplogroup ID) performed best in clustering individuals according to their input species ID. Assessments of the discriminant analyses were performed using a leave-one-out cross-validation method to check the classification results. All statistics were computed in SPSS v. 22 (IBM Corp. 2013).

3.3 Results

3.3.1 mtDNA

COI sequences were clear of double peaks, indels, frameshifts and stop codons (in the first reading frame sequenced), and once trimmed for quality resulted in an alignment 461bp long. 22 haplotypes were recovered from 44 samples which sequenced successfully, which rose to 32 haplotypes once GenBank samples were included (Fig. 3.3).

The network (Fig.3.5, 3.6) and all three tree-building methods (Figures 3.7, 3.8, 3.9) showed the same three strongly supported reciprocally monophyletic haplogroup clades in COI, corresponding to *Trichius fasciatus* (GenBank *T. fasciatus* sequences, plus novel individuals from this study), *Trichius sexualis* (GenBank *T. sexualis* sequences, plus novel sequences from this study), and *Trichius gallicus* (GenBank *T. gallicus* sequences). No individuals morphologically identified in this study as being *Trichius gallicus* had COI haplotypes clustering with the three *Trichius gallicus* sequences ((incorrectly) listed in Genbank as "*Trichius zonatus*", a synonym for *Trichius gallicus* (Krell 2012)), instead clustering with either the *Trichius fasciatus* or *Trichius sexualis* haplogroups (Figs 3.3, 3.4). Molecular diversity indices for this gene are shown in Table 3.4. Due to the complicated relationships between the morphotype and the haplotype in individuals, the following groups were used to calculate diversity indices based on the closest morphological identification compared to the museum voucher specimens and the reference sequences on GenBank:

- The *Trichius fasciatus* haplogroup, which includes 40 individuals labelled as 'F' in Supplementary Table 1. This represented 21 individuals with *T. fasciatus* morphotypes, 11 individuals with *T. sexualis* morphotypes and 8 individuals with *T. gallicus* morphotypes
- The *Trichius sexualis* haplogroup, which includes 9 individuals labelled as 'S' in Supplementary Table 1. This represented 2 individuals with *T. fasciatus* morphotypes, 5 individuals with *T. sexualis* morphotypes and 2 individuals with *T. gallicus* morphotypes
- The *Trichius gallicus* haplogroup, which includes the three sequences on GenBank, but no specimens sequenced here

61 variable sites were identified across all the *Trichius* sequences. Overall haplotype diversity was high at 0.918, as was nucleotide diversity at 0.02972 largely due to the pronounced divergence among haplogroups. Net between-haplogroup distances were 0.0668 between *T. fasciatus* and *T. sexualis*, 0.0802 between *T. fasciatus* and *T. gallicus*, and 0.0589 between *T. sexualis* and *T. gallicus*. Mean pairwise divergences within groups were 0.0053 for *T. fasciatus*, 0.0068 for *T. sexualis* and 0.0044 for *T. gallicus*.

Haplotype diversity was higher for the *Trichius fasciatus* haplogroup than for *Trichius sexualis* whether GenBank sequences were included (0.885 and 0.5619 respectively; Table 3.4) or not (0.876 and 0.7143). Nucleotide diversity however was lower in the *T. fasciatus* haplogroup than in *T. sexualis* also regardless of whether GenBank samples were included (0.0051 and 0.0065 respectively) or not (0.0054 and 0.0068). Significant levels of genetic differentiation (F_{ST}) were found between all three haplogroups (Table 3.5), but not between morphogroups.

The *Trichius fasciatus* haplogroup reported significant deviation from neutral expectations for both Tajima's *D* and Fu's *F* in COI, regardless of whether GenBank samples were included or not (Table 3.6). Significant negative values of Tajima's *D* (an excess of low frequency polymorphisms relative to expectation) are indicative of either a recent selective sweep, or a population expansion after a recent bottleneck, as does the similar Fu's *F* statistic (significant negative values from allele excess indicate genetic hitchhiking or recent population expansion). The mismatch distribution also supported a population expansion (SSD=0.0024, p=0.290, Table 3.6; Fig. 3.12 a & c) as did the raggedness statistic (Table 3.6). Based on the corresponding Tau values, and assuming a range of mutation rates in the equation $\tau=2ut$ ($u=2\mu k$, where μ=mutation rate, k=sequence length, t=time of expansion), population expansion time was estimated at between 73,000 and 121,000 years ago (Table 21), depending on the mutation rate and if GenBank samples were included. Neither *D* nor *Fs* demonstrated significant deviations from neutrality in either *Trichius sexualis* or *T. gallicus* haplogroups. Estimated times of population expansion for *T. sexualis* are between 235,000 and 429,000 years ago. This lack of recent expansion is also borne out by the Mismatch Analysis (Fig. 3.12 b & d) (not calculated for *T. gallicus* due to the small sample size) which deviated from a population expansion model.

3.3.2 Wingless

Once the sequences were trimmed for quality, the *Wingless* alignment was 149bp long. BLASTn searches for the sequences confirmed them as *Wingless*, grouping with *Wingless* sequences from *Dynastes granti* (Scarabaeoidea; Dynastidae; GenBank ID: gb|KP813412.1), the closest relative to *Trichius* for which sequences from this gene are available. The 0.6 and 0.95 threshold data from PHASE was compared by eye, and as there were no differences in the alleles recovered between the two thresholds, the 0.95 threshold data were used (Table 3.8). 11 alleles were recovered (Table 3.8). Heterozygosity was higher in *T. sexualis* regardless of the species identification method used (0.9394 using COI haplogroup (8 alleles), 0.8478 using morphological ID (9 alleles)) than in either *T. fasciatus* (0.7719 (from 9 alleles) and 0.7429 (from 6 alleles) respectively) or *T. gallicus* (0.7532 from morphology (8 alleles), no COI sequences representing putative *T. gallicus* were identified from novel samples).

There is some consensus between COI mitochondrial identification and *Wingless* identification, significant genetic differentiation (K_{ST}=0.052, p=0.003; G_{ST}=0.034) being found between the two haplogroups (Fig. 3.10), the common *Wingless* allele H.1 (n=29) being exclusive to the *Trichius fasciatus* COI haplogroup, although there is substantial *Wingless* allele sharing between the COI haplogroups. The same *Wingless* network colour coded for the morphotypes (Fig. 3.11) however shows that there is no obvious differentiation between the three morphotypes, with almost complete allele sharing and the three alleles unique to one morphotype being represented by only one or two copies. However, significant allele frequency differentiation was found between morphological *T. sexualis* and both *T. fasciatus* and *T. gallicus* (K_{ST}=0.0586, p=0.005; G_{ST}=0.024, and K_{ST}=0.0353, p=0.048; G_{ST}=0.042), but not between *T. fasciatus* and *T. gallicus* (K_{ST}=0.009, p=0.805; G_{ST}=0.0049).

The Z-test (using the modified Nei-Gojobori method) of neutral evolution ($d_N = d_S$) was not refuted for either the *T. fasciatus* COI haplogroup *Wingless* sequences (0.0704, P=0.944) or the *T. sexualis* haplogroup (0.550, P=0.583). Positive selection ($d_N > d_S$) was also not detected (0.0721, P=0.471 for *T. fasciatus*, 0.559, P=0.289 for *T. sexualis*), nor was purifying selection ($d_N < d_S$) (-0.074, P=1.00 for *T. fasciatus*, -0.568, P=1.00 for *T. sexualis*). Splitting the Wingless dataset by morphological identifications also showed that the morphogroups conformed with neutral expectations: *T. fasciatus* (-0.155, P=0.877), *T. sexualis* (0.414, P=0.679) or the *T. gallicus* group (0.110, P=0.913). Positive selection was also not detected

(*T. fasciatus* -0.153, *P*=1.00; *T. sexualis* 0.414, *P*=0.340; *T. gallicus* 0.110, *P*=0.456), nor was purifying selection (*T. fasciatus* 0.160, *P*=0.437; *T. sexualis* -0.430, *P*=1.00; *T. gallicus* -0.114, *P*=1.00).

3.3.3 Morphological Data

For the general *Trichius* analysis using all individuals (Fig. 3.13), the first three Principle Components (PCs) explained 75.77% of the variation. PC1 weighted high values of the presence of a central elytral stripe, a mesotibial tooth, and high elytral width. PC2 weighted presence of an anterior elytral band, high elytral length and relatively narrow eyes. PC3 weighted high elytral lengths, narrow eyes, and a wide pronotum (Table 3.10, Supplementary Table 2). As only *T. fasciatus*-like and *T. sexualis*-like COI haplotypes were recovered from the individuals sequenced, only these two haplogroups were used to compare between haplogroup and morphogroup congruence. The discriminant analysis coded to the COI haplogroup of the individuals showed that only 51% of individuals were correctly classified under the leave-one-out cross-validation, with *T. fasciatus* haplotypes being correctly classified (i.e. matching to individuals initially identified as *T. fasciatus*) 52.5% of the time, and a *T. sexualis* correct classification rate of 44.4% (Table 3.11).

Using only the male data and including sex-specific characters, the first three PCs explained 76.4% of the variation (Supplementary Table 3). PC1 weighted presence of an anterior elytral band, a mesotibial tooth, wide elytra and lack of sternal stripes. PC2 weighted a lack of an anterior elytral band, short elytra, wide eyes and a wide pronotum. PC3 weighted presence of a central stripe, long elytra, wide pronota and presence of sternite stripes (Table 3.12). Using the haplogroup to identify the species lead to 58.3% of the classifications being correct, whilst using morphogroup 87.2% of classifications were correct (Tables 3.13 & 3.14).

Re-running the male only analysis but removing the male-specific characteristic, the first three PCs explained 77.8% of the variation (Supplementary Table 4). PC1 weighted presence of a mesotibial tooth, wide eyes, wide pronota and wide elytra. PC2 weighted presence of an anterior elytral band, presence of a central stripe, long elytra, and narrow pronota. PC3 weighted presence of long elytra, narrow eyes, wide pronota and wide elytra (Table 3.15). Using the haplogroup to identify the species lead to 59.2% of the classifications being correct, whilst using morphogroup 64.1% of classifications were correct (Tables 3.16 & 3.17).

Using only the female data including the sex-specific character, the first three PCs explained 77.0% of the variation (Supplementary Table 5). PC1 weighted presence of a central stripe, a mesotibial tooth, wide elytra and lack of the sternal notch. PC2 weighted absence of the elytral band, short elytra, wide eyes and wide pronota. PC3 weighted presence of an elytral band, wide pronota, narrow elytra and presence of the sternal notch (Table 3.18). Due to being unable to separate *Trichius fasciatus* and *T. gallicus* using the sternal notch, the classification grouping is between a *T. fasciatus/T. gallicus* morphogroup and a *T. sexualis* morphogroup. Using the haplogroup to identify the species lead to 53.8% of the classifications being correct, whilst using morphogroup 93.8% of classifications were correct (Tables 3.19 & 3.20).

Using the female data without including the sex-specific character, the first three PCs explained 78.0% of the variation (Table 3.21). PC1 weighted presence of an elytral band, a central stripe, a mesotibial tooth, and wide elytra. PC2 weighted absence of the elytral band, short elytra, wide eyes and wide pronota. PC3 weighted presence of an elytral band, absence of a central stripe, wide eyes and narrow pronota (Table 21). Using the haplogroup to identify the species lead to 44.9% of the classifications being correct, whilst using morphogroup 75.0% of classifications were correct (Tables 3.22 & 3.23).

3.4 Discussion

Mitochondrial DNA sequence results showed European *Trichius* genetic diversity falls into three distinct clades, which might be expected to correspond to the three described species. However, the *Trichius fasciatus* haplogroup (40 individuals labelled as 'F' in Table 1), designated as such by presence of *T. fasciatus* sequences from GenBank, includes 21 individuals identified morphologically as *T. fasciatus* but also 11 individuals with *T. sexualis* morphotypes and 8 individuals with *T. gallicus* morphotypes. Likewise the *Trichius sexualis* haplogroup (9 individuals labelled 'S' in Supplementary Table 1) included 5 individuals with *T. sexualis* morphotypes but also 2 individuals with *T. fasciatus* morphotypes and 2 individuals with *T. gallicus* morphotypes. The *Trichius gallicus* haplogroup was represented only by the three *T. gallicus* sequences on GenBank with individuals morphologically identified as *Trichius gallicus* only posessing COI sequences clearly grouping with either *T. fasciatus* or *T. sexualis* (Supplementary Table 1, Fig. 3.6). Therefore, whilst the

morphological analysis conducted here resolved three distinct morphotypes there was a high level of morphological-genetic incongruence. Genetic and morphological incongruence is well described for a number of terrestrial arthropod taxa (e.g. moths (Hundsdoerfer & Wink 2006; Hundsdoerfer et al. 2011), centipedes (Giribet & Edgecombe 2006), spiders (Miller & Hormiga 2004) and flies (Kopp & True 2002)), but it has not been reported previously in scarab beetles.

Two alternative hypotheses could explain the mtDNA/morphological incongruence. Firstly, morphological variation could occur as a result of an adaptive or plastic response to environmental variation. Environmental variation has been shown to drive colour and spine changes in other invertebrates (Davis ret al. 2005; Westphal et al. 2014). However, some samples collected at the same time and place (i.e. sympatric) also were identified as two different species (Figs. 3.3 & 3.4): males from the 'TK' samples (Kimry, Russia) were identified as both *T. fasciatus* and *T. gallicus*, but only possessed *T. fasciatus* haplotypes, whilst the 'CG' (Croatian) samples were all morphologically identified as *T. sexualis*, but possessed seven COI *T. fasciatus* haplotypes and four *T. sexualis* haplotypes. The occurrence of morphological divergence in sympatry suggests that environmental factors may not be the prominent drivers of the observed incongruence.

On the other hand the patterns could be attributed to introgression. This hypothesis entails that populations (clades) have diverged allopatrically into morphospecies without attaining complete reproductive isolation with introgression occurring upon secondary contact. In this case the clades may have diverged in allopatric glacial refugia with secondary contact occurring during interglacial periods. Of the five statistics used to test for signals of past population expansions after bottlenecks, *D*, *Fs*, SSD, Raggedness and Mismatch all supported a hypothesis of recent population expansion in the *Trichius fasciatus* haplogroup, whilst only SSD and one result from D supported a hypothesis of recent population expansion in the *T. sexualis* haplogroup, with *T. sexualis* displaying a bimodal mismatch distribution indicative of longer term population stability than the unimodal pattern of expansion seen in *T. fasciatus* (Rogers & Harpending 1992; Meraner et al. 2008; Excoffier et al. 2009). Therefore COI supports a recent population expansion in *T. fasciatus*, but not *T. sexualis*. The three GenBank sequences for *T. gallicus* prevent any meaningful statistics to be computed for this species.

In addition, completing the tests on the three different species as identified morphologically shows a significant negative value for *D* in the *T. fasciatus* morphogroups, but no other significant differences from the null hypotheses for either *D* or *Fs* in the other morphogroups. However, Raggedness and SSD did support recent population expansions in all three morphogroups (Table 3.6). These patterns, and estimated times for the *T. fasciatus* haplogroup expansion (Table 3.7) fit with similar reports on other species in the northern hemisphere with population expansions in the Weichsel Early Glacial period (60,000-115,000 years ago), before the Last Glacial Maximum 21,000 years ago: Neumann et al. (2005) on European hamsters; Centeno-Cuadros et al. (2009) on Iberian water voles; Pulgarin-R & Burg (2012) on North American woodpeckers. Isolation and differentiation of the European *Trichius* in separate glacial refuges during previous glaciations, followed by population and range expansion before / during / after the last glacial maximum with consequent secondary contact, may explain the present morphological and genetic diversity observed across Europe.

To directly investigate the potential occurrence of introgression sequencing of the nuclear *Wingless* gene was performed. The gene has been used in a number of studies on insects looking at hybridisation between closely related species (e.g. Brower & DeSalle 1998; Lin et al. 2007; Solano et al. 2016) performing extremely well at reconstructing the phylogenetic history of even relatively recently diverged species pairs (Wild & Maddison 2008). Even with the short *Wingless* sequence length used, between 6 and 9 alleles were recovered within each species depending on the identification method (either morphologically or from COI haplotype, Table 3.8) from a total pool of 11 alleles. Between the two COI haplogroups, significant population subdivision (K_{ST}) was recovered (p=0.003) between the *Trichius fasciatus* and *T. sexualis* haplogroups. There was no subdivision between the *T. fasciatus* and *T. gallicus* morphogroups (p=0.805), whilst *T. sexualis* showed differentiation between both *T. fasciatus* (p=0.005) and *T. gallicus* (p=0.048) morphogroups (Table 3.9). However morphogroups possessed *Wingless* alleles throughout the network (Fig. 3.11) as did COI haplotypes (Figs. 3.4, 3.6), so despite some population subdivision indicated by this gene there was considerable allele sharing between the different clades and morphospecies.

The ‘three times rule’ (Palumbi et al. 2001) predicts that due to the theoretically 4-fold larger effective population size of diploid nuclear genes compared to mtDNA, the majority of nuclear loci should attain monophyly when the interclade/intraclade diversity ratio is larger than 3 for mtDNA. Mean within-group genetic distances for haplogroups including GenBank samples were d=0.0053 within *Trichius fasciatus* and d=0.0068 within *T. sexualis*, whilst the

between group divergence was 0.0668, approaching a 10:1 ratio. This suggests that sufficient evolutionary time has elapsed for the nuclear alleles to be sorted and therefore, the high level of allele sharing is compatible with nuclear introgression. The three common European *Trichius* are sympatric throughout much of their range, and are locally common in suitable habitat (Mannerkoski et al. 2010c, d, Mason et al. 2010). This might imply that the three species have had ample chance to meet and introgress throughout their recent evolutionary history.

Introgressive hybridisation across species boundaries has been reported for several fish species and invertebrates (Bernatchez et al. 1995; Darling 2011; Ladner & Palumbi 2012) and in many cases hybrid individuals may not be morphologically intermediate but rather indistinguishable from one or other of their parent taxa (Allendorf et al. 2001; Harper & Hart 2007). Complete analysis of introgression, and specifically the role of historical vs recurrent introgression, will require the analysis of rapidly evolving nuclear markers such as microsatellites, which have been applied to such studies in a variety of species (e.g. Darling 2011; Harris et al. 2013).

Exactly where this leaves the status of *Trichius gallicus* is unknown:

1. Using morphology to differentiate between *T. gallicus* and *T. fasciatus* is extremely unreliable without using external genital patterns in males (and are difficult to distinguish in females) (Table 3.11), suggesting that the accepted non-genital cues for identification are not reliable.
2. The species is intermediate between *T. fasciatus* and *T. sexualis* in every characteristic studied (Tables 3.1 & 3.2).
3. Morphologically identified specimens of *T. gallicus* show either *T. fasciatus* or *T. sexualis* COI haplotypes.
4. Males from the same population with the same haplogroup possess variable levels of sternite striping, from black (*T. fasciatus*-like) to a single white stripe (*T. gallicus*-like) (Table 3.1).
5. Multiple specimens used in the study represented morphological *T. gallicus* but were found far outside of its accepted distribution (Kyrgyzstan, western Russia, and central Russia) (Supplementary Table 1, Figs. 3.1, 3.3, 3.4).
6. No sequences were recovered from any specimen which grouped with *T. gallicus* COI sequences from GenBank (Figs 3.5, 3.7, 3.8, 3.9).

7. Morphological specimens of *T. gallicus* possessed alleles throughout the *Wingless* network (Fig. 3.11).

Three possible hypotheses could explain these findings.

a) *T. gallicus* is a valid species which has been the subject of a selective sweep caused by a cytoplasmic symbiont in some areas of its range, leading to the species to having both *T. sexualis* and *T. fasciatus* mitochondria in different parts of the range, but also its ancestral haplotype in areas which have not been subject to symbiont infection.
b) *T. gallicus* is an intermediate form / hybrid between *T. sexualis* and *T. fasciatus* haplogroups, facilitating gene flow between the two species.
c) *T. gallicus* is a morphological form of *T. fasciatus*, but *T. fasciatus* (*sensu lato*) has hybridised with *T. sexualis* within recent evolutionary history.

Symbiont-driven selective sweeps in mtDNA reduce mtDNA diversity, producing similar patterns to those caused by population bottlenecks and expansions (Tajima 1989; Hurst & Jiggins 2005). Though there is no evidence for an altered sex ratio in *Trichius* (males and females were approximately evenly used in the study, Supplementary Table 1), the common symbiont *Wolbachia* usually causes cytoplasmic incompatibility in insects, causing zygotes formed from eggs from uninfected females and infected males to die during early development, uninfected individuals thus being selectively killed by the symbiont (Jiggins et al. 2001; Hurst & Jiggins 2005). Sweeps caused by *Wolbachia* are known in Coleoptera (e.g. Noriyuki et al. 2014; Mazur et al. 2016), which could lead to difficulties in gene flow estimation, especially where cross-species hybridisation may have occurred (Johnstone & Hurst 1996). Thus, whilst the possibility of cytoplasmic symbionts in *Trichius* cannot be ruled out, it needs to be specifically tested for. Following from the genetic results strongly suggesting introgression between *Trichius fasciatus* and *T. sexualis*, suggestions b and c may be the most likely.

Despite many publications discussing three species of *Trichius* in Western Europe clearly defined from non-genital morphological identifications (e.g. Baraud 1992; Krell 2010; Mannerkoski et al. 2010c; Mannerkoski et al. 2010d; Mason et al. 2010; Krell 2012), this study has failed to provide strong support for species identifications for this genus without using sex-specific characteristics (Table 3.14 & 3.20). Even with these sex-specific characteristics, both the mitochondrial COI and the nuclear *Wingless* failed to provide strong support for there being three genetically distinct species within the study samples, despite

initial expectations based on samples from GenBank that three reciprocally monophyletic clades existed. No clade-specific COI haplotype was recovered from the *Trichius gallicus* specimens, and both *T. fasciatus* and *T. sexualis* haplotypes (as identified form their representative sequence on GenBank) were found throughout individuals regardless of their morphological identification. Additionally, one sample site possessed individuals with one COI haplogroup, but belonging to two different morphological groups, and another site possessed individuals all belonging to the same morphogroup, but with *T. fasciatus* and *T. sexualis* COI haplotypes.

Table 3.1. Novel DNA primers designed for this study.

Primer name	Gene targeted	Sequence
TrFaCOIF1	COI	5'-TGGTAGATGAGCAGGAATAGT-3'
TrFaCOIR1	COI	5'-TGTTGGTATAAAATTGGATCTCC-3'
TrWG60f	*Wingless*	5'-TGAAGGATAGATTCGACGGC-3'
TrWG259r	*Wingless*	5'-TTTGTGTTCGGGATTGTATGG-3'

Table 3.2: Morphological characters used in this study to identify *Trichius* species

T. fasciatus	*T. sexualis*	*T. gallicus*
No white sternite stripes	Multiple white sternite stripes on males covering four sternites	Single white sternite stripe on males, occasionally up to three stripes
Prominent mesotibial tooth	Lacking tooth	Small mesotibial tooth
Black central band on elytra reaches scutellum	Black band doesn't reach scutellum	Black band may reach scutellum
Black band covers the anterior end of the elytra	Black band restricted, remaining as a single spot if present	Black band restricted, remaining as a single spot if present, but variable
Female pygidium not indented	Female's terminal sternite with two small rounded incisions around the centre	Female terminal edge of sternite regular and convex

Table 3.3. Identification markers and scores used in this study. The shape of the female pygidium is scored from 0 to 1; the other metrics are from 0-2.

0	1	2
Males without sternite stripes	Single white sternite stripe on males	Multiple white sternite stripes on males covering four sternites
Lacking mesotibial tooth	Small mesotibial tooth	Prominent mesotibial tooth
Black central band on elytra doesn't reach scutellum	Black band may reach scutellum, but is thin	Thick black central band on elytra reaches scutellum
Black band at the anterior of the elytra not present, may just be a single spot	Large black spot approaching the scutellum, or broken band	Black band covers the anterior end of the elytra
Female pygidium not indented	Female's terminal sternite with two small rounded incisions around centre	-

Table 3.4: molecular diversity for COI across all three *Trichius* species (461bp). *Trichius gallicus* is erroneously referred to as "Trichius zonatus" [sic] on GenBank. This name is retained here to avoid confusion when referencing the morphological *Trichius gallicus*.

Haplogroup/ Species	GenBank included	Sample size	Number of haplotypes	Haplotype diversity	Nucleotide diversity	Polymorphic sites	k	θ_H	θ_π
Fasciatus - haplogroup	Yes	52	24	0.885	0.0051	27	2.351	6.196	0.0134
Fasciatus - haplogroup	No	36	18	0.876	0.0054	22	2.490	5.546	0.0120
Sexualis – haplogroup	Yes	15	5	0.5619	0.0065	13	3.010	3.998	0.0087
Sexualis – haplogroup	No	7	4	0.7143	0.0068	11	3.143	4.489	0.0097
Zonatus – haplogroup	Yes	3	3	1	0.0043	3	2.000	2.000	0.0043
Fasciatus – morphological	No	17	10	0.838	0.01163	35	5.360	10.649	0.0231
Sexualis - morphological	No	14	11	0.956	0.0314	37	14.462	11.635	0.0252
Gallicus - morphological	No	10	7	0.933	0.2642	34	12.178	12.372	0.0268

Table 3.5: Genetic differentiation (COI sequence F_{ST}) between putative *Trichius* species designations, based on haplogroup or morphological ID. Significant values of F_{ST} shown in bold.

	Fasciatus haplogroup	Fasciatus inc. GenBank	Fasciatus morphological	Sexualis haplogroup	Sexualis inc. GenBank	Sexualis morphological	Gallicus GenBank	Gallicus mophological
Sexualis haplogroup	**0.91496**	**0.91961**	-	0				
Sexualis inc. GenBank	**0.91456**	**0.91898**	-	-	0			
Sexualis morphological	-	-	0.1027	-	-	0		
Gallicus GenBank	**0.93195**	**0.93544**	-	**0.89797**	**0.89674**	-	0	
Gallicus morphological	-	-	0.0042	-	-	-0.04963	-	0

Table 3.6: Tests for COI sequence neutrality and for signals of population expansion for all three *Trichius* species.

Haplogroup/ Species	GenBank included	Tau	Theta	Raggedness	SSD	*D*	*Fs*
Fasciatus - haplogroup	No	2.543	0.0025	0.0426 *P*=0.330	0.0069 *P*=0.320	**-1.850 *P*=0.017**	**-11.66 *P*<.000**
Fasciatus - haplogroup	Yes	2.354	0.0006	0.03554 *P*=0.220	0.0024 *P*=0.290	**-1.987 *P*=0.009**	**-19.71 *P*<0.000**
Sexualis – haplogroup	No	9.233	1.583	0.1769 *P*=0.840	0.0809 *P*=0.370	**-1.623 *P*=0.016**	0.752 *P*=0.637
Sexualis – haplogroup	Yes	7.686	0.6952	0.1565 *P*=0.750	0.0298 *P*=0.680	-0.972 *P*=0.170	1.292 *P*=0.787
Zonatus – haplogroup	Yes	2.281	0.010	0.222 *P*=1.000	0.0423 *P*=0.660	0.000 *P*=1.00	-0.693 *P*=0.119
Fasciatus – morphological	No	2.688	0.0007	0.0834 *P*=0.450	0.0293 *P*=0.340	**-1.981 *P*=0.013**	-1.286 *P*=0.253
Sexualis - morphological	No	26.489	6.3200	0.0601 *P*=0.420	0.0578 *P*=0.090	1.053 *P*=0.875	-0.407 *P*=0.415
Gallicus - morphological	No	26.908	6.5754	0.1417 *P*=0.730	0.0886 *P*=0.100	0.064 *P*=0.529	1.400 *P*=0.730

Table 3.7: estimated time (years) since population expansion using Tau values from Table 21 under different minimum and maximum COI mutation rates

Species	Tau	2.34%My^{-1} Mutation rate	3.54%My^{-1} Mutation rate
Fasciatus – inc. GenBank	2.354	110916	73316
Fasciatus – No GenBank	2.543	121298	80181
Sexualis – inc. GenBank	7.686	356249	235487
Sexualis – No GenBank	9.233	427953	282884
Gallicus – GenBank only	2.595	120279	79507

Table 3.8: molecular diversity and tests for signals of population expansion for *Wingless* across all three *Trichius* species

Species	Identification	Sample size	Number of alleles	Allele diversity	Nucleotide diversity	Polymorphic sites	Tau	Theta	Raggedness	D	Fs
Trichius fasciatus	COI	29	9	0.7719	0.01078	6	0.884	0.804	0.3400	0.589 P=0.758	-1.777 P=0.206
Trichius fasciatus	Morphological	18	6	0.7429	0.00815	3	1.291	0.010	0.0400	1.519 P=0.924	-0.836 P=0.282
Trichius sexualis	COI	6	8	0.9394	0.012	6	1.981	0.010	0.2149	-0.382 P=0.376	-4.463 **P=0.001**
Trichius sexualis	Morphological	12	9	0.8478	0.0145	7	2.591	0.025	0.0520	0.475 P=0.712	-2.443 P=0.075
Trichius gallicus	Morphological	11	8	0.7532	0.0104	5	0.782	0.815	0.0476	0.386 P=0.699	-2.894 **P=0.021**

Table 3.9: measures of genetic differentiation in *Wingless* between groups. Individuals are identified by either their COI haplotype or morphotype.

	Fasciatus, COI		Fasciatus, Morphological		Sexualis, Morphological	
Sexualis, COI	K_{ST} =0.05235 ***p*=0.003**	G_{ST} =0.034	-	-	-	
Sexualis, Morphological	-	-	K_{ST} =0.0586 ***p*=0.005**	G_{ST} =0.0240	-	
Gallicus, Morphological	-	-	K_{ST} =-0.0090 =0.805	G_{ST} =-0.0049	K_{ST} =0.0353 ***p*=0.048**	G_{ST} =0.042

Table 3.10. The weighted proportion of the morphological measurements which make up each PC score from Supplementary Table 2. The most important three variables in each component are outlined in red.

	Component		
	1	2	3
Elytra	.486	.677	-.425
Central	.653	.458	.068
Meso	.775	.187	.184
Left	-.259	.822	.228
Eyes_rel	.478	-.564	-.562
Pronotum_rel	.521	-.545	.518
Top_rel	.785	-.009	.097

Table 3.11. Cross-validated classification results for the discriminant analysis for all the *Trichius*, scored and identified by their haplogroup

		COI identification	*Trichius fasciatus*	*Trichius sexualis*	Total
Cross-validated	Count	*Trichius fasciatus*	21	19	40
		Trichius sexualis	5	4	9
	%	*Trichius fasciatus*	52.5	47.5	100.0
		Trichius sexualis	55.6	44.4	100.0

Table 3.12. The weighted proportion of the morphological measurements which make up each PC score from Supplementary Table 3. The four most important variables in each component are outlined in red.

	Component		
	1	2	3
Elytra	.750	-.554	-.113
Central	.654	-.220	.462
Meso	.757	.122	.010
Left	-.241	-.755	.389
Eyes_rel	.421	.672	-.310
Pronotum_rel	.145	.880	.353
Top_rel	.729	.306	.237
Male	-.709	.491	.318

Table 3.13. Cross-validated classification results for the discriminant analysis for the *Trichius* males, scored and identified by their haplogroup.

COI identification		Output ID / Input ID	*Trichius fasciatus*	*Trichius sexualis*	
Cross-validated	Count	*Trichius fasciatus*	18	11	29
		Trichius sexualis	4	3	7
	%	*Trichius fasciatus*	62.1	37.9	100.0
		Trichius sexualis	57.1	42.9	100.0

Table 3.14. Cross-validated classification results for the discriminant analysis for the *Trichius* males, scored and identified by their morphogroup.

Morphological ID		Output ID / Input ID	*Trichius fasciatus*	*Trichius gallicus*	*Trichius sexualis*	
Cross-validated	Count	*Trichius fasciatus*	17	2	0	19
		Trichius gallicus	1	5	1	7
		Trichius sexualis	0	1	12	13
	%	*Trichius fasciatus*	89.5	10.5	.0	100.0
		Trichius gallicus	14.3	71.4	14.3	100.0
		Trichius sexualis	.0	7.7	92.3	100.0

Table 3.15. The weighted proportion of the morphological measurements which make up each PC score from Supplementary Table 4. The four most important variables in each component are outlined in red.

	Component		
	1	2	3
Elytra	.227	.886	-.197
Central	.386	.671	.041
Meso	.669	.411	-.058
Left	-.649	.507	.218
Eyes_rel	.752	-.249	-.521
Pronotum_rel	.679	-.543	.316
Top_rel	.753	.199	.513

Table 3.16. Cross-validated classification results for the discriminant analysis for the *Trichius* males, without including sexual characters, scored and identified by their haplogroup.

		Output ID / Input ID	*Trichius fasciatus*	*Trichius sexualis*	
Cross-validated	Count	*Trichius fasciatus*	23	17	40
		Trichius sexualis	3	6	9
	%	*Trichius fasciatus*	57.5	42.5	100.0
		Trichius sexualis	33.3	66.7	100.0

.

Table 3.17. Cross-validated classification results for the discriminant analysis for the *Trichius* males, without including sexual characters, scored and identified by their morphogroup.

		Output ID / Input ID	*Trichius fasciatus*	*Trichius gallicus*	*Trichius sexualis*	
Cross-validated	Count	*Trichius fasciatus*	12	5	2	19
		Trichius gallicus	2	3	2	7
		Trichius sexualis	0	3	10	13
	%	*Trichius fasciatus*	63.2	26.3	10.5	100.0
		Trichius gallicus	28.6	42.9	28.6	100.0
		Trichius sexualis	.0	23.1	76.9	100.0

Table 3.18. The weighted proportion of the morphological measurements which make up each PC score from Supplementary Table 5. The four most important variables in each component are outlined in red.

	Component		
	1	2	3
Elytra	.644	-.465	.365
Central	.765	-.087	.010
Meso	.830	.172	.025
Left	.288	-.874	.269
Eyes_rel	.371	.722	.032
Pronotum_rel	.282	.626	.620
Top_rel	.844	.051	-.285
Female	-.860	.015	.334

Table 3.19. Cross-validated classification results for the discriminant analysis for the *Trichius* females, scored and identified by their haplogroup.

		Output ID / Input ID	*Trichius fasciatus*	*Trichius sexualis*	
Cross-validated	Count	*Trichius fasciatus*	6	5	11
		Trichius sexualis	1	1	2
	%	*Trichius fasciatus*	54.5	45.5	100.0
		Trichius sexualis	50.0	50.0	100.0

Table 3.20. Cross-validated classification results for the discriminant analysis for the *Trichius* females, scored and identified by their morphogroup

		Output ID / Input ID	*T. fasciatus / gallicus*	*T. sexualis*	
Cross-validated	Count	*T. fasciatus / gallicus*	10	1	11
		T. sexualis	0	5	5
	%	*T. fasciatus / gallicus*	90.9	9.1	100.0
		T. sexualis	.0	100.0	100.0

Table 3.21. The weighted proportion of the morphological measurements which make up each PC score from Supplementary Table 6. The four most important variables in each component are outlined in red.

	Component		
	1	2	3
Elytra	.642	-.473	.540
Central	.786	-.097	-.386
Meso	.848	.161	-.012
Left	.322	-.880	.121
Eyes_rel	.379	.717	.447
Pronotum_rel	.357	.619	.032
Top_rel	.810	.042	-.312

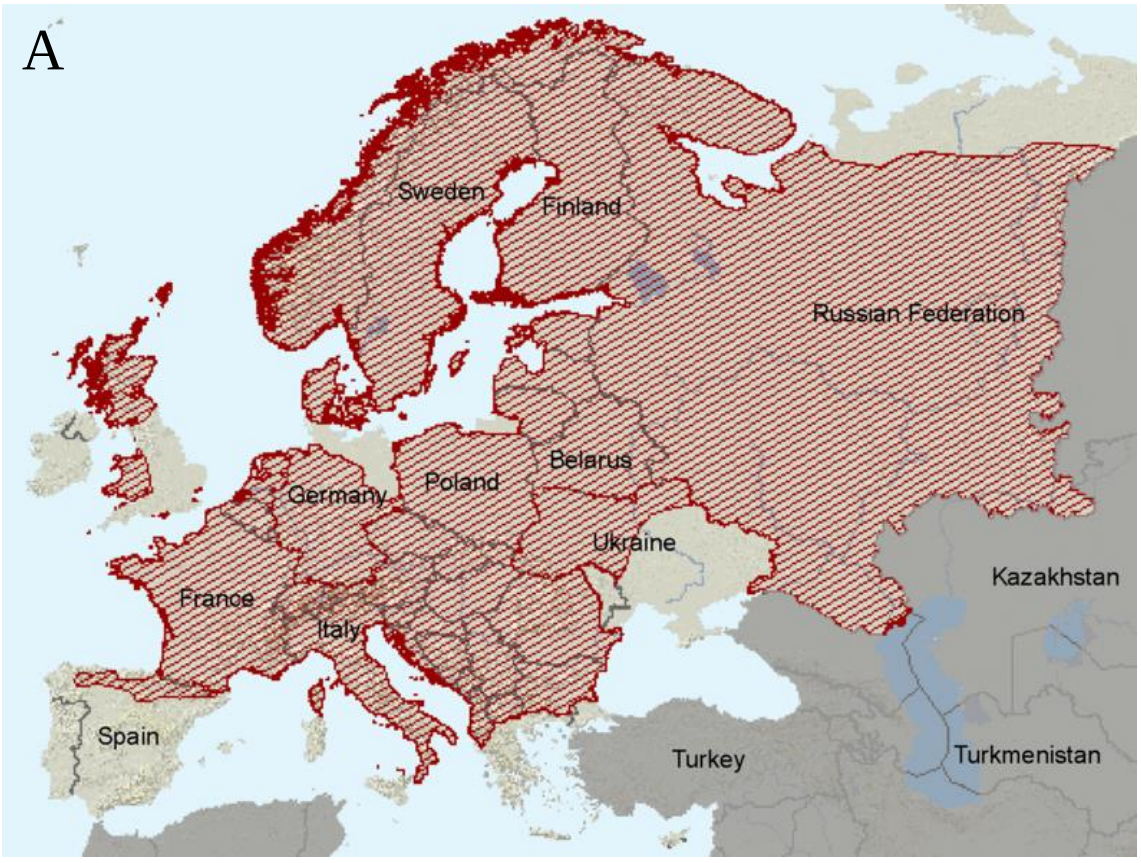
Table 3.22. Cross-validated classification results for the discriminant analysis for the *Trichius* females without including sex-specific characteristics, scored and identified by their haplogroup.

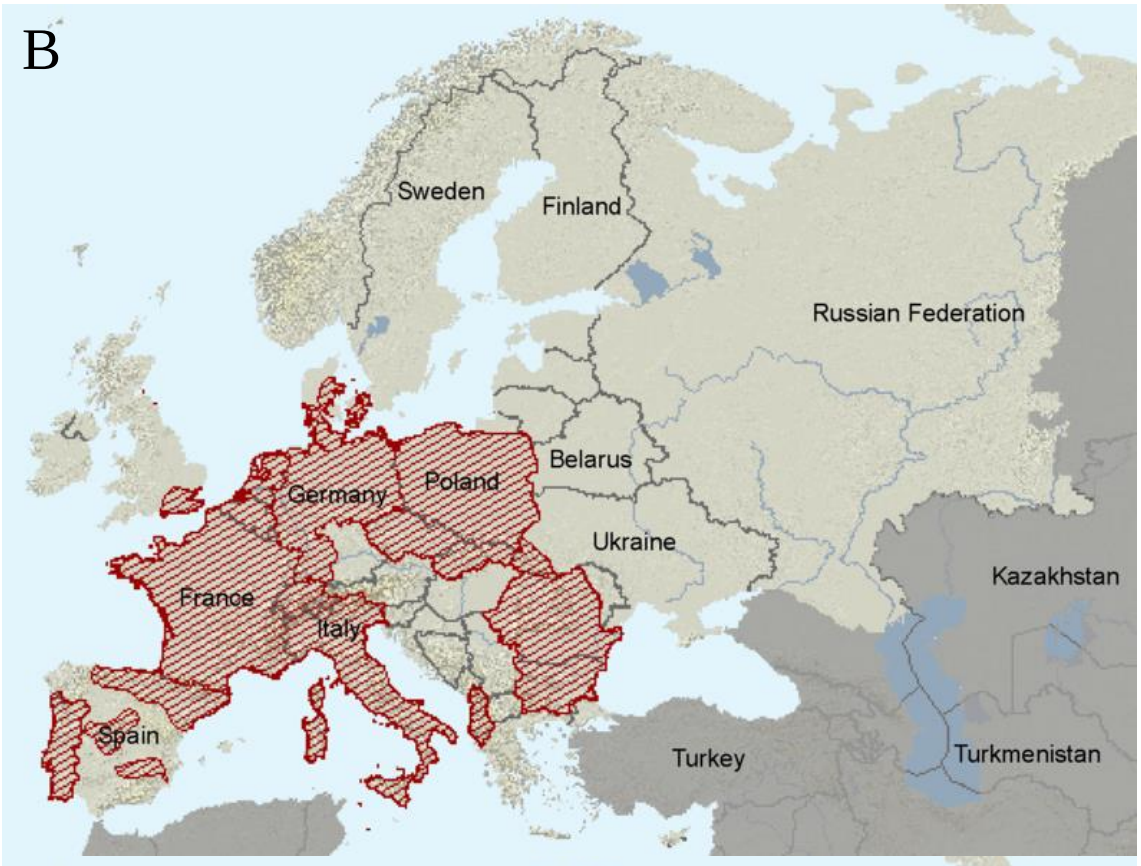
		Output ID / Input ID	*Trichius fasciatus*	*Trichius sexualis*	
Cross-validated	Count	*Trichius fasciatus*	18	22	40
		Trichius sexualis	5	4	9
	%	*Trichius fasciatus*	45.0	55.0	100.0
		Trichius sexualis	55.6	44.4	100.0

Table 3.23. Cross-validated classification results for the discriminant analysis for the *Trichius* females including sex-specific characteristics, scored and identified by their morphogroup.

		Output ID / Input ID	*T. fasciatus / gallicus*	*T. sexualis*	
Cross-validated	Count	*T. fasciatus / gallicus*	8	3	11
		T. sexualis	1	4	5
	%	*T. fasciatus / gallicus*	72.7	27.3	100.0
		T. sexualis	20.0	80.0	100.0

Figure 3.1: distributions (in the red shaded area) of common *Trichius* spp. in Europe from the IUCN Red List Regional Assessments for Europe (Mannerkoski et al. 2010c.,d.; Mason et al. 2010): A) *Trichius fasciatus*; B) *Trichius gallicus* (referred to as '*Trichius zonatus*' in the Red List); C) *Trichius sexualis*. The distribution of *T. fasciatus* extends considerably further into Asia, including to Kazakhstan and lake Baikal in Russia




B
Sweden
Finland
Russian Federation
Belarus
Germany
Poland
Ukraine
France
Kazakhstan
Italy
Spain
Turkey
Turkmenistan


C
Sweden
Finland
Russian Federation
Belarus
Germany
Poland
Ukraine
Kazakhstan
France
Italy
Spain
Turkey
Turkmenistan

Figure 3.2. Example of a C-T mixed peak in *Trichius Wingless*.

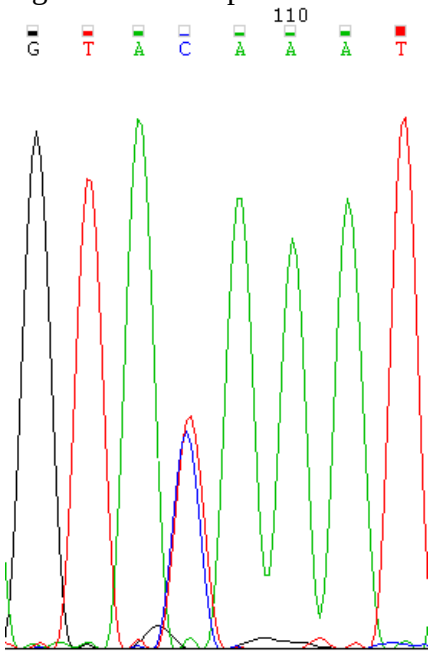


Figure 3.3. Sample map for *Trichius* as identified by their COI haplogroup, and including GenBank samples

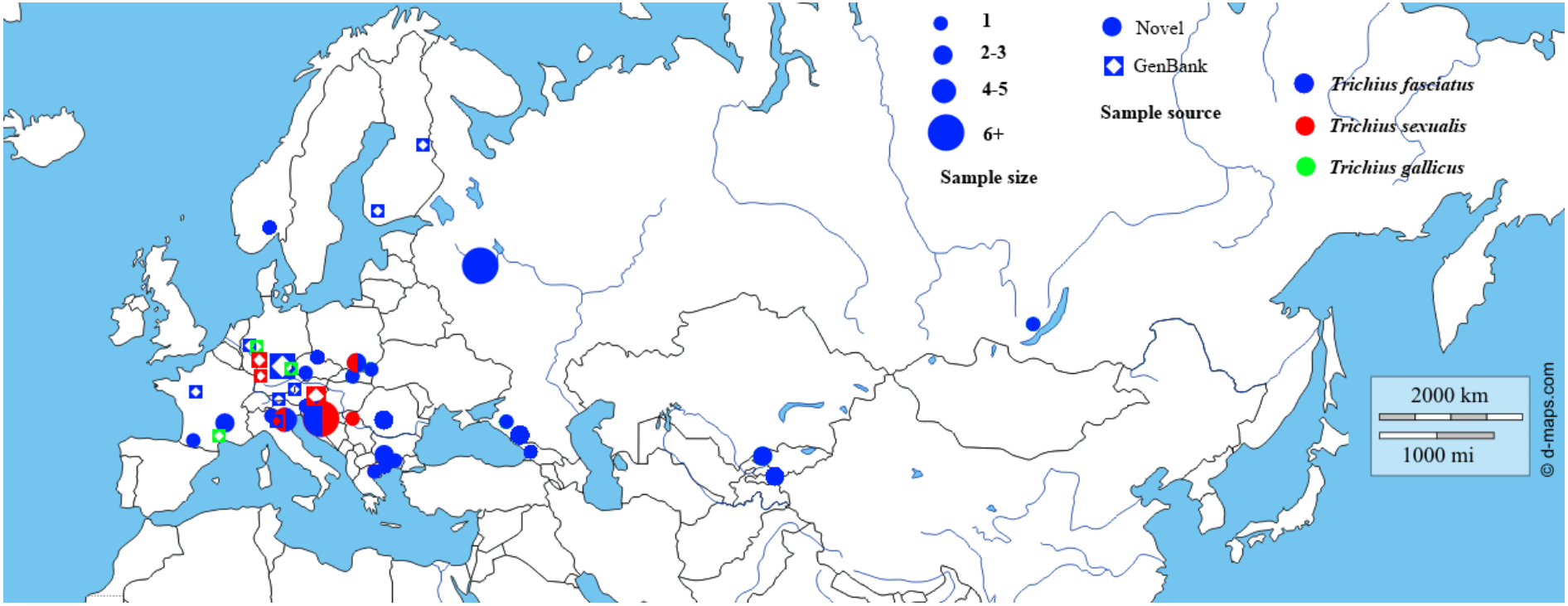


Figure 3.4. Sample map for *Trichius* as identified by their morphogroup

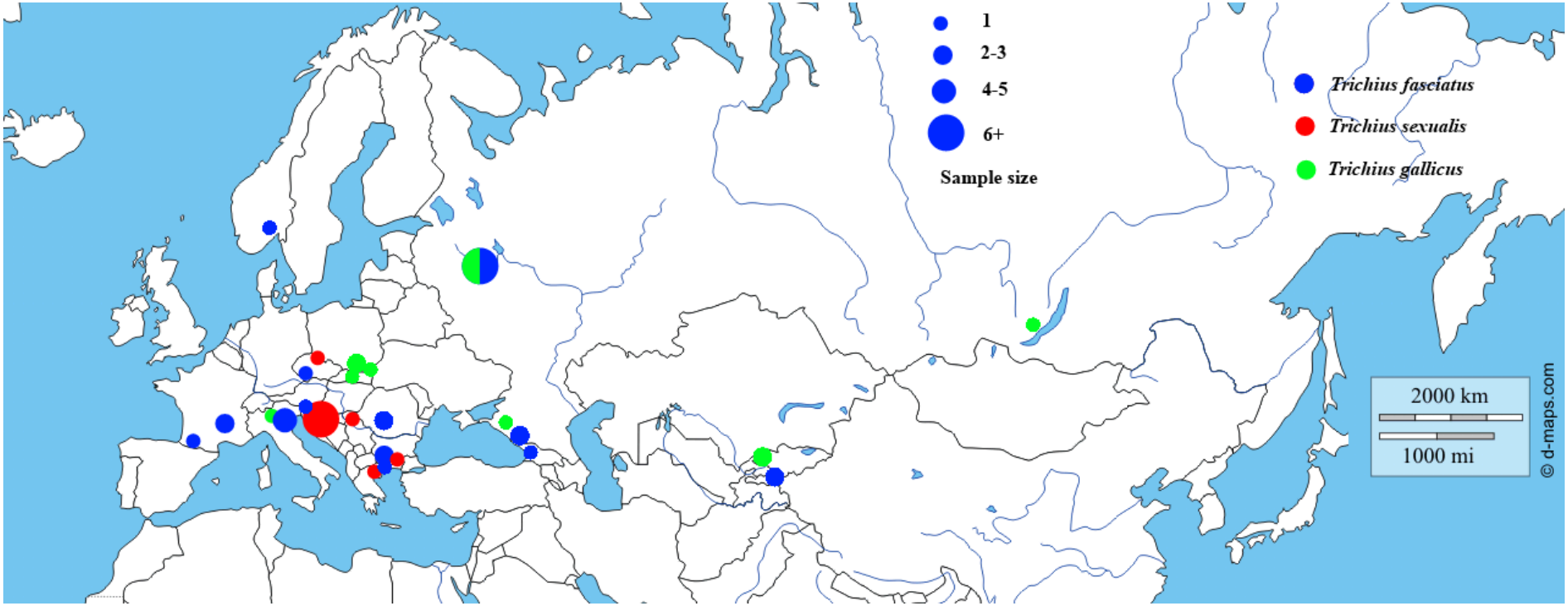


Figure 3.5. Haplotype network for COI for *Trichius* including sequences from GenBank. Haplogroups are labelled to species according to the presence of GenBank type sequences, as described in the text

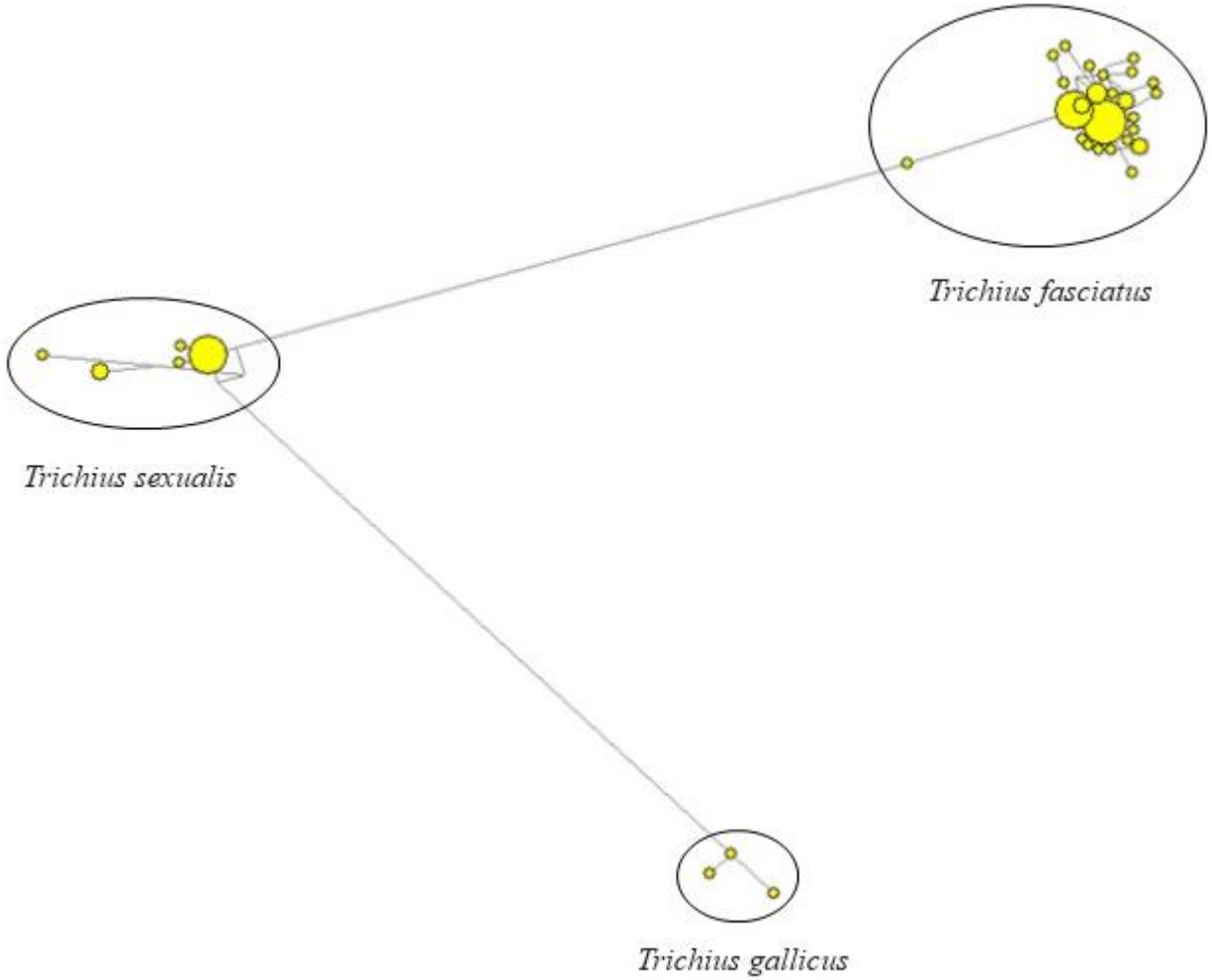


Figure 3.6. Haplotype network for COI for *Trichius* without sequences from GenBank, with haplogroups labelled as described in text. The haplotypes are coloured by the morphological identification of the individuals therein: *T. fasciatus* in blue, *T. sexualis* in red, and *T. gallicus* in green. The distance between the two haplotype clades has been resized for clarity.

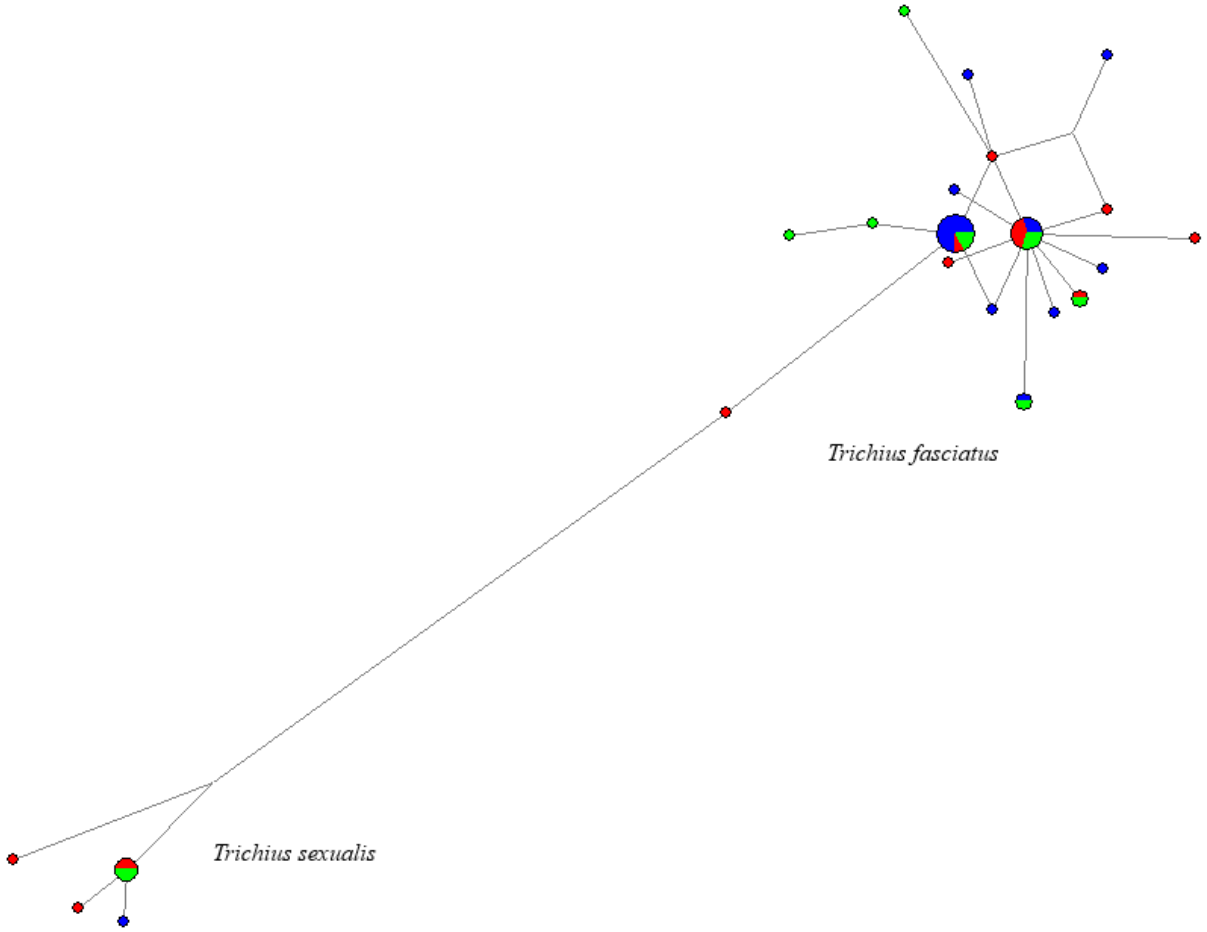


Figure 3.7: Maximum Likelihood tree computed for COI for *Trichius*. Haplogroups are labelled to species according to presence of named GenBank sequences. (% bootstrap support indicated against nodes)

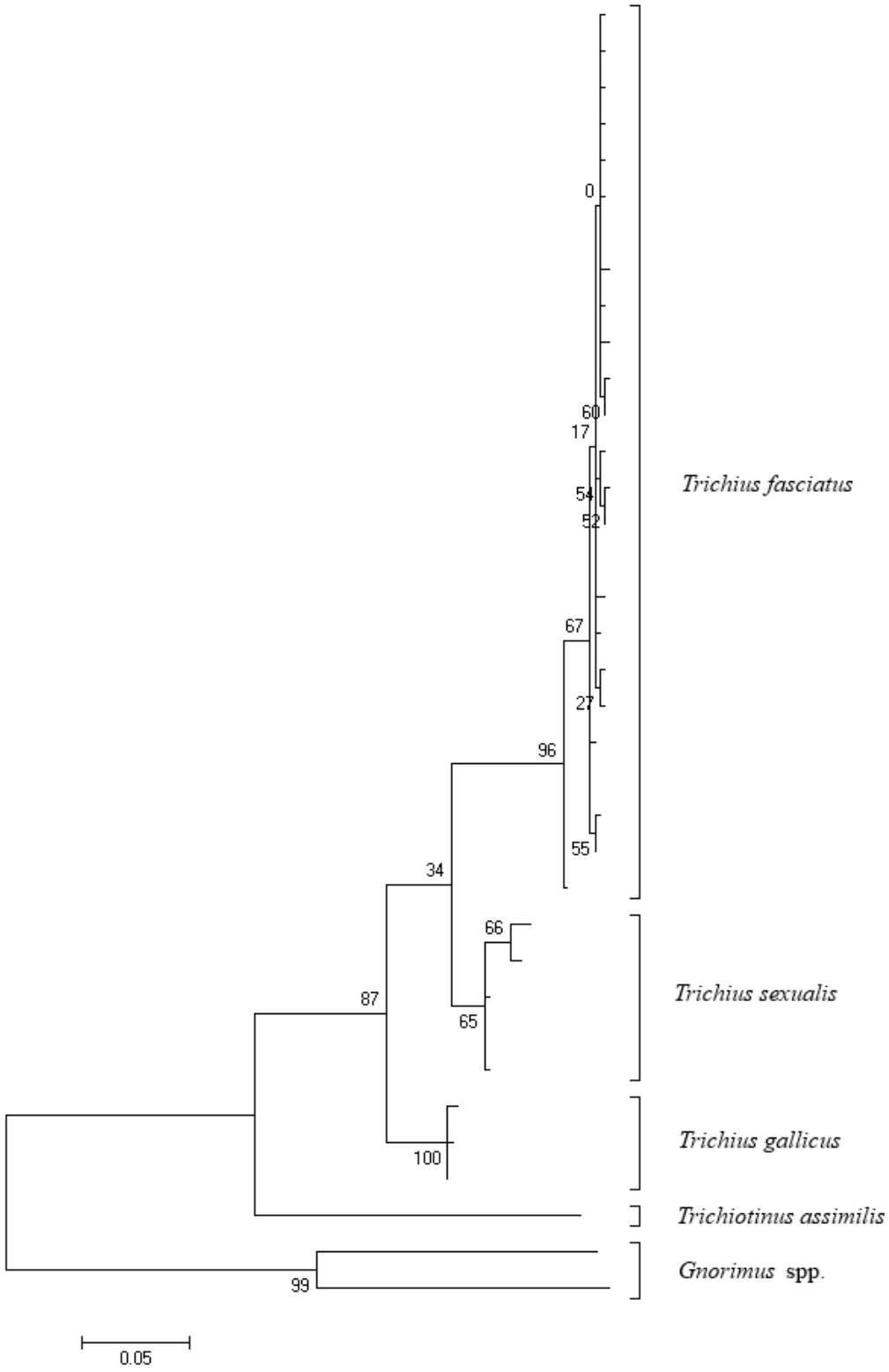


Figure 3.8: Neighbour Joining tree computed for COI for *Trichius*. Haplogroups are labelled to species according to the presence of GenBank type sequences (% bootstrap support indicated against nodes)

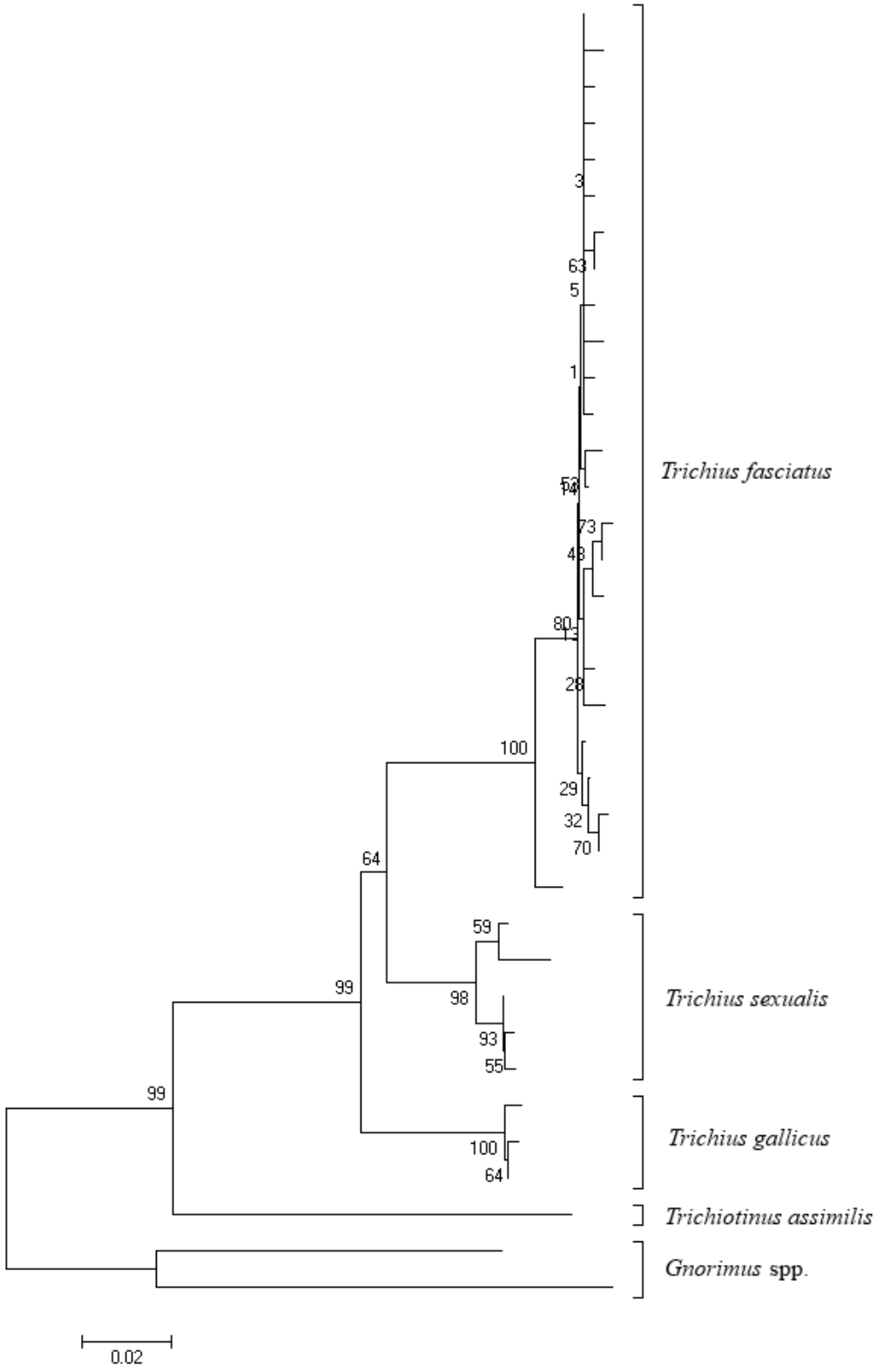


Figure 3.9: Bayesian Inference tree computed for COI for *Trichius.* Haplogroups are labelled to species according to the presence of named GenBank type sequences. Posterior Probabilities for each node are shown.

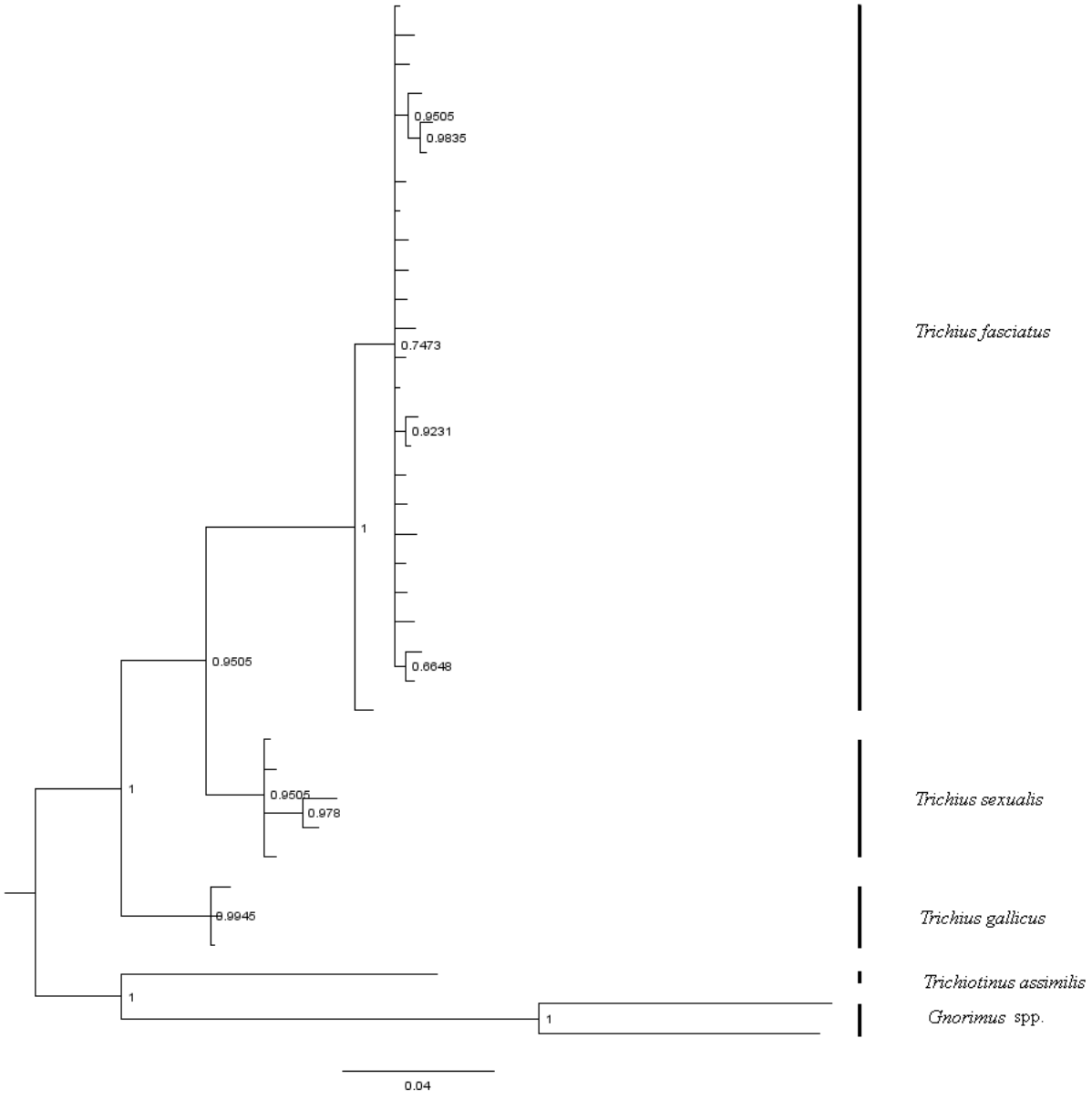


Figure 3.10. Allele network for *Wingless* sequences from *Trichius*, produced using data through PHASE, with individual composition colour coded according to COI clade: blue = *T. fasciatus*, red = *T. sexualis.*

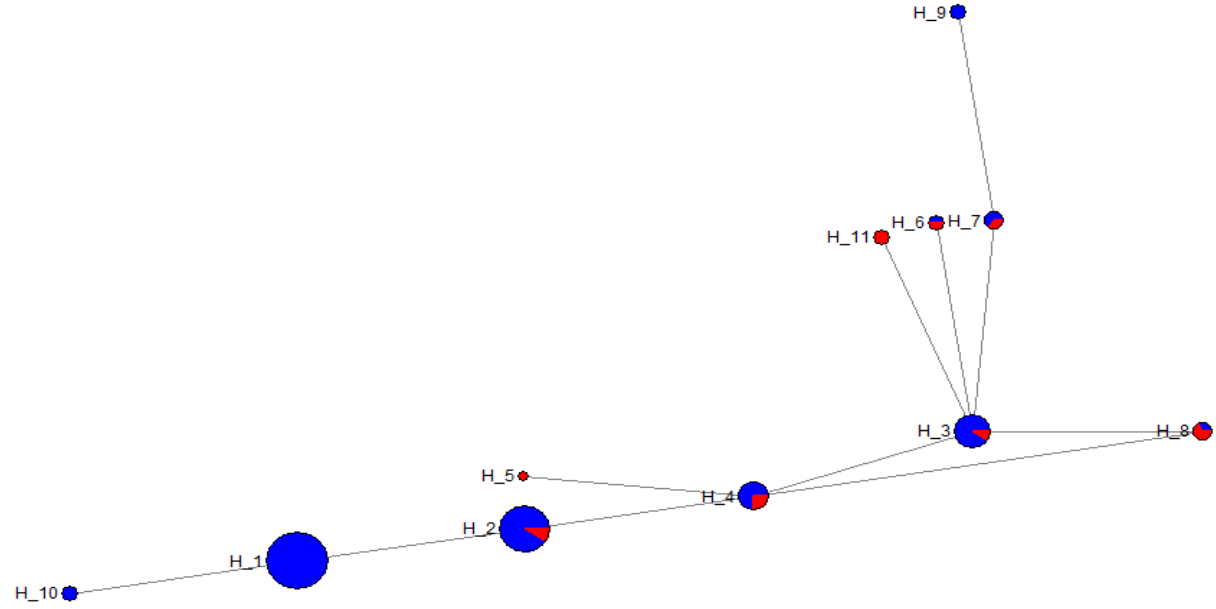
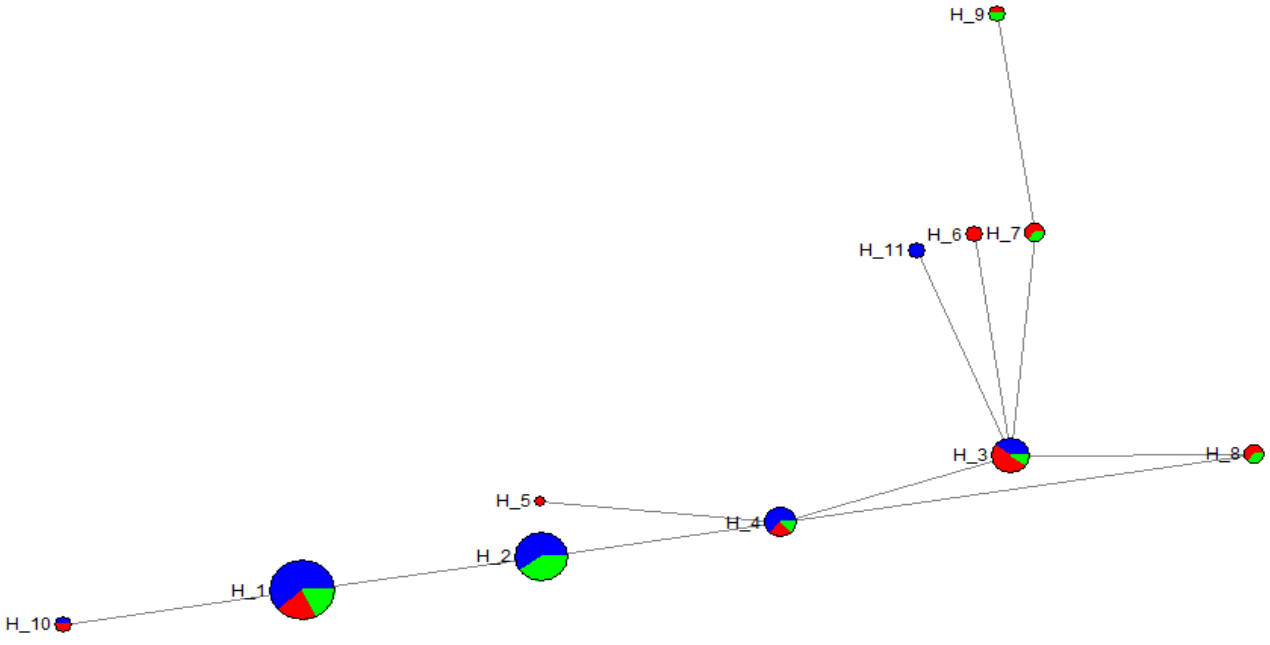


Figure 3.11. Allele network for *Wingless* from *Trichius*, produced using data through PHASE, with individual composition colour coded according to morphological identification: blue = *T. fasciatus*, red = *T. sexualis* and green = *T. gallicus*



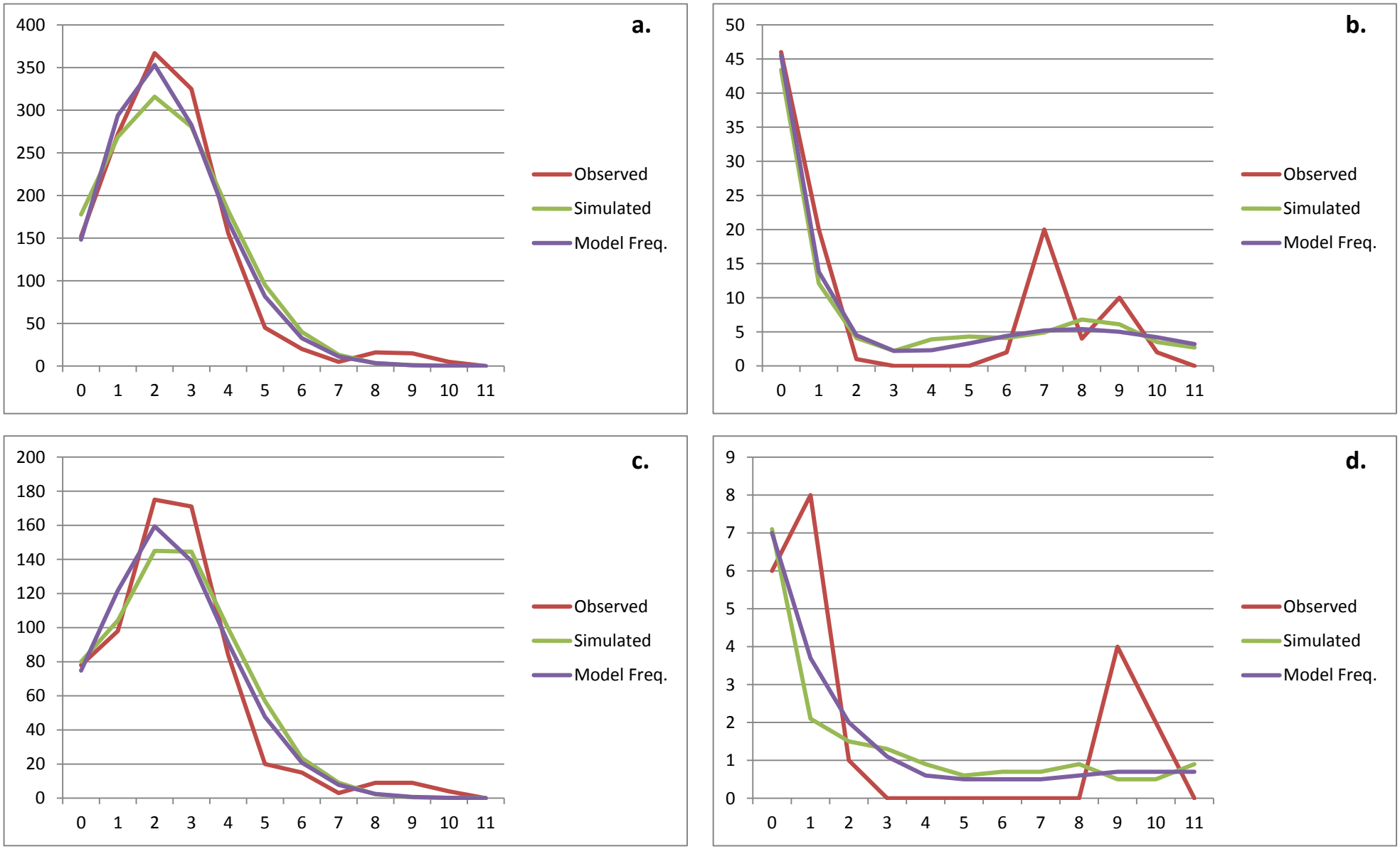
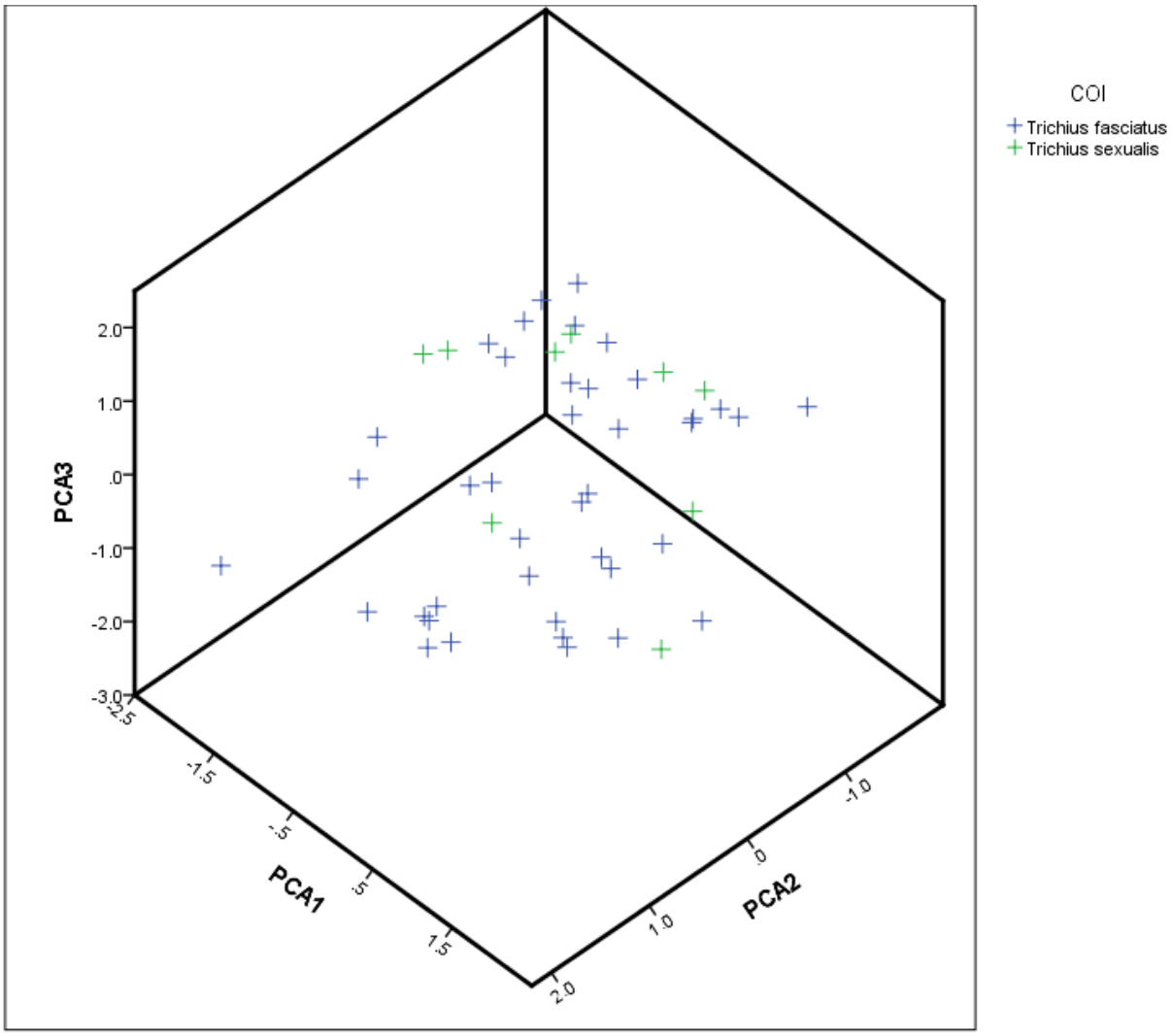


Figure 3.12: Mismatch Analysis for *Trichius* COI haplogroups: a. *T. fasciatus* including GenBank samples; b. *T. sexualis* including GenBank samples; c. *T. fasciatus* without GenBank samples; d. *T. sexualis* without GenBank samples.

Figure 3.13. 3D scatterplot showing the three PC values for each individual of both sexes colour coded by their COI haplogroup



Chapter 4

Geometric morphometrical and morphological variation across European *Gnorimus*

4.1 Introduction

The advent of DNA based taxonomy has in many cases contradicted traditional views on systematics and species identification based on morphological traits (Sites & Marshall 2003). As accurate species delimitation is fundamental to research in macroevolution, biogeography, ecology and conservation biology studies comparing traditional taxonomic approaches focussing on morphology with DNA-based inferences are needed (Wiens & Penkrot 2002; Raupach et al. 2016). Combined analysis of phenotypic and genetic divergence is also a central topic in evolutionary biology (Slatkin 1987). Phenotypic based approaches are regaining importance due to mapping and conservation efforts being increasingly driven by 'citizen science' (observations and data used with little or no verification) and the associated necessity of identifications based on external morphology or from a single photograph (Gardiner et al. 2012; Lindenmayer et al. 2013).

Photographs have also been recently adopted as holotypes for species descriptions (Marshall & Evenhuis 2015), and although controversial (Krell & Wheeler 2014), the practice is supported by the International Code of Zoological Nomenclature, and therefore might have a greater role to play in describing the 90% of biota yet to be named (Pape 2016). Knowledge of how well accessible, non-invasive methods of identification (such as external morphology) match with identifications based on more specialist methods (DNA barcoding or internal morphology) will become increasingly important for entomology and conservation as data from citizen science (collected with accessible, non-invasive identification methods) becomes more widely used.

As body size can significantly influence the shape of arthropods (Chown & Gaston 2010), geometric morphometrics can be used to describe the shape of organisms using a series of landmarks to define discrete anatomical loci homologous across the study specimens (Zelditch et al. 2012). A Procrustes superimposition is usually employed to find landmark configurations irrespective of organism size, and the orientation and position of the shape used (Rohlf & Slice 1990; Klingenberg et al. 2012). Once all individuals have had their shape extracted and superimposed, the data are subjected to a canonical variate analysis (CVA) to maximise the differences between pre-defined units relative to the variation within taxa (Campbell & Atchley 1981), thus being an efficient method to detect between units, including species and sexes. Detection of significant pairwise differences in mean shape can then be tested using Mahalanobis distance (the distance in standard deviations of point P from

the mean of D) (Klingenberg et al. 2012). Geometric morphometrics have been applied to studies of Coleoptera body shape to answer a range of biological questions, in particular looking at variation within and between species and sexes (Bai et al. 2014), or investigating changes in body and weapon shape under sexual selection (Eldred et al. 2016). Work on *Carabus* ground beetles has demonstrated significant variation in shape between the sexes and between populations at scales under 200km (Alibert et al. 2001), whilst the technique has also allowed for the identification of cryptic, previously unidentified ground beetle species (Roggero et al. 2013). Geometric morphometrics becomes increasingly powerful when combined with molecular approaches to identify subspecies and between-species relationships (Garnier et al. 2005; Zinetti et al. 2013; Ober & Connolly 2015), and analysing potential hybridisation between sister species (Pizzo et al. 2006).

The present study uses two different approaches to morphological analysis (landmark-based geometric morphometrics, and scored phenotypic characteristics) to examine the morphological variation within and between two closely related and widespread European saproxylic scarabs: the noble chafer (*Gnorimus nobilis*) and the variable chafer (*G. variabilis*). Both species are phenotypically variable, with many named colour variants (Tauzin 2004b). However, the described variants most likely represent continuous variation across the species' range. Though apparently easy to discriminate (*G. nobilis* has a cuticle reflecting 'metallic' circularly polarised light and males with a spoon shaped mesotibia, *G. variabilis* is matt, slightly larger and with curved male mesotibia) the two species share alleles at nuclear genes (Chapter 5), which may result from inter-specific hybridisation, retention of ancestral polymorphism, or specimen misidentification or regional morphological variation, so there are potential questions about the taxonomic status of the species and how they can be delimited morphologically.

Both *Gnorimus* species show an east-west split in mitochondrial DNA haplogroups (clades) across Europe, which may indicate the presence of different subspecies or evolutionarily distinct units. In addition, *G. nobilis* has three recognised subspecies: *G. nobilis nobilis* found throughout Europe, *G. nobilis bolshakovi* (Gusakov 2002) from one locality in Western Russia, and *G. nobilis macedonicus* (Baraud 1992) from Macedonia. Though the genetic divergence of *G. nobilis bolshakovi* individuals supports recognition of subspecies status, genetic affiliation of *G. nobilis macedonicus* individuals is much less clear and so the merit of separate of this subspecies open to question (Chapter 5). Additionally, the two new subspecies were named on the basis of colouration and lack of the spoon-shaped metatibia

found on male *G. n. nobilis*, thus there could be subtle differences in shape and scored phenotype which have not been identified. However, Tauzin (2004a) suggested the two new subspecies be synonymised due to their similarity. As a comparison to both sympatric European species, the eastern Palaearctic *G. subopacus* was also included in part of the analysis.

The two morphological analyses applied differ in their applicability and sample requirements. The general phenotype analysis uses phenotype assessable from non-standard conditions (i.e. field observations and/or photographs) and can be applied to living specimens anywhere. The geometric analysis uses phenotype only assessable from dead specimens in a standardised position. Using both methods together allows for a combined analysis of the power of morphological analyses to correctly identify previously assessed groups (mtDNA clades, subspecies, sex) together with clinal data and spatial differences which have not been identified by genetic study. Thus the power of both methods to describe interspecific and intraspecific heterogeneity can be analysed and discussed in light of other studies and utility for easy specimen classification.

4.2 Methods

4.2.1 Sample Collection

Specimens of *Gnorimus* spp. were obtained from researchers, the Natural History Museum (London, UK), Museo Zoologico La Specola (Florence, Italy) and commercial beetle suppliers (Supplementary Table 7). Species and subspecies included in the study were *G. nobilis nobilis, G. nobilis bolshakovi, G. nobilis macedonicus, G. variabilis,* and *G. subopacus*. *G. nobilis bolshakovi* is the only known *G. nobilis* subspecies from Russia, whilst *G. nobilis macedonicus* is only described in Macedonia (Baraud 1992; Rozner *pers. comm.*). Russian samples, plus additional specimens from Eastern Ukraine, shared the same COI haplotypes (Chapter 5) and so are here grouped into one 'Eastern *G. nobilis* clade' which may represent *G. nobilis bolshakovi*. The three specimens of *G. nobilis macedonicus* collected possess either COI haplotypes common to *G. nobilis nobilis* (n=2) or to the Eastern clade (n=1), and so were included in either the 'Western' (*G. nobilis nobilis*) or 'Eastern' clades respectively. Though most specimens were used both in the morphological analysis and in the

genetic analysis (Chapter 5), the need for intact undamaged specimens in the morphological analysis prevented some individuals being included in the geometric morphometrics analysis.

Each specimen was photographed with a Canon EOS 7D (Canon Inc., Tokyo, Japan) mounted to either a Canon 50mm f/2.5 Compact Macro, or a Sigma 105mm f/2.8 DG EX Macro (Sigma Corporation, Kanagawa, Japan) lens. This was set above the specimens on a tripod with a spirit level to avoid lens-induced warp. Specimens were set and repositioned to avoid any shape warping if this had not already been done. Individuals were sexed based on the shape of the mesotibia (spoon shaped in males), the shape of tergite 5 (indented in females) and presence/absence of a sternal furrow (present in males).

24.2 Phenotypic Scoring

A range of morphological characters described in the literature as separating *G. nobilis, G. variabilis, G. subopacus,* the different subspecies of *G. nobilis,* and representing some of the variation apparent within species (Baraud 1992; Tauzin 2004a, Tauzin 2004b) were selected and scored for each individual (see Table 4.1, Fig. 4.1). Colour was initially considered as a variable, considering that this is the primary consideration used by Tauzin (2004a; 2004b) to separate *G. nobilis* into different named variants/aberrations. However, due to the reflection of circularly polarised light by this species, depending on the direction of the light individuals can often appear to be two colours simultaneously (most commonly red and green, but also bronze or purple) due to their rotation relative to the eye or lens. Because the photographs used had been taken in a variety of conditions, including different photographic stages in the laboratory and in museums, it was impossible to standardise photographic conditions across all individuals and to therefore get a true representation of the colour of the individuals in question. Therefore, colour was not considered in the analysis. Additionally, the length of the left elytra (as a proxy for body size, as it is not affected by the position the beetle has been mounted/killed in) was measured using a digital calliper.

Three datasets were then built: both sexes together, males only including the shape of the metatibia, and females only. Other than the metatibia, no character used is known to vary based on sex. Each dataset was tested for normality in SPSS (version 22) using the Shapiro-Wilk test. As the data were shown to be a mixture of normal and non-normal distributions, Principle Components Analysis was used as the most appropriate way to reduce the complexity of the dataset. Components were computed and checked for co-variation. The

three components explaining the greatest proportion of variation within the dataset were taken forward for the rest of the analysis. Individuals were grouped into the 'haplogroups' (Chapter 5, henceforth referred to as taxa) *Gnorimus variabilis* (n=15), *G. subopacus* (n=7), Western *G. nobilis* (n=64) and Eastern *G. nobilis* (n=11) as inputs into the discriminant analysis, and the success of the discriminant analysis was checked using a leave-one-out cross-validation method, followed by a one way ANOVA and a Tukey post-hoc test to assess the ability of each PC to identify each morphogroup. As the analysis hinged on the haplotype identification, no museum specimens were used for *G. nobilis* and *G. variabilis*, but due to a small sample size (n=1) in *G. subopacus*, 6 museum individuals for this species were included.

In addition, as both *Gnorimus nobilis bolshakovi* and *G. n. macedonicus* were described based on the lack of a spoon-shaped callus on the male metatibia, the variation in this characteristic across Europe was tested, along with analysis of shape variation correlating with leg shape. The metatibia shape varies from being straight and "normal", similar to female legs and the legs of other male *Gnorimus* species, to having a large, hairless, black spoon-shaped callus. These two extremes were scored as '1' and '4' respectively, with intermediate forms scoring 2 or 3. In addition to the samples used above, geotagged photographs of male *G. nobilis* from internet resources were used to obtain additional scores for the metatibia to improve sample size. These scores were then added to a map to investigate the geographic pattern of variation in male leg shape. If only specimens from Russia and Macedonia possess the straight leg shape, then this would indicate that this is a suitable marker for identifying these two subspecies. Finally, the workflow from S2.2 was followed (ANOVA, CVA, Discriminant analysis) to see if there were any differences in shape between males with different leg shapes, and if the different male shape correlated with leg shape.

4.2.3 Landmark Geometric Morphometric Analysis

For the geometric morphometrics, 23 landmarks were chosen for the dorsal (Fig. 4.2) and 31 landmarks were chosen for the ventral (Fig. 4.3) sides and scored using tpsDig2 (ver. 2.28; Rohlf 2015). Landmarks chosen were similar to those used in other geometric studies for beetles (e.g. Benítez et al. 2013, Zinetti et al. 2013; Eldred et al. 2016), but without including certain features which could be easily warped (the position of the head for instance is very

variable and difficult to set) or could be missing from certain specimens (many individuals had a leg removed for DNA extraction). In addition, landmarks hypothesised to capture some sex-specific shape differences, such as the three landmarks surrounding the terminal tergite to analyse variation in the female indent, were also included. All landmarks were used at definitive hard points, such as the apex of a shape, or the maximum point on a curved surface, to capture variation on as many hard parts of the animals as possible (such as the elytra, pronotum, scutellum, etc.) without relying on characters that may be difficult to score or highly variable, such as the distribution of hairs on the underside, or variation in cretaceous surfaces. All three putative subspecies of *Gnorimus nobilis* were grouped as one species in the analysis against *G. variabilis*. Due to low sample sizes and issues with warp in some of the photographs, *Gnorimus subopacus* was excluded from this analysis.

From the landmark scores, Canonical Variates Analysis (CVA) was implemented in MorphoJ (ver. 1.06; Klingenberg 2011) to determine if pre-defined groups can be statistically distinguished based on multivariate data. CVA therefore summarizes the description of differences among groups relative to within-group variation (irrespective of how this relates to variation across all specimens). To test the ability of CVA to identify between the different groups, a discriminant analysis was completed on the data. Four separate analyses were completed: one to test for shape differences between the Western and Eastern clades of *Gnorimus nobilis* (as grouped in S2.2), and one to test for shape differences between *G. nobilis* and *G. variabilis*, each repeated for the dorsal and ventral landmarks separately. The two clades of *G. nobilis* were grouped for the comparison with *G. variabilis*. Each group was also split by sex, thus four groups were included in each of the dorsal and ventral analyses.

The groups suitable for the CVA on dorsal landmarks were male *G. nobilis nobilis* (n=77), female *G. nobilis nobilis* (n=43), male *G. nobilis bolshakovi* (n=9), female *G. nobilis bolshakovi* (n=2), female *G. nobilis macedonicus* (n=1), male *G. variabilis* (n=23), and female *G. variabilis* (n=24). The groups suitable for the CVA on ventral landmarks were male *G. nobilis nobilis* (n=50), female *G. nobilis nobilis* (n=26), male *G. nobilis bolshakovi* (n=9), female *G. nobilis bolshakovi* (n=2), female *G. nobilis macedonicus* (n=1), male *G. variabilis* (n=13), and female *G. variabilis* (n=14). No male *G. nobilis macedonicus* was included. Because many museum specimens were pinned to card they could not be used in the ventral analysis, hence the discrepancy in sample sizes.

A Procrustes ANOVA was used to check each dataset for asymmetry, a key source of error in landmark analysis (Palmer & Strobeck 1986; Klingenberg et al. 2002). The statistical significance of any differences in mean shape for both the dorsal and ventral surfaces between sexes and species/subspecies was assessed using 10,000 permutations of the Mahalanobis statistic computed in MorphoJ. A discriminant analysis was then performed for each subgroup comparison to analyse the ability of the two CVs to identify each species and sex.

4.3 Results

4.3.1 Phenotypic Scoring Analysis

4.3.1.1. Differences between species, subspecies and sexes

For *Gnorimus* with no sex-related splits in the data, the first three principle components (PCs) together explained 72.8% of the variation in the dataset. The three PCs used correlated with: high scores for spotting on the pronotum and tergite 5, and extensively punctuated scutella in PC1, long elytra, low levels of elytra ribbing, and extensively spotted elytra in PC2, and long elytra, strong ribbing and few spots on tergite 5 in PC3 (Table 4.2). Individuals were correctly identified to the input taxon identification in the dataset 53.9% (cross validated) of the time (Supplementary Table 8) via the discriminant analysis. *G. subopacus* was correctly identified 100% of the time, *G. variabilis* was correctly identified 96.3% of the time, Western *G. nobilis* were correctly identified 35.0% of the time, and Eastern *G. nobilis* were correctly identified 64.3% of the time (all cross-validated). The ANOVA showed that the mean differences between each groups PC values were significantly greater than the mean differences within in all three PCs from the whole population ($P<0.001$, Supplementary Table 9), with all group pairs other than Western and Eastern *G. nobilis* being discriminated by the analysis (Supplementary Table 10; 4.4)

For female *Gnorimus*, the first three principle components (PCs) together explained 73.3% of the variation in the dataset. The three PCs used correlated with: high scores for spotting on the pronotum and tergite 5, and extensively punctuated scutella in PC1; long elytra, high scores for elytra spotting, and smooth scutella in PC2; and long elytra, presence of strong

elytral ribs, and few spots on tergite 5 in PC3 (Table 4.3). Individuals were correctly identified to their input 'haplotype' (Western *G. nobilis*, Eastern *G. nobilis*, *G. variabilis* and *G. subopacus*) identification in the dataset 59.3% (cross validated) of the time (Supplementary Table 11) via the discriminant analysis. *G. subopacus* was correctly identified 100% of the time, *G. variabilis* was correctly identified 93.3% of the time, Western *G. nobilis* were correctly identified 39.4% of the time, and Eastern *G. nobilis* were correctly identified 25% of the time. The ANOVA showed that the mean difference between groups was significantly greater than the mean differences within (P<0.001, Supplementary Table 12), and Tukey post-hoc testing showed that PC1 split (with P values <0.05) Western *G. nobilis* from *G. subopacus* and *G. variabilis*, Eastern *G. nobilis* from *G. subopacus* and *G. variabilis*, and *G. subopacus* from *G. variabilis*. PC2 split Western *G. nobilis* and *G. variabilis*, and Eastern *G. nobilis* and *G. variabilis*. PC3 split Western *G. nobilis* and *G. subopacus*, and Western *G. nobilis* and *G. variabilis* (Supplementary Table 13; Fig. 4.5).

For male *Gnorimus*, the first three principle components (PCs) together explained 67.7% of the variation in the dataset. The three PCs used correlated with: high scores for spotting on the elytra, pronotum and tergite 5, and extensively punctuated scutella in PC1, long elytra, low levels of elytra ribbing, smooth scutella, and non-spoon shaped mesotibia in PC2, low levels of elytra ribbing, low levels of pronota spotting, high levels of spotting on tergite 5, and spoon shaped legs in PC3 (Table 4.4). Individuals were correctly identified to the input haplotype identification in the dataset 80% (cross validated) of the time (Supplementary Table 14) via the discriminant analysis. *G. subopacus* was correctly identified 83.3% of the time, *G. variabilis* was correctly identified 91.7% of the time, Western *G. nobilis* were correctly identified 76.1% of the time, and Eastern *G. nobilis* were correctly identified 90% of the time. The ANOVA showed that the mean difference between groups was significantly greater than the mean differences within in all three PCs (P<0.001, Supplementary Table 15). Tukey post-hoc testing showed that PC1 split (with P values <0.05) Western *G. nobilis* from *G. subopacus*, Eastern *G. nobilis* from *G. subopacus*, and *G. subopacus* from *G. variabilis*. PC2 split Western *G. nobilis* and *G. variabilis*, Eastern *G. nobilis* and Western *G. nobilis*, Eastern *G. nobilis* and *G. variabilis*, *G. subopacus* and *G. variabilis*, and Eastern *G. nobilis* and *G. variabilis*. PC3 split Western *G. nobilis* and *G. subopacus*, Western *G. nobilis* and *G. variabilis*, Eastern *G. nobilis* and *G. subopacus*, Western *G. nobilis* and *G. variabilis*, and *G. variabilis* and *G. subopacus* (Supplementary Table 16; Fig. 4.6).

4.3.1.2 Male leg shape

Contrary to expectations, investigating the male leg shape showed that many more samples than just the putative non-nominate subspecies possessed straight mesothoracic legs (Fig. 4.7). No asymmetry was detected in the dataset ($P<0.0001$). Once scaled by the inverse of the within-group variation, CV1 in the dorsal dataset explained 63.718% of the within-group variation, whilst CV2 explained 24.707%. The variates in morphospace are displayed in Supplementary Figure 1, grouped by mesotibia score, along with 95% confidence ellipses. The most extreme values uncovered in the data are displayed in Supplementary Figures 2 & 3. CV1 shows high scores with wider individuals with wider, longer pronota, and short abdomens. CV2 also weighs wider pronota, and pointed posterior elytra edges. The discriminant analysis showed that scores '1' and '4' could be differentiated ($P=0.0135$) as could '2' and '4' ($P=0.0102$) (Supplementary Table 17)

For the ventral analysis, no asymmetry was detected in the dataset ($P<0.0001$). Once scaled by the inverse of the within-group variation, CV1 in the dorsal dataset explained 62.415% of the within-group variation, whilst CV2 explained 26.690%. The variates in morphospace are displayed in Supplementary Figure 4, grouped by mesotibia score, along with 95% confidence ellipses. The most extreme values uncovered in the data are displayed in Supplementary Figures 5 & 6. CV1 shows high scores with wider individuals with wider, longer pronota, and narrower, longer abdomens. CV2 weighs wider and longer pronota, and a longer thorax. The discriminant analysis showed that only scores '2' and '4' could be differentiated ($P=0.017$). All scores of 3 & 4 belonged to individuals from the Western *G. nobilis* taxon, and all individuals from the Eastern *G. nobilis* taxon possessed scores of 1 & 2. However, many individuals from the Western *G. nobilis* taxon clade in the east of the range possessed legs with scores of 1 & 2 (compare Fig. 5.6 to Fig. 4.9) (Supplementary Table 18).

4.3.2 Landmark Geometric Morphometric Analysis

4.3.2.1 Dorsal *Gnorimus nobilis* against *G. variabilis*

No asymmetry was detected in the dataset ($P<0.0001$). Once scaled by the inverse of the within-group variation CV1 explained 59.79% of the within-group variation, whilst CV2 explained 35.39%. The variates in morphospace are displayed in Figure 4.8, grouped by sex

and subspecies, along with 95% confidence ellipses. The most extreme values uncovered in the data are displayed in Figures 4.9 & 4.10. CV1 is very powerful at splitting the two species, with *G. variabilis* (generally high scores) showing longer pronota that are wide at the anterior, the longest part of the elytra closer to the centre of the body, and the widest point of the elytra toward the anterior. CV2 split the sexes, with females showing longer abdomens with a notch, wider pronota, and shorter elytra. The discriminant analysis showed that all groups could be differentiated from each other ($P<0.0001$, Table 4.5).

4.3.2.2 Ventral *Gnorimus nobilis* against *G. variabilis*

No asymmetry was detected in the dataset ($P<0.0001$). Once scaled by the inverse of the within-group variation CV1 explained 72.62% of the within-group variation, whilst CV2 explained 23.12%. The variates in morphospace are displayed in Figure 4.11, grouped by sex and subspecies, along with 95% confidence ellipses. The most extreme values uncovered in the data are displayed in Figures 4.12 & 4.13. CV1 is very powerful at splitting the two sexes, with females (generally low scores) showing slightly wider pronota at the anterior and slimmer at the posterior, longer slimmer abdomens with greater separation between the segments and a large 'notch', and a shorter, narrower thorax. CV2 showed that *Gnorimus variabilis* has a slightly narrower pronotum at the anterior but wider at the posterior, and a wider shorter abdomen. The discriminant analysis showed that all groups could be differentiated from each other ($P<0.05$), other than *G. variabilis* males and females ($P=0.903$, Table 4.6).

4.3.2.3 Dorsal shape of *Gnorimus nobilis* subspecies

No asymmetry was detected in the dataset ($P=0.0007$). Once scaled by the inverse of the within-group variation CV1 explained 77.31% of the variation within the dataset, whilst CV2 explained 14.16%. The variates in morphospace are displayed in Figure 4.14, grouped by sex and subspecies, along with 95% confidence ellipses. The most extreme values uncovered in the data are displayed in Figures 4.15 & 4.16. CV1 is very powerful at splitting the two sexes, with females (generally high scores) showing wider pronota, shorter slimmer elytra and a longer abdomen. Though there was a lot of variation within both sexes, CV2 showed that both *G. nobilis macedonicus* and *G. nobilis bolshakovi* differ in shape to *G. nobilis nobilis* by having wider pronota, slightly longer abdomens, and considerably slimmer elytra. The

discriminant analysis showed that *G. nobilis nobilis* males and females were significantly different in overall shape (P<0.0001), as were *G. nobilis nobilis* and *G. nobilis bolshakovi* males (P=0.002), and *G. nobilis nobilis* females and *G. nobilis bolshakovi* males (P<0.0001, Table 4.7). Analysing the distribution of individuals in morphospace by eye showed there to be no clinal variation within each group: individuals from the north and south, and east and west, were found throughout morphospace with no sub-grouping.

Additionally, data were also split into other groups to test additional phylogeographic hypotheses, such as differences between western *G. nobilis* sub-populations (Italy, Spain, the Balkans). None of these additional phylogeographic hypotheses showed any subsequent physical subdivision within western *G. nobilis*.

4.3.2.4 Ventral shape of *Gnorimus nobilis* subspecies

No asymmetry was detected in the dataset (P=0.017). Once scaled by the inverse of the within-group variation CV1 explained 85.56% of the within-group variation, whilst CV2 explained 9.19%. The variates in morphospace are displayed in Figure 4.17, grouped by sex and subspecies, along with 95% confidence ellipses. The most extreme values uncovered in the data are displayed in Figures 4.18 & 4.19. CV1 is very powerful at splitting the two sexes, with females (generally low scores) showing slightly wider pronota, longer slimmer abdomens with greater separation between the segments, and a shorter, narrower thorax. CV2 showed that both *G. nobilis macedonicus* and *G. nobilis bolshakovi* differ in shape to *G. nobilis nobilis* by having wider pronota, slightly longer slimmer abdomens, and a slimmer, shorter thorax. The discriminant analysis showed that *G. nobilis nobilis* males and females could be told apart by shape (P<0.0001), as could *G. nobilis nobilis* and *G. nobilis bolshakovi* males (P<0.0001), *G. nobilis nobilis* females and *G. nobilis bolshakovi* males (P=0.016), *G. nobilis nobilis* males and *G. nobilis bolshakovi* females (P=0.0003), and *G. nobilis nobilis* males and *G. nobilis macedonicus* females (P=0.003, Table 4.8).

4.4 Discussion

The geometric analysis demonstrates clearly that *Gnorimus nobilis* and *G. variabilis* can be distinguished from one another by shape alone to a high degree of confidence. *G. nobilis* had wider and longer elytra than *G. variabilis*, as well as pronota that were wider at the front than *G. variabilis*, but slimmer at the back, and wider, shorter abdomens (Figs. 4.9, 4.10, 4.12,

4.13). Both species could be differentiated based on the dorsal and ventral geometric analyses and in every morphological scoring analysis, as were both from *G. subopacus* (not included in the landmark analysis).

However, cross-validated results from the general phenotype analysis were less consistent than for the landmark: from the female analysis, one *G. variabilis* (out of 15) was classified as a Western *G. nobilis*, two Western *G. nobilis* were classified as a *G. variabilis* (out of 33); in the male analysis 4 Western *G. nobilis* were classified as *G. subopacus* (out of 67), and one *G. variabilis* was classified as Eastern *G. nobilis* (out of 12). Though this performance is relatively strong, in the combined analysis more individuals were misclassified: 15 Western *G. nobilis* as *G. subopacus* (out of 100), 2 Western *G. nobilis* as *G. variabilis*, and one *G. variabilis* as Eastern *G. nobilis*. Therefore the characters described as being diagnostic in separating the three species (elytral ribbing and scutellum punctuation in particular) are here confirmed as being useful to separate species/subspecies regardless of sex, but they are weaker than the landmark analysis as they show greater within-species variation then previously supposed. Combining the results across the methods shows that:

Gnorimus nobilis has wide and long shaped elytra, a wide anterior pronotum, a wide, short abdomen, well punctuated scutella, absolutely shorter elytra, with spoon shaped mesotibial in some male populations.

G. variabilis had short, slim shaped elytra, a wider posterior to the pronotum, slimmer longer abdomens, scutella with punctuations largely restricted to the anterior, absolutely longer elytra, and ‘normal’ male mesotibial throughout the range.

G. subopacus was not included in the geometric morphometric analysis, but showed the highest levels of scutellum punctuation and strong elytra ribbing.

The interest in shape and phenotypic differences between *Gnorimus nobilis* and *G. variabilis* was sparked by the finding of a number of gene alleles being shared between the two species (Chapter 5), contrary to findings from ‘standard’ barcoding genes (such as COI) which indicate that the two species are distinct. One hypothesis that may explain this result was that there are melanic forms of *G. nobilis* which lack the metallic cuticle, and may have been mistaken for black *G. variabilis*. However, the shape and phenotypic characteristic differences between the two species can refute this hypothesis. In addition, though there is

some overlap between *G. nobilis* and *G. subopacus*, the combination of strong elytral ribs and intensely punctuated scutella allows *G. subopacus* to be reliably identified.

As no shape variation between the subspecies had been mentioned by their description-authors, it was an unexpected finding that the subspecies differed in shape. Though hampered by small sample sizes for some sub-groups (n=2 for *Gnorimus nobilis bolshakovi* females, n=1 for *G. nobilis macedonicus* females, and no *G. nobilis macedonicus* males), which limit some of the conclusions that can be drawn, male *G. nobilis nobilis* and *G. nobilis bolshakovi* males can be distinguished based on shape (P=0.0018 dorsal, P=0.0319 ventral), *G. nobilis bolshakovi* males having narrower longer elytra (Figs. 4.15, 4.16, 4.18, 4.19). The subspecies could not be differentiated based on phenotypic characteristics (Supplementary Table 13) in the female-only analysis, but could be differentiated in the male-only analysis, though with 10-17.9% of individuals being misidentified in the discriminant (Supplementary Table 14).

Most other analyses showed confusion and misclassification between the two subspecies: more Western *G. nobilis* were classified as Eastern *G. nobilis* in the combined morphological analysis than were correctly classified as Western *G. nobilis* (48 misclassified, 35 classified). The geometric morphometric ventral analysis split more groups in the discriminant analysis, again showing differences between the sexes and the subspecies, though *G. nobilis bolshakovi* and *G. nobilis nobilis* females could not be distinguished. Only PC2 reliably split male *G. nobilis bolshakovi* and *G. nobilis nobilis* in the phenotypic characteristic analysis. As PC2 strongly weighs low scores for mesotibial shape (i.e. non-spoon shaped) this may explain its use in splitting the two taxa. However, as PC2 also weighs longer elytra, a lack of elytral ribbing, and smooth scutella, these other characteristics may also differentiate the eastern *G. nobilis* from the western clade, though more research is needed. The low sample size and large level of variation within the species suggests that to correctly identify between the two putative subspecies, genetic and morphological markers should be used.

The leg shape analysis also indicated that there may be additional morphological variation within *G. nobilis*. There is a clear east/west split in male leg shapes; black, hairless, spoon-shaped mesotibia being largely confined to western and central Europe, and straight mesotibia in Eastern Europe. This leg shape was supposedly confined to *G. nobilis bolshakovi* and *G. nobilis macedonicus*, but finding it to be much more widespread among *G.nobilis nobilis* was unexpected, which may explain some of the confusion between male *G. nobilis nobilis* and *G. nobilis bolshakovi* in the phenotypic character analysis (above).

Following the discriminant analysis from the geometric-morphometric analysis, using both the dorsal and ventral analyses, males with leg shapes ‘1’ and ‘4’ can be differentiated (*P*=0.0135 dorsal) and shapes ‘2’ and ‘4’ can be differentiated (*P*=0.0102 dorsal and *P*=0.0166 ventral). This implies that there may be more morphological variation within *G. nobilis* than initially supposed. Though the function of these protuberances on the legs is poorly known, they may help males to hold on to females during mating (Bates *pers. comm.*), though why such an apparently useful structure isn’t found throughout the whole population is unknown. It may be that during the last glacial maximum both Balkan and Caucasian refugia for *G. nobilis* held males with normal leg shapes, but Iberian, French and Italian refugia (Chapter 5) held males with spoon-shaped protuberances. As the species spread following the northern spread of its habitat the western COI haplotype clade may have become fixed in the populations where the two expanding range edges met, then spread south.

Male *Gnorimus nobilis* have shorter abdomens without the notch, a longer thorax, and slimmer pronota at the anterior (Fig. 8). *G. nobilis bolshakovi* also have slimmer abdomens, a shorter thorax, and a slight shift forward to the position of the pro-legs (Fig. 9). The phenotypic scoring analysis did not show any differences between the two subspecies, other than PC2 between male *Gnorimus nobilis* subspecies (Tables 12 and 15), which may be related to this component strongly weighting the shape of the mesotibia.

Male and female *Gnorimus* can also be distinguished based on shape (Figs. 4 & 7), females having longer abdomens with a notch, shorter and slimmer elytra, wider pronota, and slimmer abdomens with greater spacing between the sclerites (Figs. 6 & 8). Within-species analysis for *G. nobilis* shows this shape difference to hold (Figs. 11 & 14). There are a number of reasons why shape differences between the sexes may have evolved. Wider pronota in females may have evolved to allow for more space within the prothoracic cavity for extrinsic coxal muscles (Chapman 1998), allowing for greater leverage on the coxa, and therefore on the proleg as a whole (Grey 1944). This is known from other insects that spend a degree of their lives underground (Villani et al. 1999), but is usually found along with other adaptations for digging. Female *Gnorimus* appear to have slightly shorter and broader tibiae with longer teeth toward the tarsi (Blake *pers. obvs.*), which matches knowledge from other Cetoniinae (Lachaume 1983; Allard 1985, 1986, 1991; Ratcliffe & Warner 2011). Thus a wider pronotum could be an adaptation to help females dig through pabula in rot holes to lay eggs. Slimmer elytra and bodies could also be an adaptation for this (Villani et al. 1999),

compensating for the loss of internal abdominal space by elongating it. Males could have evolved wider elytra to sheath larger wings for better dispersal; wing length to body size ratios are correlated with better dispersal abilities across Coleoptera (Harrison 1980). Males fly more often and further than females in other scarab beetles, or have adaptations likely to help in this role (such as larger wings, Kawano 1995; Rink & Sinsch 2007; Svensson et al. 2011).

Though possibly more of an issue for species which feed on food that needs to be slowly digested (*Gnorimus* feed on pollen and nectar), differences in the internal composition of the abdomen may explain the elongation detected in female *Gnorimus* compared to males and the increased level of spacing between the segments (Fig. 4.12 & 4.18). Dissection of young females preserved in ethanol has shown that mature eggs take up a large volume of space in the abdomen of females (Blake *pers. obvs.*) but there are no data showing how egg number changes during the adult life span of this genus. *Gnorimus* feed as adults on energy and nutrient-rich foods that require little digestion, but it is likely that (as in Lepidoptera; Wickman & Karlsson 1989) adults eclose with abdomens full of fat bodies as a nutrient store and with large reproductive organs. Scarabs, generally being large for insects, usually produce relatively small numbers of large eggs (Berrigan 1991, citing Iwata 1966; Blake *pers. obvs*.), and *Gnorimus* are no exception. The large eggs and fat bodies, though comprised of large quantities of water, may still take up large volumes of space when compared to old individual *Gnorimus* that have used up most of their stored resources. This may result in discrepancies between old and young individuals, and between females and males. Thus the reliability of the shape of the abdomen in identifying male or female *Gnorimus* may need more attention, but the shape of hard structures (such as the groove in the terminal female tergite) do appear to be reliable characteristics to sex *Gnorimus*.

Though the approach revealed differences in phenotype and shape between the different sexes and species investigated, conforming to prior hypotheses, some flaws in the analysis were evident, though these did not seem to affect the results. Some causes of error could include:

- using liquid preserved specimens alongside dried specimens may lead to differences in the size and shape of the soft internal organs, which could take up additional space in the abdomen, forcing it to telescope

- females may have eggs retained within the body, which as semi-hard structures, may take up space in the abdomen leading to telescoping. Additionally, young adult females may have more eggs held within their bodies than older females, which may also take up abdominal space
- well-fed individuals could also have large volumes of space taken up by food they've eaten

Despite some gene-sharing between *Gnorimus nobilis* and *G. variabilis*, the two species can be reliably identified based on shape and general phenotypic characteristics which are useful both on set specimens in a standardised environment (in museums, or the laboratory) and from high-resolution photographs in the field. Thus if melanistic variants of *G. nobilis* are found in some populations, they can be differentiated from *G. variabilis* due to their shorter elytral length, poorly punctuated scutella, and slimmer elytra and abdomens. In addition there exist some differences in shape and characteristics between the western (*G. nobilis nobilis*) and eastern (*G. nobilis bolshakovi*, and individuals which share their COI and CR haplotype (Chapter 5)) populations of *G. nobilis*, and an unexpected finding that male *G. nobilis* exist in two allopatric morphs, one in the west with large spoon-shaped metatibia, and an eastern morph with normal metatibia (Fig. 4.7). However, these populations don't largely differ in shape (Supplementary Figures 1 & 4), though there may be some subtle differences which require further study (Supplementary Tables 17 & 18). Finally, male and female *Gnorimus* can be reliably differentiated with shape, which is likely to have adaptive significance based on sex-specific differences in behaviour and biology. Additionally, these differences should be useful for accurate identification of sexes and species from photographs taken in non-controlled settings (including from data collected by citizen science initiatives), improving the data that can be collected from such schemes.

Table 4.1: Characters used in scoring morphological variation in *Gnorimus* specimens, with the maximum achievable score

Character	Explanation	Score
Elytra Ribbing	Presence and level of longitudinal ridges on the elytra	4
Elytra Spotting	Presence and level of cretaceous spots on the elytra	6
Pronotum Spotting	Presence and level of cretaceous spots on the pronotum	4
T5 Spotting	Presence and level of cretaceous spots on tergite 5	5
Scutellum Punctuation	Distribution of punctuation on the scutellum	5
Male metatibia shape	Shape of the male metatibia from straight and hairy to spoon-shaped, black and hairless	4

Table 4.2. PCA Component Matrix for all *Gnorimus*. The 3 most important variables for each component are listed in red.

	Component		
	1	2	3
Size	-.231	.768	.424
Rib	.337	-.445	.726
Espot	.593	.630	.046
Pspot	.693	.119	.272
T5Spot	.651	.263	-.425
Scutellum	-.726	.412	.118

Table 4.3. PCA Component Matrix for female *Gnorimus*. The 3 most important variables for each component are listed in red.

	Component		
	1	2	3
Size	-.449	.750	.259
Rib	.523	-.182	.744
Espot	.474	.732	-.096
Pspot	.608	.317	.232
T5Spot	.683	.219	-.430
Scutellum	-.769	.334	.096

Table 4.4. PCA Component Matrix for male *Gnorimus*. The 4 most important variables for each component are listed in red.

	Component		
	1	2	3
Size	.194	.779	-.061
Rib	.144	-.407	-.675
Espot	.786	.322	.149
Pspot	.717	-.165	-.181
T5Spot	.580	-.124	.628
Scutellum	-.531	.621	.119
Leg	-.322	-.542	.556

Table 4.5. Mahalanobis distance between groups, and the significance of the Discriminant analysis to differentiate between the dorsal groups. Significant differences between groups are marked in bold

Species and sex	*G. nobilis* M	*G. nobilis* F	*G. variabilis* M
G. nobilis F	3.5560 ***P*<0.0001**	-	-
G. variabilis M	5.0641 ***P*<0.0001**	6.4057 ***P*<0.0001**	-
G. variabilis F	6.2888 ***P*<0.0001**	5.9751 ***P*<0.0001**	5.3756 ***P*<0.0001**

Table 4.6. Mahalanobis distance between groups, and the significance of the Discriminant analysis to differentiate between the ventral groups. Significant differences between groups are marked in bold

Species and sex	*G. nobilis* M	*G. nobilis* F	*G. variabilis* M
G. nobilis F	4.9430 ***P*<0.0001**	-	-
G. variabilis M	3.9139 ***P*=0.0319**	7.1839 ***P*=0.0021**	-
G. variabilis F	7.6272 ***P*<0.0001**	7.0561 ***P*=0.0010**	5.591 *P*=0.9026

Table 4.7. Mahalanobis distance between subspecies and sex groups, and the significance of the Discriminant analysis to differentiate between the dorsal groups. Significant differences between groups are marked in bold

Subspecies and sex	*G. nobilis nobilis*, M	*G. nobilis nobilis*, F	*G. nobilis bolshakovi*, M	*G. nobilis bolshakovi*, F	*G. nobilis macedonicus*, F
G. nobilis nobilis, F	3.5330 ***P*<0.0001**	-	-	-	-
G. nobilis bolshakovi, M	2.9800 ***P*=0.0018**	5.5005 ***P*<0.0001**	-	-	-
G. nobilis bolshakovi, F	4.3754 *P*=0.2046	4.1929 *P*=0.6382	2.6383 *P*=0.9742	-	-
G. nobilis macedonicus, F	6.2315 *P*=0.1879	8.1169 *P*=0.1388	4.3592 *P*=0.9109	0.6581 *P*=0.7235	-

Table 4.8. Mahalanobis distance between ventral groups, and the significance of the Discriminant analysis to differentiate between the subspecies groups. Significant differences between groups are marked in bold

Subspecies and sex	*G. nobilis nobilis*, M	*G. nobilis nobilis*, F	*G. nobilis bolshakovi*, M	*G. nobilis bolshakovi*, F	*G. nobilis macedonicus*, F
G. nobilis nobilis, F	7.0767 ***P*<0.0001**	-	-	-	-
G. nobilis bolshakovi, M	3.9139 ***P*=0.0319**	12.4072 ***P*=0.0163**	-	-	-
G. nobilis bolshakovi, F	12.2141 ***P*=0.0003**	10.5705 *P*=0.9373	5.0728 *P*=0.8012	-	-
G. nobilis macedonicus, F	15.2859 ***P*=0.0026**	31.3945 *P*=0.6108	5.3553 *P*=0.8460	0.3576 *P*=0.8191	-

Figure 4.1. Examples of the phenotypic characters used in the study, with an example of an individual with the lowest score to the left, and a specimen with the highest achievable score on the right. a & b: elytra ribbing. c & d: elytra spotting. e & f: pronotum spotting. g & h: scutellum punctuation. i & j: male metatibia.

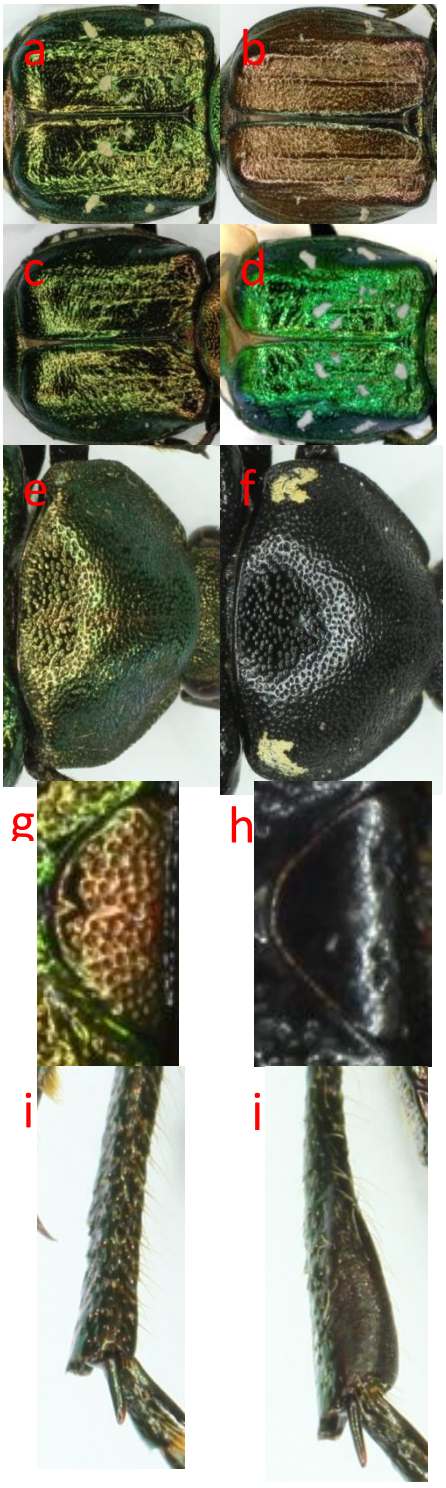


Figure 4.2: Geometric morphometric landmarks used for the dorsal surface of *Gnorimus*

Figure 4.3: Geometric morphometric landmarks used for the ventral surface of *Gnorimus*

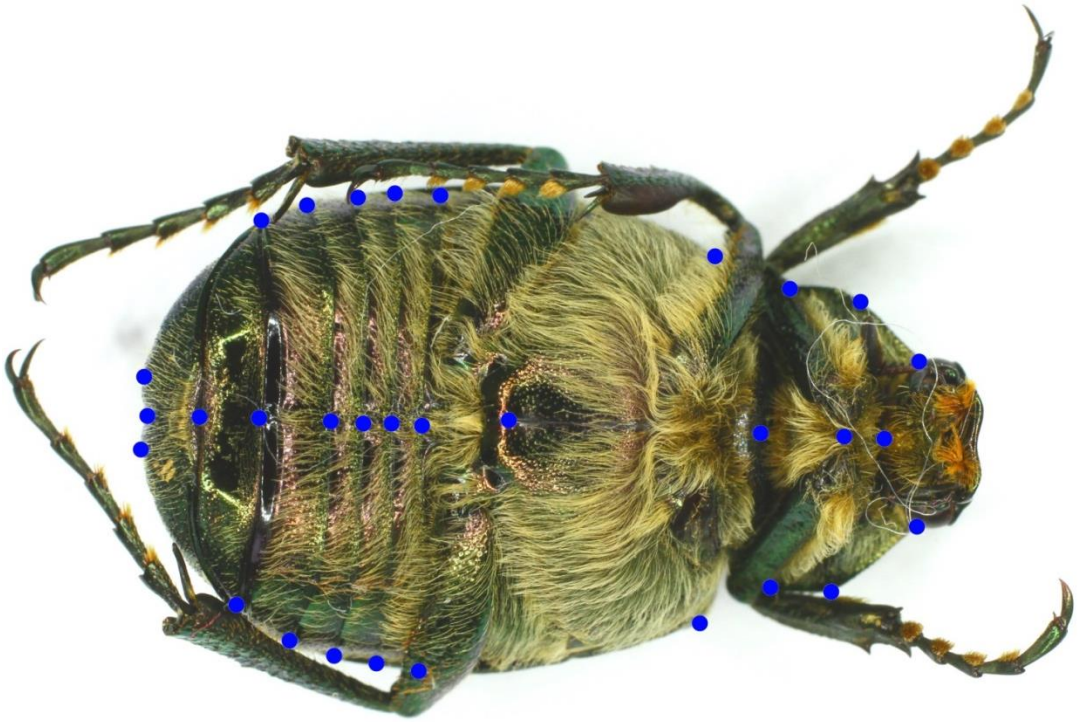


Figure 4.4 – 3D scatter plot showing the three principle components computed for *Gnorimus*

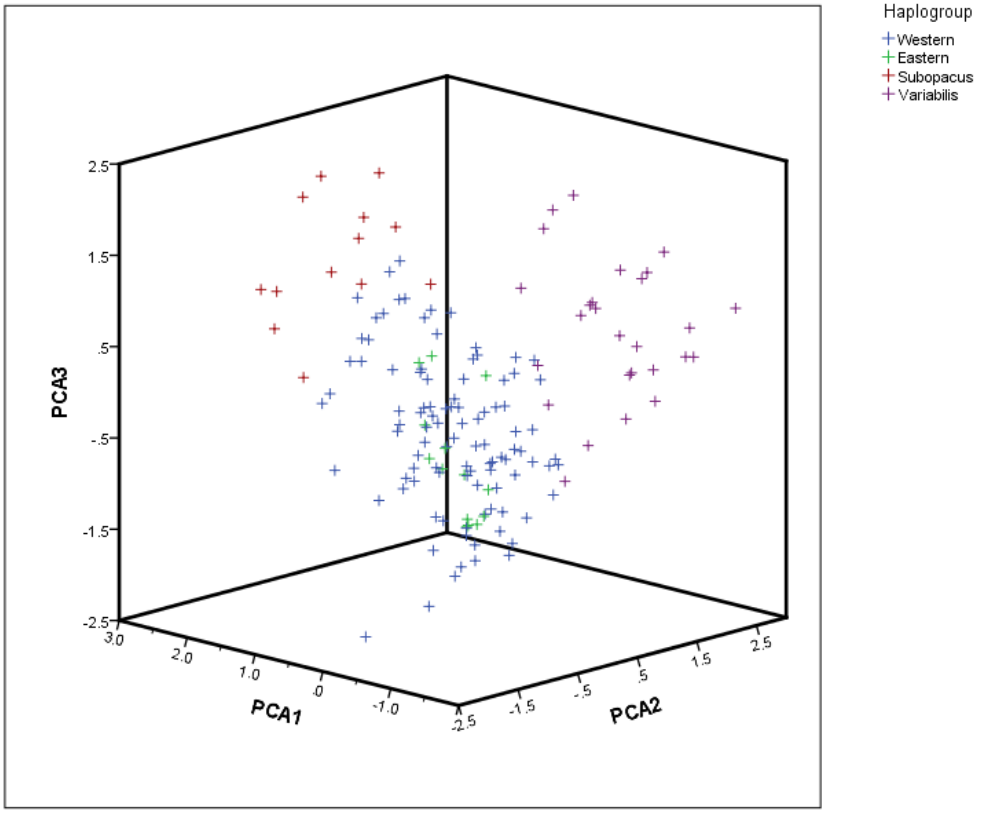


Figure 4.5 – 3D scatter plot showing the three principle components computed for the female *Gnorimus*

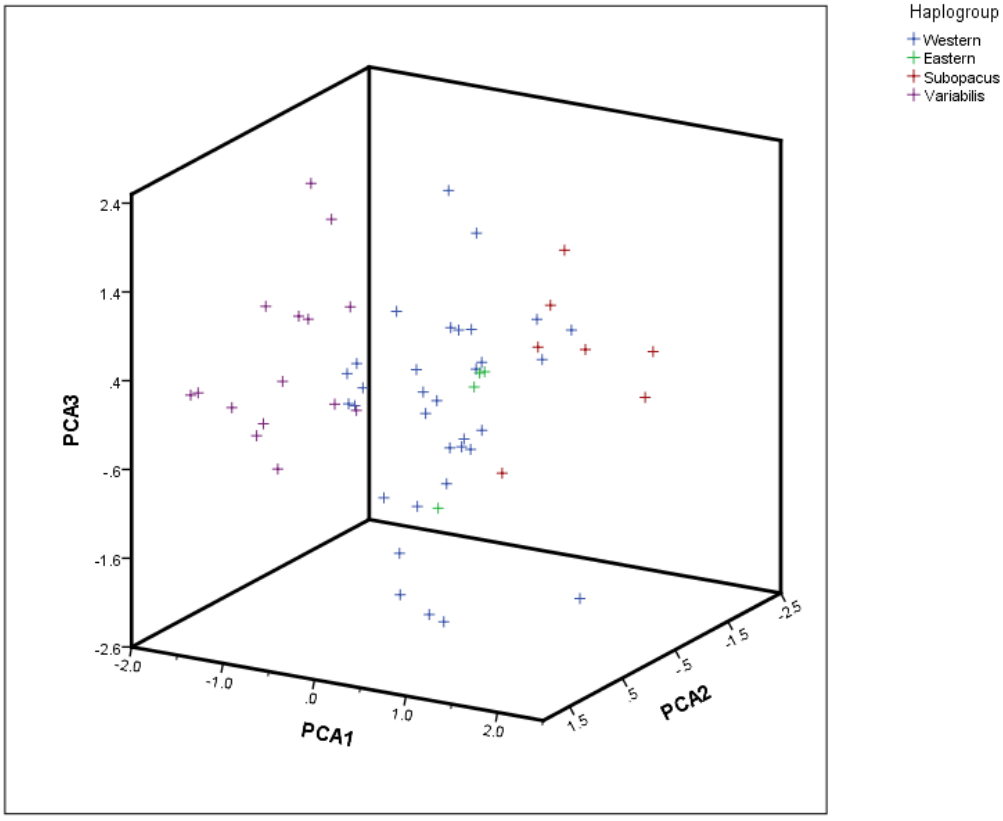


Figure 4.6 – 3D scatter plot showing the three principle components computed for the male *Gnorimus*

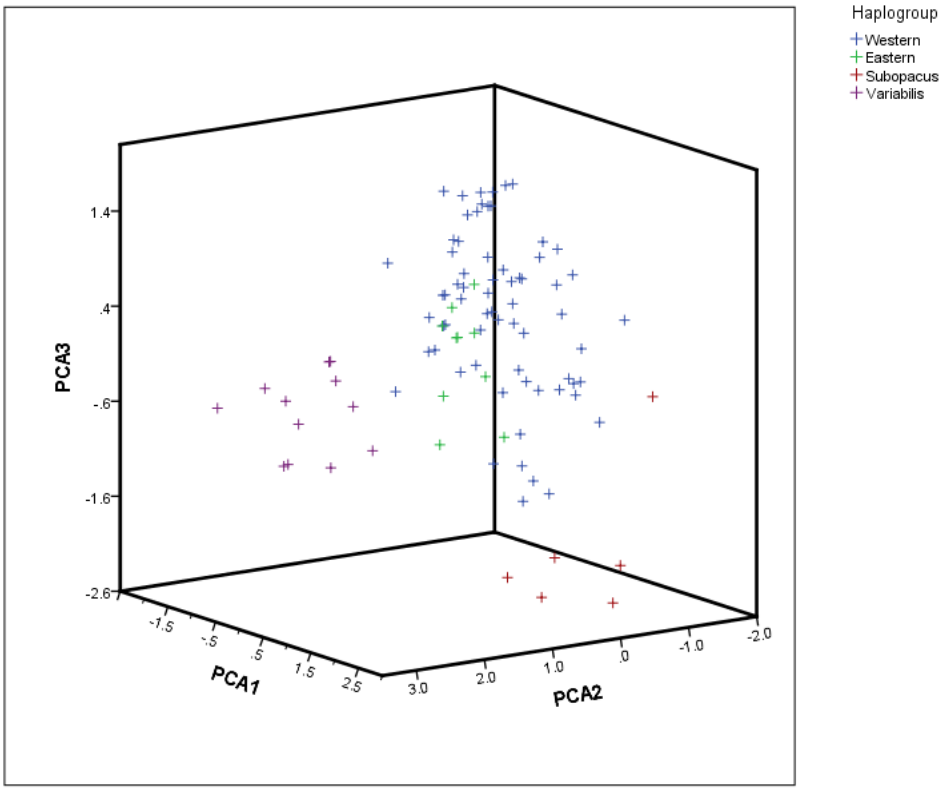


Figure 4.7: *Gnorimus nobilis* distribution showing variation in male metatibial shape.

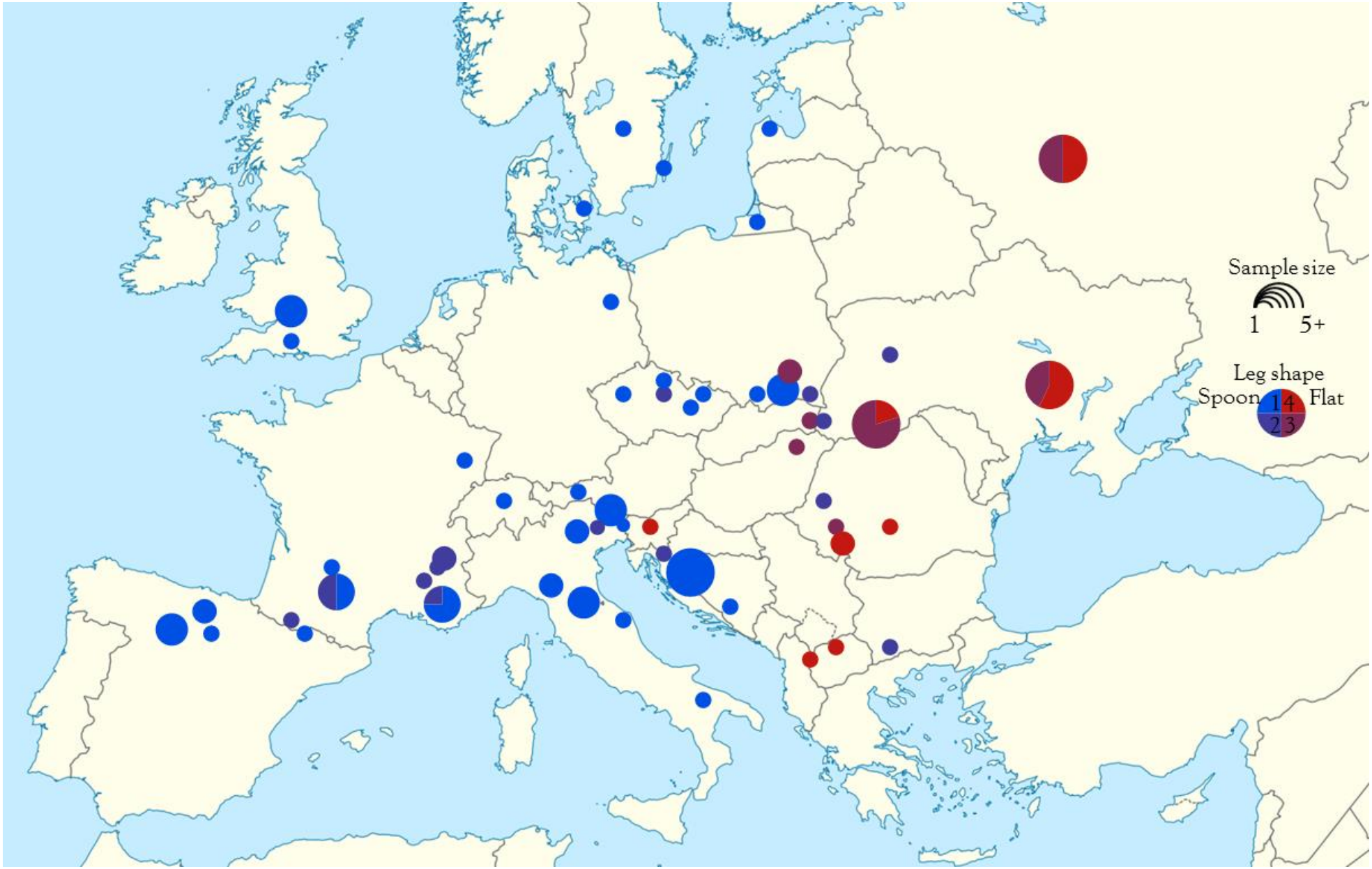


Figure 4.8. Plot of canonical variate 1 against canonical variate 2 for the *Gnorimus nobilis* against *G. variabilis* dorsal dataset. Individuals were also split by sex. The confidence ellipses show 95% probability from equal frequency

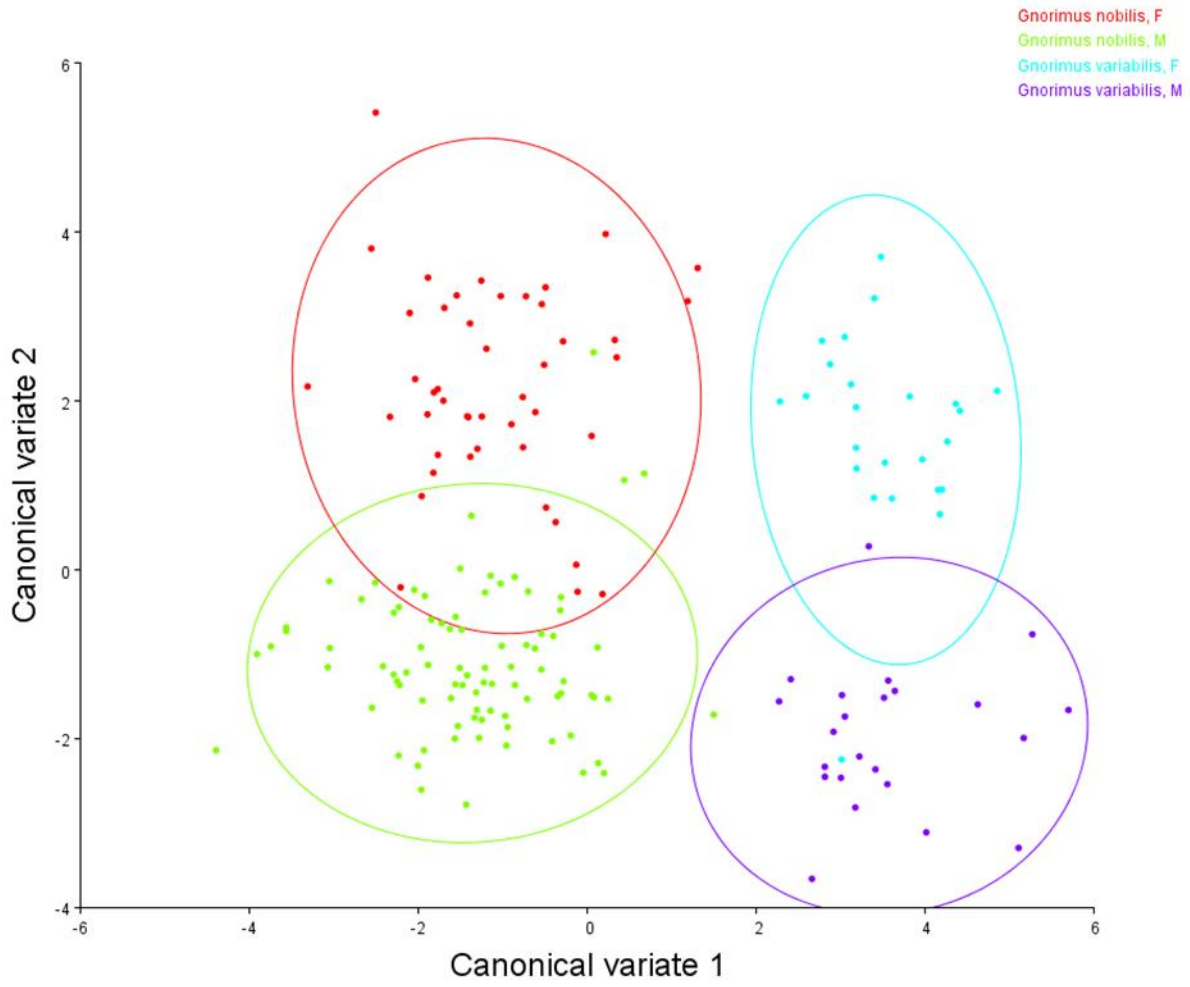
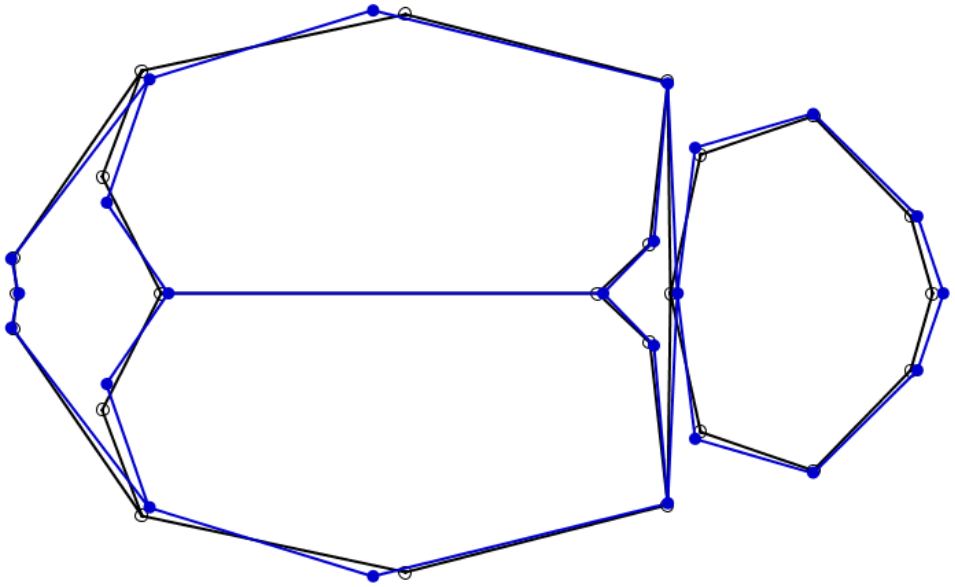
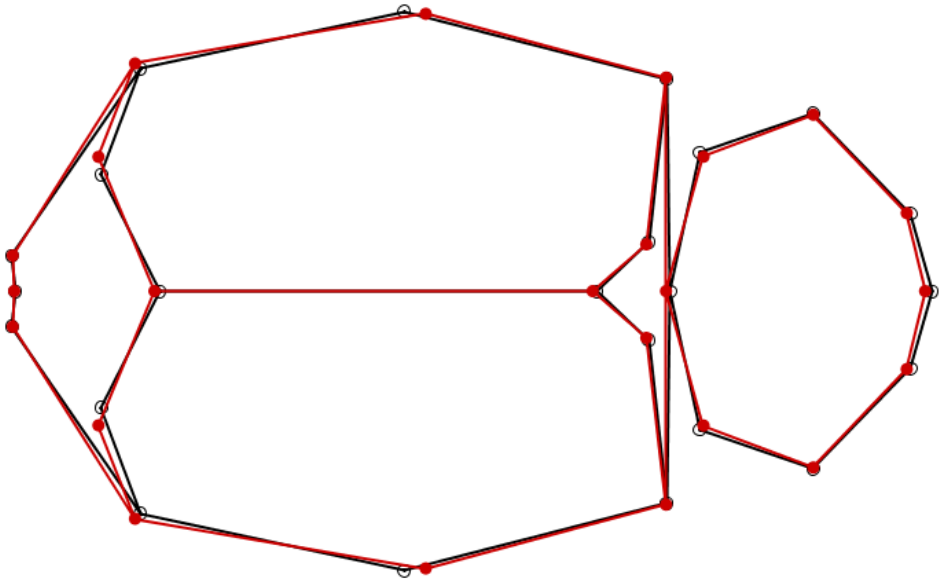


Figure 4.9. Positive (blue) and negative (red) shifts in canonical variate 2 for the dorsal species level analysis. These represent the most extreme values from the data for each variate, showing how the scores vary.

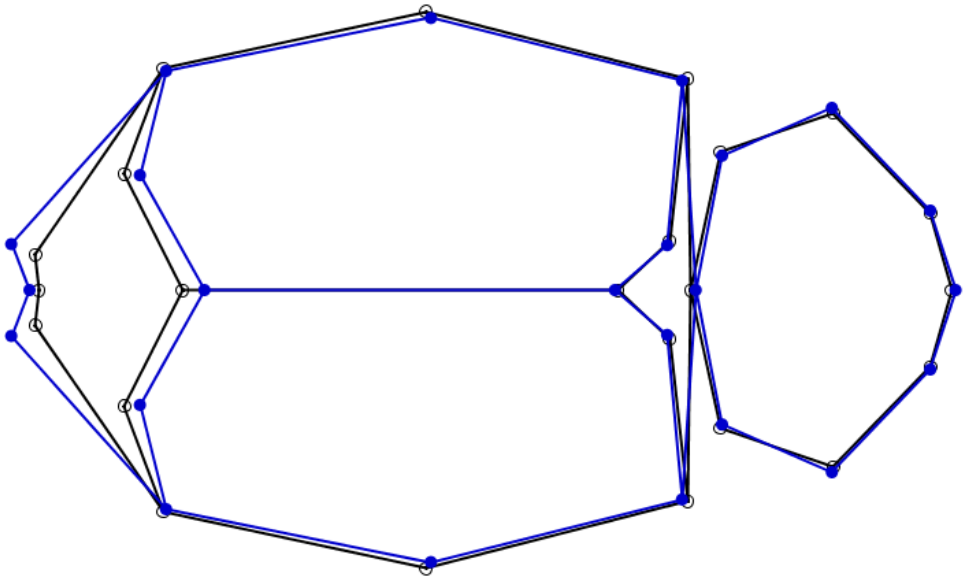


CV1

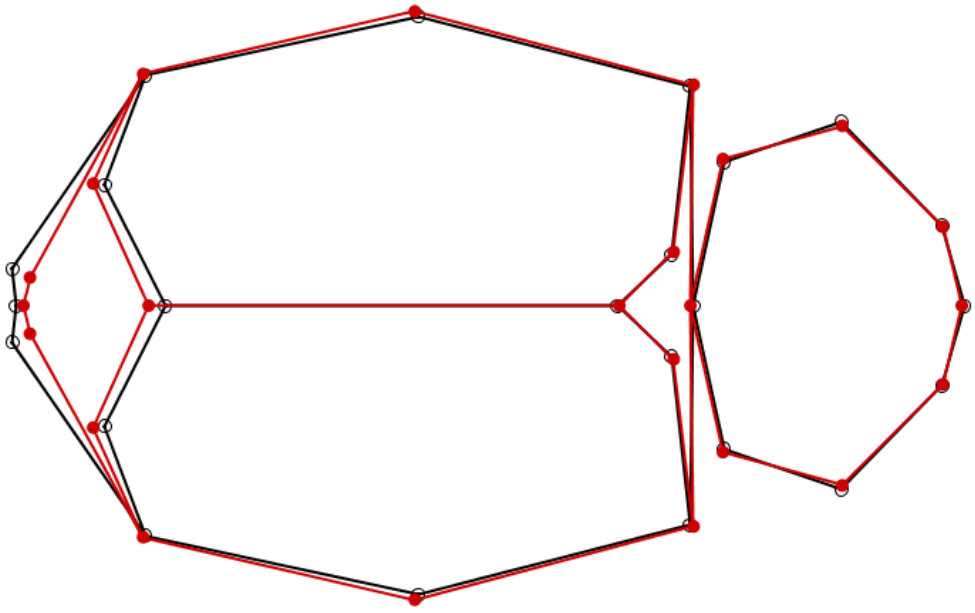


CV1

Figure 4.10. Positive (blue) and negative (red) shifts in canonical variate 2 for the dorsal species-level analysis. These represent the most extreme values from the data for each variate, showing how the scores vary.



CV2



CV2

Figure 4.11. Plot of canonical variate 1 against canonical variate 2 for the *Gnorimus nobilis* against *G. variabilis* ventral dataset. Individuals were also split by sex. The confidence ellipses show 95% probability from equal frequency

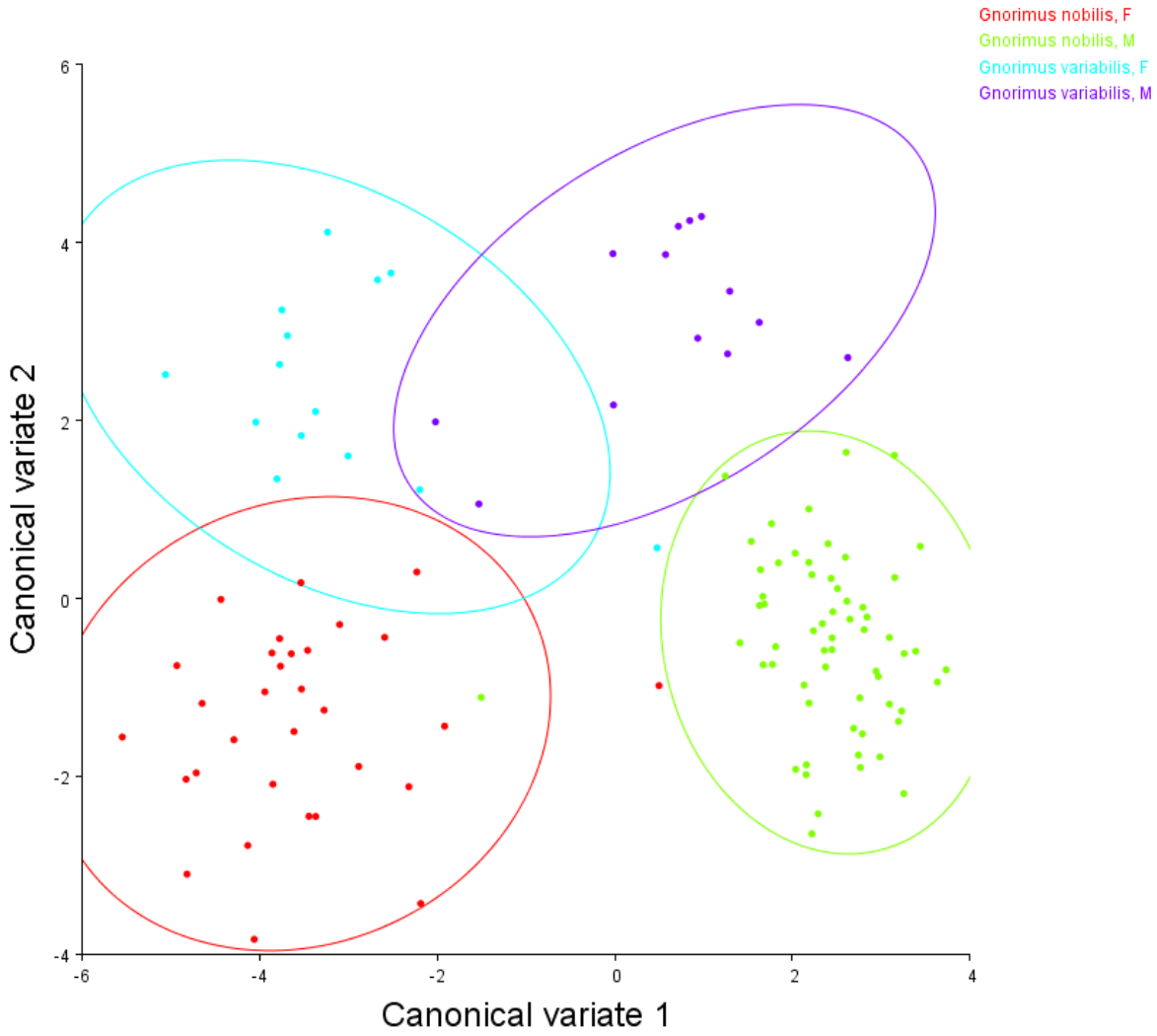
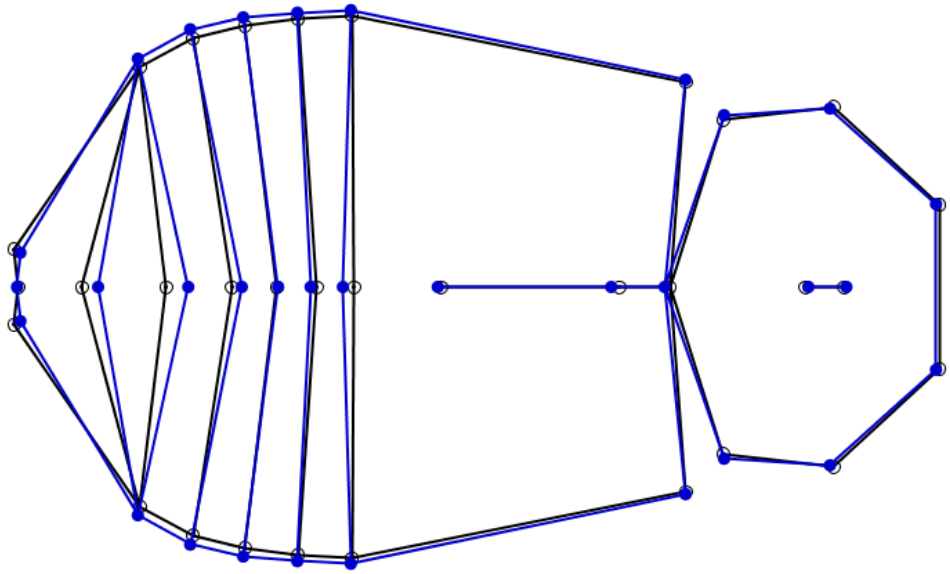
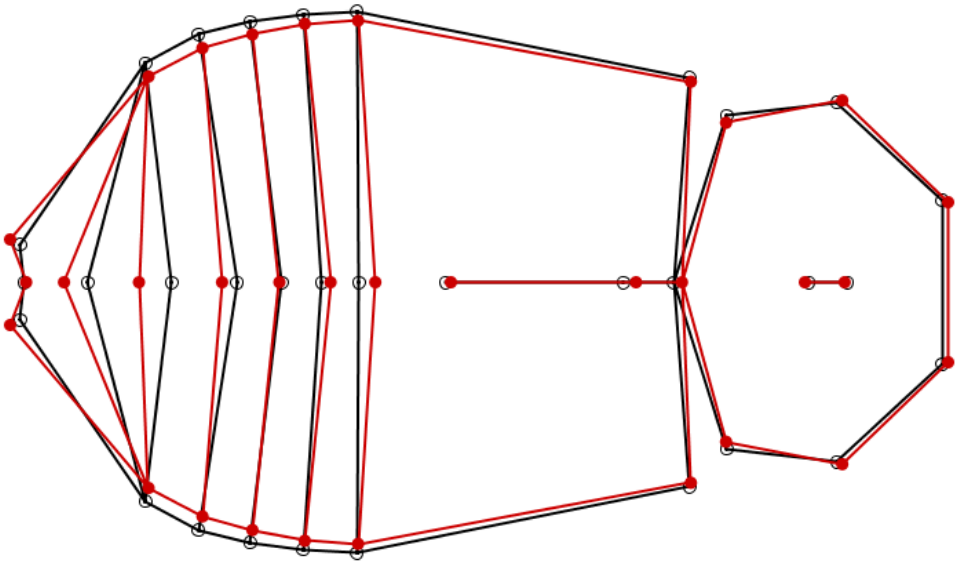


Figure 4.12. Positive (blue) and negative (red) shifts in canonical variate 2 for the ventral species-level analysis. These represent the most extreme values from the data for each variate, showing how the scores vary.

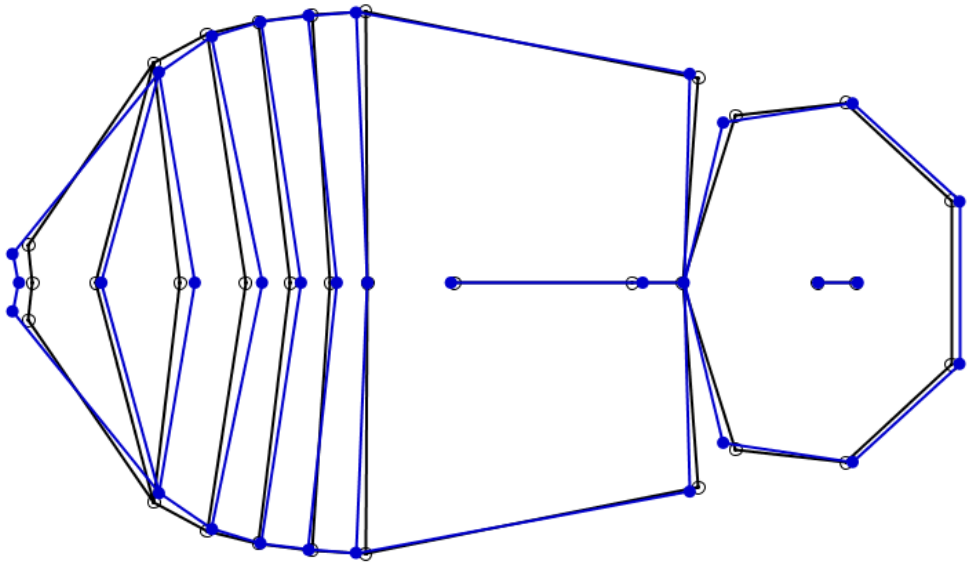


CV1

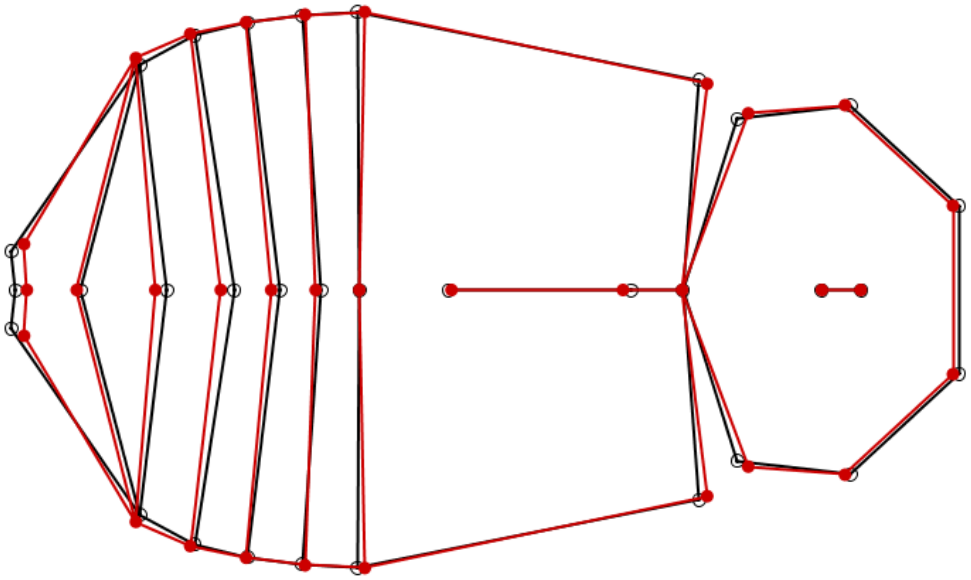


CV1

Figure 4.13. Positive (blue) and negative (red) shifts in canonical variate 2 for the ventral species-level analysis. These represent the most extreme values from the data for each variate, showing how the scores vary.



CV2



CV2

Figure 4.14. Plot of canonical variate 1 against canonical variate 2 for the *Gnorimus nobilis* subspecies dorsal dataset. Individuals were also split by sex. The confidence ellipses show 95% probability from equal frequency

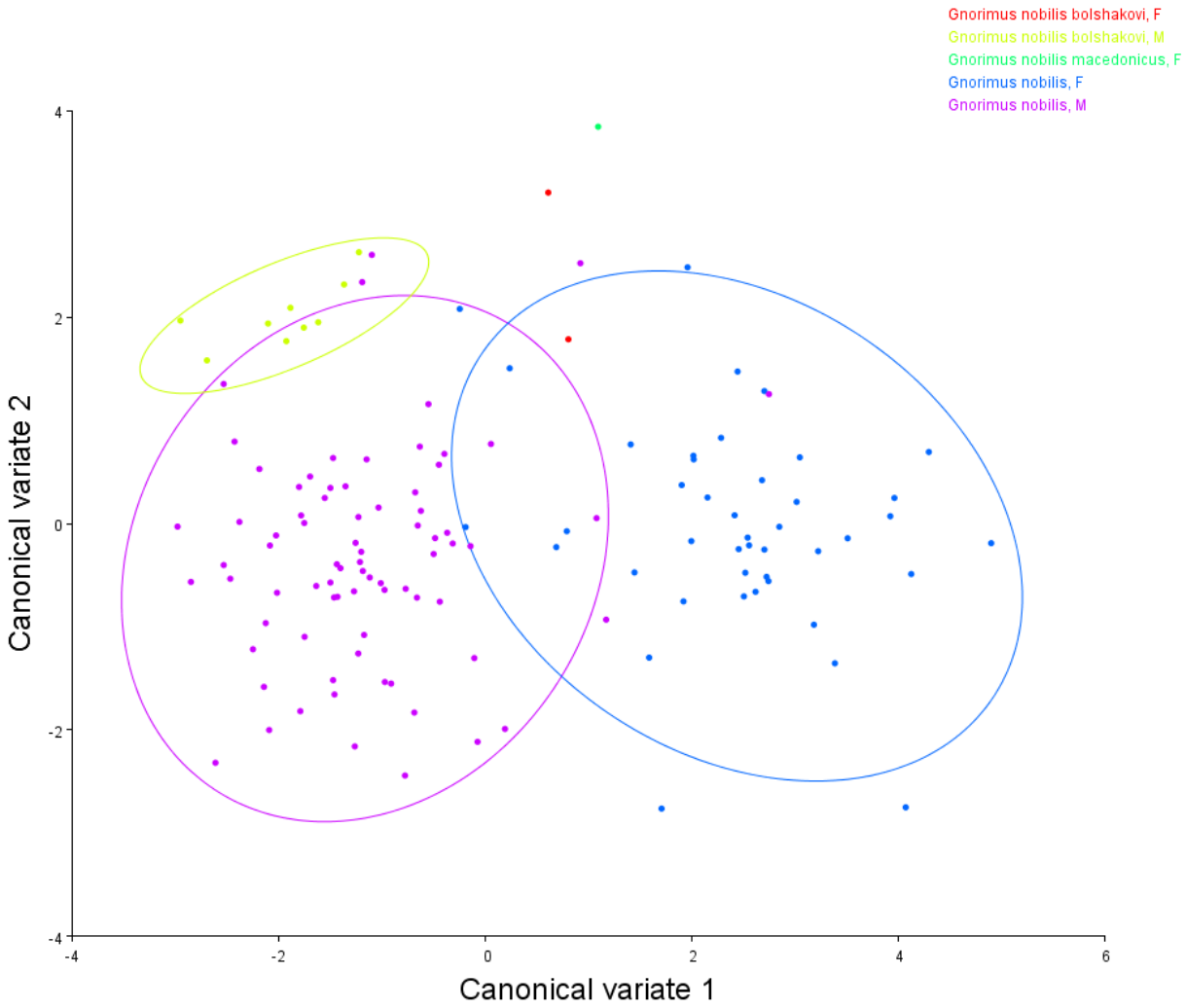
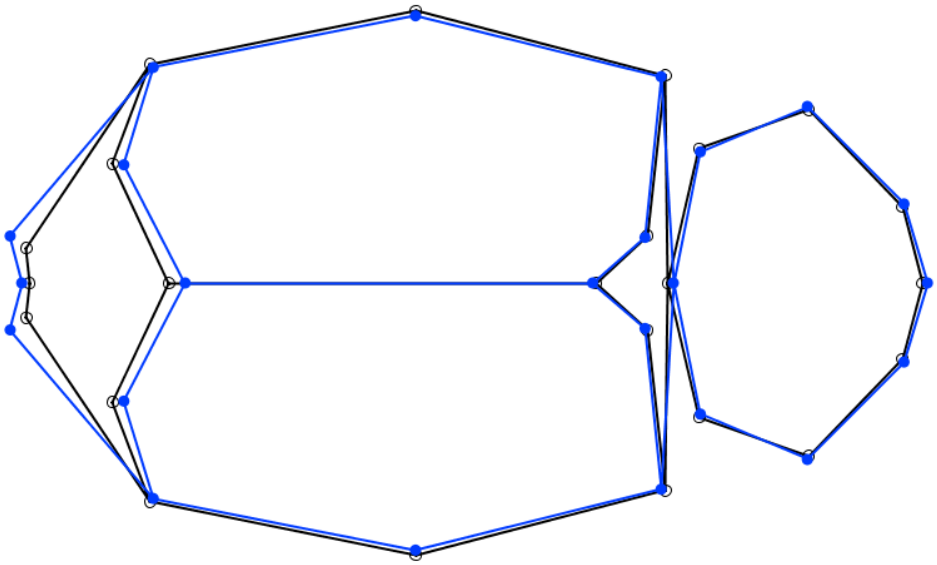
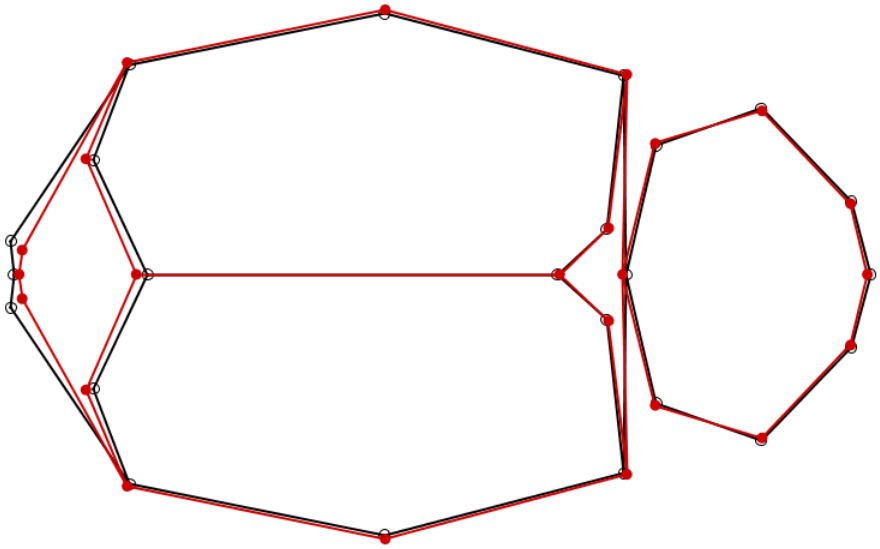


Figure 4.15. Positive (blue) and negative (red) shifts in canonical variate 1 for the dorsal subspecies analysis. These represent the most extreme values from the data for each variate, showing how the scores vary.

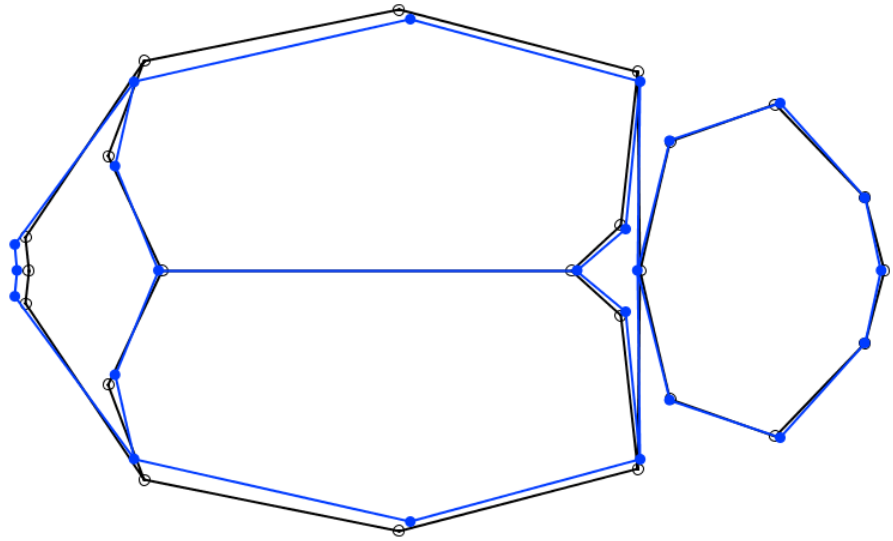


CV1

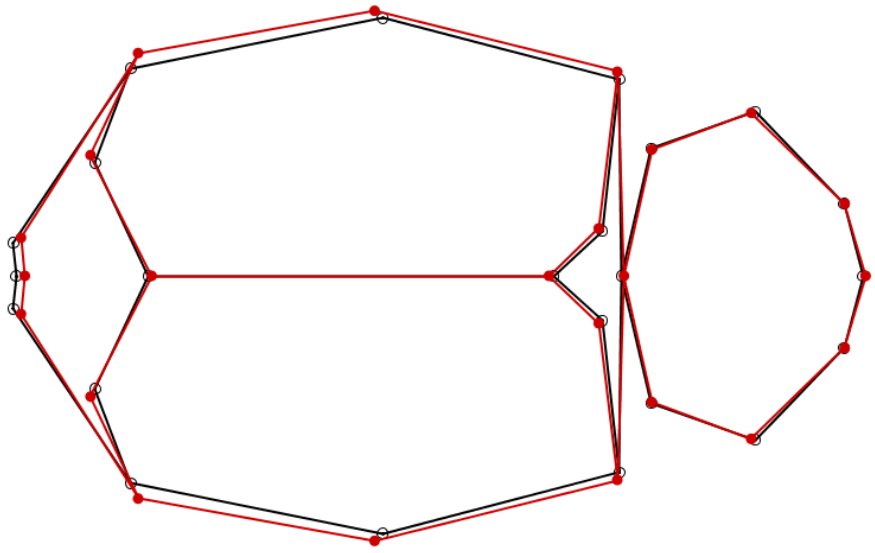


CV1

Figure 4.16. Positive (blue) and negative (red) shifts in canonical variate 2 for the dorsal subspecies analysis. These represent the most extreme values for each variate, showing how the scores vary.



CV2



CV2

Figure 4.17. Plot of canonical variate 1 against canonical variate 2 for the *Gnorimus nobilis* subspecies ventral dataset. Individuals were also split by sex. The confidence ellipses show 95% probability from equal frequency

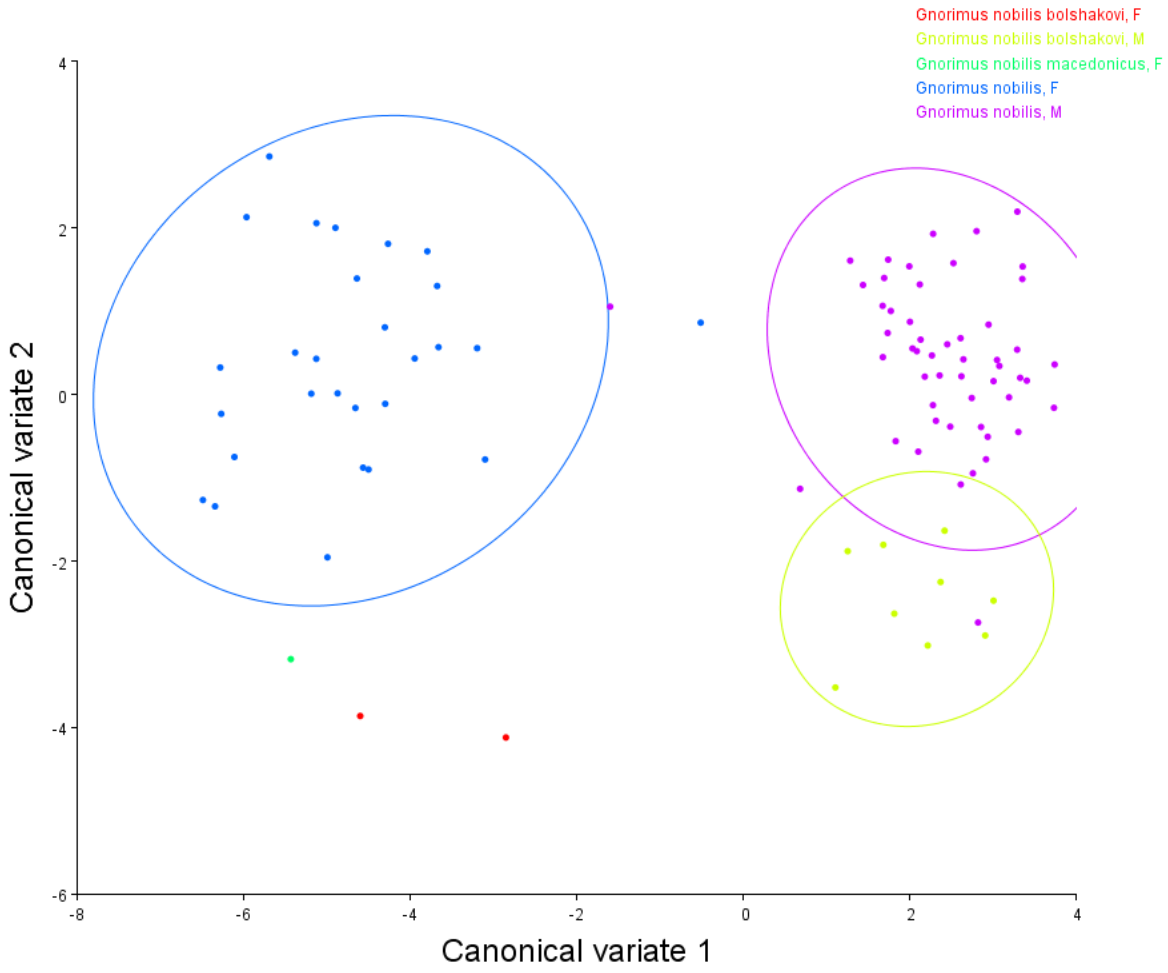
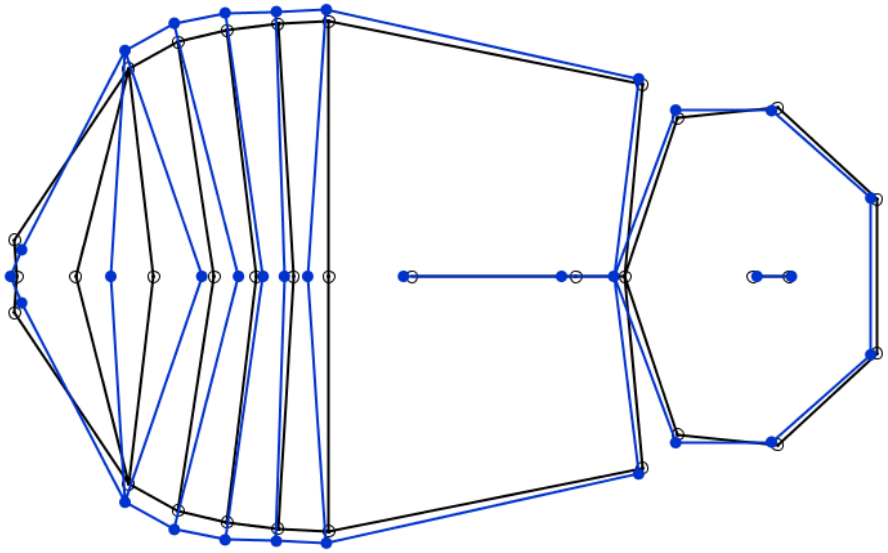
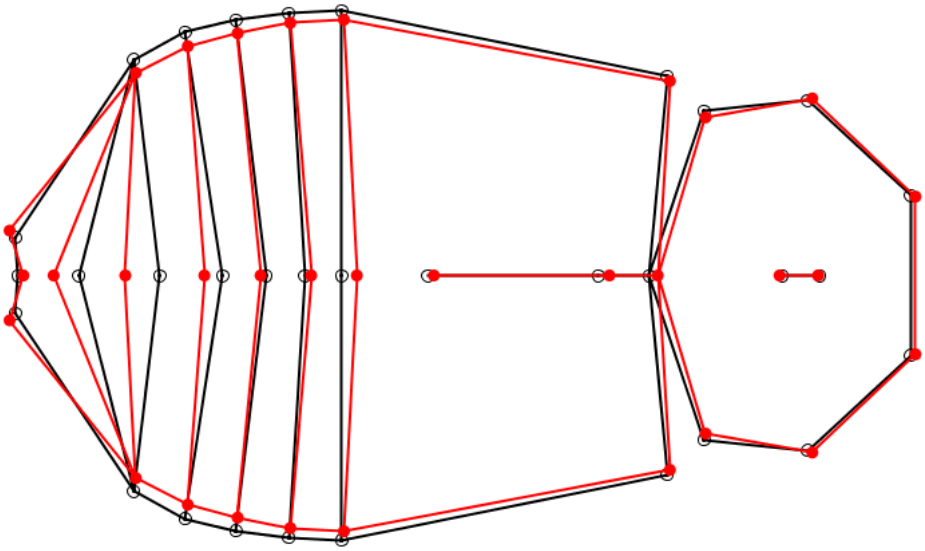


Figure 4.18. Positive (blue) and negative (red) shifts in canonical variate 1 for the ventral subspecies analysis. These represent the most extreme values from the data for each variate, showing how the scores vary.

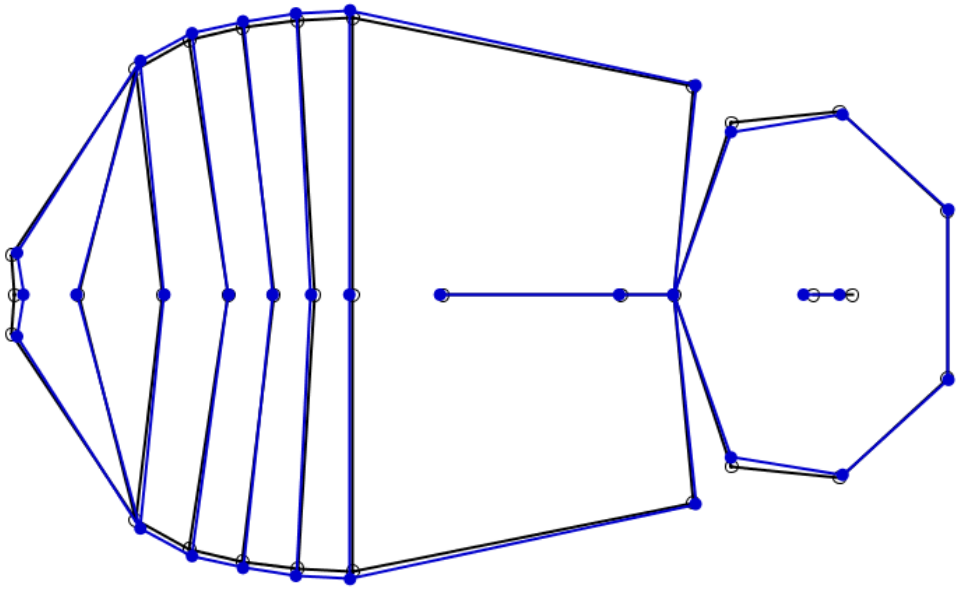


CV1

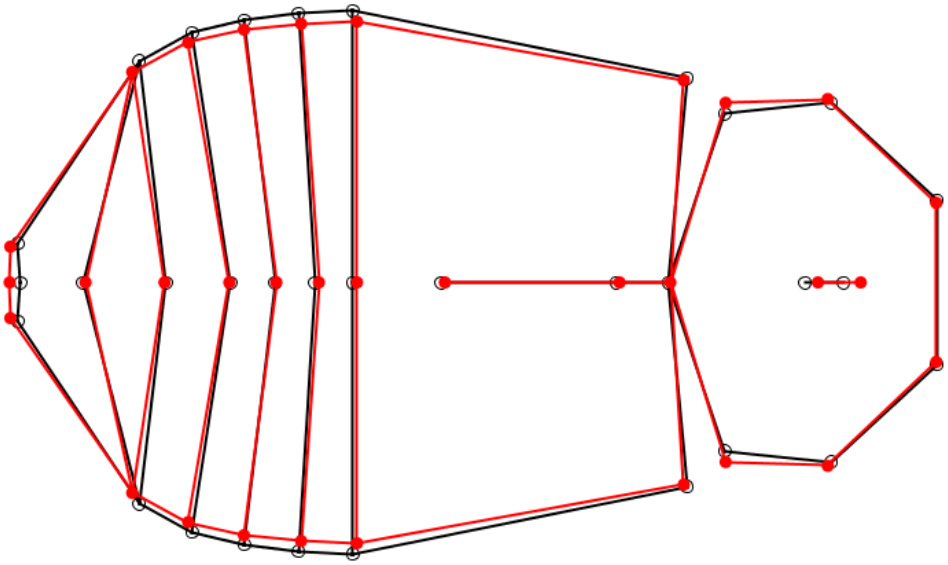


CV1

Figure 4.19. Positive (blue) and negative (red) shifts in canonical variate 2 for the ventral subspecies analysis. These represent the most extreme values from the data for each variate, showing how the scores vary.



CV2



CV2

Chapter 5

Systematics and phylogeography of *Gnorimus* in Europe

5.1 Introduction

Gnorimus is a small genus of 11 species distributed throughout the northern hemisphere (Tauzin 2004a), though only the two most widespread species in Europe (*G. nobilis* (L. 1758) and *G. variabilis* (L. 1758)) have been the recipients of much research and conservation attention (Schenke 2010; Trizzino et al. 2013; Bates et al. 2014). These two species are both specialists feeding on heartwood rot of veteran trees with some overlap in their chosen hosts, oaks (*Quercus* spp.), sweet chestnuts (*Castanea sativa*) and fruit trees (*Prunus* spp., *Malus* spp.) in particular, although *G. nobilis* is also found in willow (*Salix* spp.) and poplar (*Populus* spp.) and lives at higher altitudes in the southern part of its range (Mannerkoski et al. 2010a; Trizzino et al. 2013). *Gnorimus variabilis* prefers lower altitudes in the southern reaches of its range, and will also use beech (*Fagus sylvatica*) and alder (*Alnus* spp.) as hosts (Mannerkoski et al. 2010b; Trizzino et al. 2013).

Though widespread, the two species are poorly documented from a phylogeographical standpoint. Both species have a large number of named colour variations (Tauzin 2004b) which almost certainly represent continuous variation (see Chapter 4), but *G. nobilis* has three recognised subspecies: *G. nobilis nobilis* found throughout Europe, *G. nobilis bolshakovi* (Gusakov 2002) from one locality in eastern Russia, and *G. nobilis macedonicus* (Baraud 1992) from Macedonia. It is worth noting that the range of *G. n. macedonicus* is surrounded by that of *G. n. nobilis*. The relationships between these three subspecies are unknown: both new subspecies were named on the basis of colouration (see Chapter 4) and lack the spoon-shaped metatibia typical of male *G. n. nobilis*. Tauzin (2004a) suggested the two minor subspecies be synonymised due to their morphological similarity, although no comment was made on the disjunct distribution or biogeographic separation between Russia and Macedonia for the definition of *G. n. macedonicus*.

Little is known about the changing distributions and phylogeography of saproxylic beetles, despite these species being under serious pressure from habitat loss and degradation (Horák et al. 2012). Rot holes (also called tree cavities) are the rarest saproxylic habitat in Europe (Müller et al. 2014), taking decades or centuries to form. Once rot holes have formed they can persist for as long as the tree stays standing (up to many tens of years), thus acting as reservoirs in the absence of any suitable holes in surrounding trees. Though most large European tree species hosting saproxylic beetles conform to a ‘standard’ “southern richness, northern purity” model of phylogeography (Petit et al. 2002; Hewitt 2004; Magri et al. 2006),

evidence strongly suggests that some tree species persisted in northern European refugia during the last glacial maximum (Provan & Bennett 2008; Schmitt & Varga 2012). The extent to which saproxylics utilised such habitats during the last glacial maximum is unknown (though some evidence for northern rot-hole refugia exists - Coope 1998), as is the extent to which the post-glacial recolonisation pattern of rot-hole specialist beetles across Europe matches that of their tree hosts.

To analyse patterns of postglacial recolonisation and genetic predictions of past distributions, plus the genetic status of the named subspecies and potential presence of other cryptic subspecies, of European species of *Gnorimus* a suite of genetic markers were chosen for their suitability and reliability to address these questions. Mitochondrial genes have been widely utilised in phylogeographic studies of insects (e.g. Hewitt 2004; Habel et al. 2005; Garrick et al. 2006; Ahrens et al. 2013), though such genes need to be complemented by studies on nuclear genes to better understand potential hybridisation / introgression and to avoid reading too much into results from a single gene (Rand & Harrison 1989; Knowles & Maddison 2002; Shaw 2002; Flanders et al. 2009). Combinations of mtDNA and nDNA markers can provide extremely detailed pictures of the pattern of phylogeography, phylogeny and post-glacial expansion in the study species (e.g. Holderegger et al. 2006; Vila et al. 2006; Říčan et al. 2008). Additionally, markers under selection can also provide insights when combined with selectively neutral markers.

Cytochrome Oxidase 1 (COI) is the most commonly utilised mitochondrial DNA marker in phylogeography (Moore 1995; Hebert et al. 2003), and as such has a vast quantity of literature associated with its analysis. Studies typically either use primers designed from conserved regions in the gene to amplify a large section for analysis (Simon et al. 1994), or develop species- or genus-specific primers using sequences on publically accessible databases such as GenBank. The COI gene has a reported mutation rate between 3.54%My^{-1} (Papadopoulou et al. 2010) and 2.34%My^{-1} (Brower 1994) in insects, allowing for analysis of population divergence events down to a scale of approximately 5KYA using typical sequence lengths (250bp+).

Control Region (CR, also known as the AT-rich region in insects) is a large non-coding portion of the mitochondrial genome which controls the initiation of replication and transcription of mtDNA (Saito et al. 2005).The CR is extraordinarily long and AT rich in insects (usually ~90% AT; Zhang & Hewitt 1997), varying in size in Cetoniidae from 761bp

in *Osmoderma opicum* (Kim et al. 2016) to 5654bp in *Protaetia brevitarsis* (Kim et al. 2014), the latter being one of the longest Control Regions of any sequenced beetle. This extended length of CR is often composed of large (100bp+) tandem repeats (e.g. in *P. brevitarsis*) or shorter microsatellite-like sequences (e.g. in *O. opicum*). For the few species for which population-level data are available, CR can be highly variable in size due to variable numbers of repeat units (Mancini et al. 2008), and therefore represents an extremely challenging DNA region to amplify reliably within Coleoptera (and arthropods in general), especially those with very large CRs. For those species where a section of CR can be reliably amplified, the gene shows a high level of nucleotide variation, making it suitable for use in phylogeography (Mardulyn et al. 2003), but in general the region often proves to be so difficult to amplify reliably in arthropods that little population-level data is available (e.g. Snäll et al. 2002; Mancini et al. 2008; Chen et al. 2012). Additionally, there has been very little work on arthropods to characterise genetic distances between lineages (e.g. species) and the rate of evolution in CR. Most studies have used microsatellite-like repeat units within CR rather than nucleotide differences, an approach suited to species with small, simple CRs, but one that becomes near impossible when the repeat units are long (30bp+) or the region highly complex (Hwang & Kim 1999; Navajas et al. 2002; Ma et al. 2009). One of the few studies to have computed CR divergence levels showed pairwise uncorrected distances among *Halocaridina* shrimp were approximately 9.9-18% per million years (2-3 times higher than other regions of the mitochondrial genome) (Justice et al. 2016).

In addition to mtDNA markers, neutral nuclear markers are often employed to analyse population connectivity, speciation, hybridisation, and assessment of phylogeographic patterns (Roy et al. 1994; Barluenga et al. 2006; Palstra & Ruzzante 2008). The most commonly employed markers are microsatellites, short simple sequences composed of tandemly repeated motifs of non-coding DNA, which are popular due to their high mutation rate and common occurrence in genomes (Sunnucks 2000). Though homoplasy, null alleles and allele dropout caused by high mutation rates can make microsatellites challenging to study (Hendrick 1999; Brito & Edwards 2009), they offer considerable power to investigate fine scale processes, such as dispersal and subpopulation structure (Paetkau et al. 1995; Keller et al. 2005).

Evidence of adaptation in the genome is being increasingly targeted in studies of conservation genetics and phylogeography (Pauls et al. 2013; Frankham et al. 2014), typically being studied by investigating sequence variation within functional genes and

comparing the patterns to those from neutrally evolving markers (though see Ballard & Whitlock 2004). Heat Shock Protein 70 (HSP70 hereafter) mediates cellular protection under environmental stress via interfering with apoptosis (Beere et al. 2000; Li & Srivastava 2004). Though the utility of heat shock genes for investigating phylogeographic pattern has been little investigated, the work that has been done suggests that they are useful markers with which to discriminate between lineages along with other nuclear and/or mitochondrial markers (e.g. Lee & Boulding 2009 on the related HSC70; Hou et al. 2014 using HSP70; Baringou et al. 2016 using HSP70).

The *Orco* (Odorant co-receptor, previously known as Or83b, Vosshall & Hansson 2011) gene region codes for an atypical receptor which is co-expressed with other odorant receptors in *Drosophila melanogaster*, but is also highly conserved across other insect orders (Larsson et al. 2004; Jones et al. 2005; Malpel et al. 2008). *Orco* binds to a specific odorant receptor, then forms a heterodimer which acts as an ion channel (Jones et al. 2011). Along with this general function, it appears to be an important receptor for detecting host cues (the gene is expressed to a greater degree in female emerald ash borer beetles (*Agrilus planipennis*), which need to find rot holes to lay their eggs, than males (Mamidala et al. 2013)) and/or male-produced pheromones (e.g. *Bactrocera cucurbitae*, Diptera; Shen et al. 2011). The gene has been used in phylogeny reconstruction due to its conserved nature (e.g. Yang et al. (2012) in Orthoptera and Endopterygota; Zhao et al. (2013) in Hymenoptera; Macharia et al. (2016) in Diptera), however despite the gene being useful to reconstruct phylogenies across orders, families and genera, no study has investigated this gene for its use in phylogeographic studies within genera.

The objectives of this study were to: (i) analyse genetic diversity in the two common European species of *Gnorimus* (*G. nobilis* and *G. variabilis*), and to investigate their relationships to other *Gnorimus*; (ii) to resolve the systematic status of both *G. n. bolshakovi* and *G. n. macedonicus* subspecies of *G. nobilis*; and (iii) to investigate the post-glacial recolonisation patterns and historical demographic changes of both species. An earlier objective was to extend the Europe-wide phylogeographic analysis to investigate smaller scale connectivity between *Gnorimus* populations in UK orchards using microsatellite loci. Due to technical difficulties in isolating microsatellites and extremely low levels of variation at the loci developed (see Chapter 6) led to this objective being dropped and replaced with a comparative study of population genetics of UK populations of another saproxylic beetle, the stag beetle *Lucanus cervus*.

5.2 Methods

5.2.1 Sampling Protocol

A total of 131 specimens of adult *Gnorimus nobilis* were obtained for this study (Supplementary Table 7), plus four larvae from Kent, UK. Due to the elusive nature and low abundance status of species of the genus large numbers of samples across the full range of each species are difficult to gather (and not ethically justifiable), so small numbers of existing beetle samples were collected from a variety of sources including museums, commercial insect suppliers, scientists and private sellers on eBay (only reliable sources used, with collection locality verified). Due to the focus on *G. nobilis*, samples were collected from throughout its range representing all three named subspecies and from nearly every potential southern glacial refugial location. The only putative refugium that could not be sampled for this study was Turkey. In addition, samples of *G. variabilis* (n=17) and *G. subopacus* (n=2) were included as outgroups for comparison. No individuals were sampled from areas where they are legally protected. Where extra fresh samples were collected in the UK, this was done in collaboration with local conservation groups and the People's Trust for Endangered Species with landowner permission obtained where necessary. Samples were stored on arrival to Aberystwyth University either in 100% ethanol or individually bagged and dry frozen at -20°C.

5.2.2 DNA Extraction Protocol

DNA was extracted from all samples using a modified CTAB phenol-chloroform-isoamyl alcohol protocol (Winnepenninckx et al. 1993) (see Chapter 6): tissue was initially taken from a meta- or mesothoracic leg up to the trochanter from preserved adults. For delicate specimens or those on loan from an institution only the tarsi and tibia were used. For specimens which yielded very little DNA via this method, a separate leg was used up to and including the coxa, or cutting into the thorax and using preserved flight muscles instead. The tissue was washed in distilled water, dried by blotting on lab roll, and placed into a 1.5ml Eppendorf tube. 350µl of CTAB buffer and 10µl of Proteinase K was added, followed by 10s on a vortex. This was incubated overnight (18-22 hours) at 37°C. 350µl of equilibrated 6.7/8.0pH Phenol-Chloroform-Isoamylalcohol was added into each tube before being shaken

by hand for 10 minutes. The tubes were then centrifuged for 10 minutes at 13,000RPM, after which the top layer was pipetted into a new tube, and the waste discarded. 990µl of 100% ethanol was added to the new tube, and the solution was incubated at -20°C for at least two hours. This was then centrifuged at 13,000RPM for 10 minutes; the liquid solution poured off, with small remaining volumes removed using a pipette. DNA pellets were then dried in open tubes in a fume cupboard for 20 minutes, after which 50µl of distilled water was added. DNA was then left overnight at 4°C to go into solution before assessment by running 5µl DNA stock on 1% agarose gels.

5.2.3 mtDNA genotyping and statistical analysis

Depending on the concentration of DNA visible on a gel, stock DNA was diluted to between 1/10 and 1/100. Most samples were diluted to either 1/20 or 1/50. Species-specific primers (Table 5.1) were designed from a sequence from *Gnorimus variabilis* on GenBank (DQ155821.1) to amplify the 5' end of COI, one of the most variable regions of COI and useful for phylogeographic study and species identification (Lunt et al. 1996). These primers were used to amplify and sequence this region from *G. nobilis*, from which internal species-specific primers were designed to amplify a 512bp region. Using these primers, all specimens were screened using 20µl PCRs consisting of 10µl of Biomix (Bioline, London, UK), 1µl of each primer at 10µM, 5µl of dH_2O, and 3µl of diluted DNA. PCR conditions were 95°C/3 minutes, 55X (95°C/30s, 52°C/45s, 72°C/45s), 72°C/3 minutes. For specimens that didn't amplify (most notably non-*G. nobilis Gnorimus*) the annealing temperature was reduced to 48°C. All primers novel to this study were designed in Primer 3 (Koressaar & Remm 2007).

For Control Region, primers were initially developed from recommendations by Simon et al. (2006), sequencing in from the 12s and ND2 regions. Primers used were the 12s primers SR-J14197 (5'- ATAAGYCTACTTTGTTACGACTT-3') and SR-J14610(5'-ATAATAGGGTATCTAATCCTAGT-3'), and the ND2 primer N2-N993 (5'-GGTAAAAATCCTAAAAATGGNGG-3'). Modifications to the primers were made following suggestions from Simon et al. (2006) to as closely match Coleoptera as possible. Amplicons approximately 6-7kb in size were initially produced using the same PCR mix as above, but under the following PCR conditions: 95°C/3 minutes, 55X (95°C/30s, 55°C/60s, 70°C/90s), 72°C/5 minutes. Additional tests using lower annealing and extension temperatures proved unreliable, despite Su et al. (1996) suggesting that this might help with

A+T rich DNA templates. The amplicons were cleaned and sequenced with both forward and reverse primers. These provided approximately 600bp from 12s, and 400bp from Nad2, from which internal primers were designed. These were used to continue to sequence along the fragment, developing new primers when possible. However, this proved to be extremely difficult due to the repetitive nature of the CR (see below). Primers developed often turned out to be from regions replicated in the centre of the CR, and would regularly produce a mixture of bands of different sizes with a reverse primer corresponding to replicated forward primer sites. This, along with the extremely large size of the region and its extremely AT rich nature (generally between 80-95%) resulted in dozens of forward and reverse primers being developed and utilised to eventually find a short region (Table 5.1; 350bp) which reliably amplified across all samples, including museum specimens, and was free of mixed (overlapping) peaks. The final PCR conditions for these primers were: 95°C/3 minutes, 55X (95°C/30s, 54°C/30s, 72°C/30s), 72°C/5 minutes.

Amplicons were checked on a 2% agarose gel, then cleaned with SureClean Plus (Bioline) following the manufacturer's protocol (but increasing the initial centrifugation step to 20 minutes), and sequenced with AB BigDye technology on an AB3500.

Sequences were checked and edited in Chromas Lite (Version 2.1; 2012; Technelysium Pty Ltd), checked for identity using the BLAST algorithm (Altschul et al. 1990), then aligned in BioEdit (Version 7.1.11; 2013; Hall 1999) using the CLUSTAL W algorithm with default settings (Thompson *et al*. 1994). Sequences of COI from *Gnorimus nobilis* were aligned to a sequence of *G. nobilis* accessioned to GenBank (JX234208.1) from southern Greece (N39°52'E22°44', Sipek *pers. comm*. 2015); no sequences to *Gnorimus* CR were available on GenBank, so *de novo* alignments were created. All sequences were carefully checked by eye for double peaks, then checked in the 'ORF Finder' (http://www.bioinformatics.org/sms/orf_find.html) for indels, frameshifts and stop codons which could indicate the presence of nuclear mitochondrial pseudogenes (NUMTs; Song et al. 2008).

Intra-population estimates of genetic diversity were calculated for both mtDNA regions: number of haplotypes (*H*), haplotype (h) and nucleotide (π) diversities, and the number of polymorphic sites (*P*) were calculated in ARLEQUIN (3.5.2.2 2015, Excoffier & Lischer 2010). Additionally, statistics of neutral sequence evolution and of population expansion were computed under a sudden spatial expansion model: Tau, Theta, Harpending's

Raggedness index (Harpending et al. 1993), Tajima's *D* (Tajima 1989), Fu's *F* (Fu & Li 1993) and Mismatch Distribution analysis (Rogers & Harpending 1992) were computed in ARLEQUIN. Simulated results of pairwise nucleotide differences can be compared to the observed distribution by using the sum of squared deviations (SSD), which was also computed in ARLEQUIN. Values of Tau were used to estimate population expansion times under 2.34%MY^{-1} and 3.54%MY^{-1} COI mutation rates (Brower 1994; Papadopoulou et al. 2010). In addition, the mean diversity within subpopulations was calculated alongside the mean diversity of all populations for COI.

The most appropriate model for the analyses was chosen using jModelTest (V. 2.1.10; Guindon & Gascuel 2003; Darriba et al. 2012) for each gene using the model with the lowest value for the Akaike Information Criterion (AIC) statistic. The Tamura 3-parameter substitution model including a gamma-distributed amongst site variation was shown to be the most appropriate method for COI (P=61, AIC=2175.01), whilst the General Time Reversible model including a gamma-distributed amongst site variation was the most appropriate method for CR (P=55, AIC=1371.92). Both Maximum Likelihood phylogenies were built using an extensive subtree-pruning-regrafting (SPR) tree inference method, and 1000 bootstrap replications.

Neighbour Joining (1000 bootstrap replications) trees were built for both genes using MEGA6 (6.06; 2013; Tamura et al. 2013), using models with the lowest AIC statistic in jModelTest where appropriate within MEGA6 (Tamura 3-parameter for both COI and CR). A Bayesian approach using MrBayes (v3.2.6; Huelsenbeck & Ronquist 2001; Ronquist & Huelsenbeck 2003) (assuming an inverse gamma distribution of site variations, 100,000 generation burn in and 1,000,000 total generations) was also used to explore phylogenetic relationships within the COI and CR datasets. A sequence each from *Trichius fasciatus* (GenBank accession JX234253.1) and *T. zonatus* (GenBank accession EF487734.1) were used as Trichiini outgroups.

Haplotype networks were built using DNASP to produce an .rdf file, which was then visualised in Network (5; 2015; fluxus-engineering.com) using Median Joining networks (Bandelt *et al*. 1999) and Maximum Parsimony networks (Polzin & Daneshmand 2003). Inter-population tests of genetic diversity were calculated. Estimates of genetic differentiation among lineages were calculated using pairwise F_{ST} in ARLEQUIN.

5.2.5 Nuclear gene sequencing and analysis

Stock extracted DNA was diluted to between 1/5 and 1/100 for nuclear gene amplification. For Heat Shock Protein 70, the Clarke HSP primers were used to amplify a large portion of this gene from *Gnorimus nobilis*, before species-specific primers (Table 5.1) were developed from within this larger sequence. PCR cycling protocols were: 95°C/3 minutes, 45X (95°C/30s, 52°C/30s, 72°C/30s), 72°C/5 minutes.

For *Orco*, primers were developed from an alignment of three *Holotrichia* (Scarabaeidae, Melolonthinae) species uploaded to GenBank (*H. plumbea*, HQ110087.1; *H. oblita*, JF718662.1; *H. parallela*, JF826514.1). The alignment was 1428bp long, and showed some highly conserved sites between the three *Holotrichia* species and the non-Scarab Coleoptera outgroups *Tenebrio molitor* (KP296755.1) and *Tribolium castaneum* (AM689918.1). Primers (Table 5.1) were designed from the most conserved regions in this alignment, and were then used to amplify and obtain a sequence of the gene from *G. nobilis*. This was aligned to the other *Orco/Or83b* sequences, and shorter primers were developed to more reliably amplify the gene from a variety of samples. PCR conditions for the short, species-specific primers were: 95°C/3 minutes, 45X (95°C/30s, 57°C/30s, 72°C/30s), 72°C/5 minutes.

Amplicons were checked on a 2% agarose gel, then cleaned with SureClean Plus (Bioline) following the manufacturer's protocol (but increasing the initial centrifugation step to 20 minutes), and sequenced with AB BigDye technology.

Sequences of both genes were checked and edited in Chromas Lite (Version 2.1; 2012; Technelysium Pty Ltd), then checked for identity using the BLAST algorithm against the GenBank database. Mixed peaks were edited to include degenerate base codes if these were present. To be classed as a mixed peak, points of inflection in both peaks had to match exactly, whilst the smaller peak could be no less than one third of the size of the larger (to account for PCR and sequencing biases; most mixed peaks were near equal in size). Recombination Detection Program 4 (RDP4, Martin et al. 2015) was used to detect recombination in HSP70 and *Orco* using the following statistical methods: RDP, GENECONV, MaxChi, BootScan and SiScan. Due to the short sequence lengths, the window size for the recombination detection methods was set to 20bp (Miraldo et al. 2011). No recombination was detected in any statistical analysis.

Following this, PHASE was used to reconstruct alleles assuming no recombination (Stephens et al. 2001; Stephens & Scheet 2005), implemented in DNASP (version 5.10, Librado & Rozas 2009). To test for differences between haplotype thresholds (following recommendations from Garrick et al. (2010)), the data were run in separate analyses at 0.6 and 0.95 thresholds (1000 iterations, 1000 burn in). Comparing both haplotype networks and Maximum Parsimony trees in Network (version 5, Bandelt et al. 1999; Polzin & Daneshmand 2003; fluxus-engineering.com) showed no differences in the alleles constructed for either gene, whilst no unresolved genotypes were detected. For future analyses, the 0.95 threshold data were used for both genes.

An issue occasionally found in arthropod nuclear gene phylogeography is segregation of a large number of alleles at a locus, resulting in low power to reconstruct haplotypes (Garrick et al. 2010; Walter et al. 2014). Whilst a common solution is to remove apparent heterozygotes from the analysis, here most individuals are heterozygous in HSP, thus making this solution impractical. Alcaide et al. (2011) suggest that for PHASE to accurately reconstruct nuclear haplotypes in the Major Histocompatibility Complex (MHC) there needs to be a minimum allele to individual ratio of 1:2. Though our HSP data violate this rule due to most individuals being heterozygous (allele to individual ratio: 84:93, 1:1.107) and a low percentage of common alleles (the five most common alleles hold 57 of the total pool of 186 alleles) that PHASE always resolved all genotypes suggests that this is not an issue with our analysis. *Orco* shows somewhat more structure, with an allele to individual ratio of 27:76 (1:2.815).

Alleles were then aligned in BioEdit (Version 7.1.11; 2013; Hall 1999) using the CLUSTAL W algorithm with default settings (Thompson et al. 1994). Nucleotide-sequence-based estimates of genetic differentiation (K_{ST} and K_{ST}*) for both genes were calculated in DNASP (5.10.01; 2010; Librado & Rozas 2009), assessing statistical significance by permutating sequences among samples (with 1000 permutations) for both K_{ST} and K_{ST}*. K_{ST}* is a modification of K_{ST} which adds a logarithmic function to "$d_{ij,jk}$ (which) denotes the number of differences (restriction sites or nucleotide sites) between the *jth* sequence from locality *i* and the *kth* sequence from locality *j*", with the effect of reducing the weighting given to large numbers of nucleotide differences between alleles (Hudson et al. 1992). Tests of selection were computed in MEGA6 using a codon-based Z-test of selection under the modified Nei & Gojobori method (Nei & Gojobori 1986; Zhang et al. 1998): using the ratio of synonymous substitution per synonymous site (d_S) and the number of nonsynonymous substitutions per

nonsynonymous site (d_N). The variances of d_S and d_N were computed with bootstrap analysis (1000 replicates), and the following hypotheses examined: neutral evolution predicts $d_S=d_N$, whilst positive selection predicts $d_S<d_N$, and purifying selection predicts $d_S>d_N$ (Hughes & Nei 1989).

5.3 Results

5.3.1 Mitochondrial data

5.3.1.1 COI sequences

COI sequences were clear of double peaks, indels, frameshifts and stop codons. Once the low-quality ends of the sequences were removed, 318bp of COI was used in the analysis from 108 individuals of *Gnorimus nobilis* (Table 5.2; Fig. 5.1; Supplementary Table 7). Though most sequences were longer than this (up to 512bp), the reliance on old museum samples stored dried and pinned from certain glacial refugia (many samples from Spain and southern Italy were over 20 years old, some were over 50 years old) or samples not preserved with an emphasis on retaining high-quality DNA (most of the eBay samples over 5 years old) meant that some of the amplicons were faint, and sequence quality sometimes fell below acceptable levels, thus leaving the trimmed alignment at 318bp. However, this should be sufficient to identify divergences between predicted glacial refugia based on the 3.54%My^{-1} (≈11 base pair mutations in 318bp between lineages per million years; Papadopoulou et al. 2010) mutation rate for Coleoptera COI, assuming reciprocally monophyletic groups.

The haplotype network and the Maximum Likelihood, Neighbour Joining and Bayesian Inference trees all supported a separation within *Gnorimus nobilis* into two groups representing eastern and western geographical clades (Figs.5.2, 5.3, 5.4 & 5.5), though branch bootstrap values were low from two phylogeny methods (ML bootstrap=45, BI bootstrap 57), whilst high in the NJ (bootstrap=98). Individuals of the “western clade” were found distributed from southern France, Germany and the UK in western Europe, and throughout central Europe from Latvia/Poland in the north to Greece in the south, and including south of the Alps in Italy (Fig. 5.6). Individuals of the “eastern clade” were distributed in Russia and Ukraine (plus one individual in Macedonia – see below). All trees

indicated a distinct clade in southern Italy that was recovered as the sister clade to the main western-eastern clades within *G. nobilis* with very strong support from two trees (ML bootstrap=90, BI bootstrap 100), but low support in the NJ (bootstrap=45), but did not recover the Spanish samples as being a distinct clade. One sample from Greece belonged to the most common haplotype, whilst the other sample grouped closely to a GenBank sample from Greece. These latter Greek haplotypes were nearly equidistant between the western clade and the eastern clade, being 6bp from the most common haplotype and 8-10bp from the Eastern clade. The Greek clade was identified as being more closely related to the Eastern clade in the MP, NJ and BI analyses. Individuals previously identified as *G. nobilis bolshakovi* had a distinctive eastern clade haplotype, which was shared by samples from central Ukraine, and one sample identified as *G. nobilis macedonicus*. The other two *G. nobilis macedonicus* samples displayed the common (western clade) haplotype. In addition, all analyses showed clear separation between *G. nobilis* and *G. variabilis*, and a well-supported divide within *G. variabilis* between samples from the eastern and western areas of the distribution (ML bootstrap=100, NJ bootstrap=100, BI bootstrap 100).

Within-group distances between COI haplotypes were considerably lower for the Western population than in any other due to large numbers of a single common haplotype (Table 5.3). Between population distances of the Eastern clade were 0.0486-0.0553 compared to the Western, Spanish and Italian clades, whilst *G. variabilis* differed by 0.1866-0.2038 to the *G. nobilis* populations. In addition, overall Approximate ages of clade divergence were 1.88-2.44MY (2.34%MY^{-1}, Brower 1994) or 1.24-1.61MY (3.54%My^{-1}, Papadopoulou et al. 2010) between the Eastern and Western *G. nobilis* haplogroups, and 5.90-6.84MY / 3.90-4.52MY between *G. nobilis* and *G. variabilis*.

Both Tajima's *D* and Fu's *F* statistics differed significantly from neutrality (significant negative values) in COI in the Western haplotype group (which includes Spanish and Italian samples), rejecting a hypothesis of constant population size (Table 5.2). Both statistics did not reject the hypothesis of constant population size in the Eastern haplogroup. Significant negative values of both statistics indicate recent selective sweeps, genetic hitchhiking and/or a population expansion after a bottleneck. Mismatch analysis showed a unimodal distribution in Western *G. nobilis* (Fig. 5.11a), indicating a recent population expansion (Excoffier et al. 2009) whilst the bimodal distribution in the Eastern *G. nobilis* indicates an older expansion and/or a stable population (Rogers & Harpending 1992). The 'twin peak' mismatch distribution in *G. variabilis* (Fig. 5.11e) implies that there may be a greater east-west division

in the species than has reported here, though the small sample size limits what can be said. SSD results were not significant for any haplogroup, failing to reject the null hypothesis of population expansion. Raggedness indices were similar in both haplogroups, implying similar expansion times. Western *Gnorimus nobilis* had an estimated expansion time of 224000-344000 years before present, Eastern *G. nobilis* 440000-670000 years before present, and *G. variabilis* 408000-622000 years before present.

5.3.1.2 Control Region

Once the low-quality ends of the sequences were removed, 330bp of CR was used in the analysis from 80 individuals of *Gnorimus*. Sequences used were free of double peaks, indels and frameshifts in the internal primer combinations used (see Methods). The repetitive nature of the CR in Cetoniids (Kim et al. 2014; Kim et al. 2016) makes it extremely difficult to reliably amplify and sequence a clean region across multiple individuals not immediately preserved for genetic analysis. Certain longer primer combinations yielded sequences which did have occasional mixed peaks (hinting at the presence of NUMTs) and so were not used for phylogeographic analysis. Most of the difficulties in sequencing CR outlined by Mancini et al. (2008) were found with *Gnorimus*, including the high A+T nucleotide bias, homopolymer stretches and repeated elements. Similar to other Cetoniid CR sequences (*Protaetia brevitarsis* Kim *et al*. 2014; *Osmoderma opicum* Kim et al. 2016), the Control Region of *Gnorimus nobilis* is littered with start (AAT, ATA and ATT in particular, following the suggestion of Sheffield et al. (2008) of AAT and AAC as start codons in Polyphaga) and stop codons in all reading frames: in the section of CR used seven Open Reading Frames (ORFs) were detected, the longest CR ORF being 141bp. The region used in this study did not have any indels present between individuals. The region used aligns approximately to bases 19145-19477 in the *Protaetia brevitarsis* mitochondrial genome (GenBank accession number: KC775706.1 (Kim et al. 2014)). The approximation is due to the large number of indels between the two sequences.

The haplotype network (Fig. 5.7) and Maximum Likelihood (Fig. 5.8), Neighbour Joining (Fig. 5.9) and Bayesian Inference (Fig. 5.10) trees all strongly supported a separation within *Gnorimus nobilis* into two distinct haplogroups representing eastern and western geographical clades. As with COI individuals from northern Spain were distinctive, as were individuals from southern Italy, but their exact relationships to the two main clades (central

European and Eastern) were unclear, being either more closely rooted with central Europe or the Eastern group depending on the tree building method used. In addition, the ML and NJ trees weakly supported a basal clade within the Western and Central clade exclusive to English *G. nobilis*. Individuals from Kent had a unique haplotype, as did those from the north Worcestershire site, but the southern Worcestershire site possessed two different haplogroups.

Due to the difficulty in sequencing the CR, fewer samples were sequenced than for COI (Table 5.2). The Greek sample with the central European COI haplogroup could not be amplified, but the Greek specimen with a distinctive haplotype in COI had a CR haplotype belonging to the central European CR haplogroup. The Eastern group was also supported, with all individuals with an Eastern COI haplotype also possessing an Eastern CR haplotype, including the *G. nobilis macedonicus* sample with the Eastern COI haplotype. Possibly due to nucleotide differences in the primer binding sites, no *G. variabilis* samples could be sequenced. Within-group distances between CR haplotypes were higher for the Western population than in any other due (Table 5.3). Between population distances of the Eastern clade were 0.0852-0.0872 compared to the Western, Spanish and Italian clades, approximately twice that of COI.

Neither Tajima's *D* nor Fu's *F* statistics differed significantly from neutrality in CR in the Eastern haplogroup, which couldn't reject a hypothesis of constant population size (Table 5.2). In the Western haplogroup, Fu's *F* differed significantly from neutrality, whilst Tajima's *D* approached the significance cut off (P=0.059). Mismatch analysis showed a unimodal distribution in Western *G. nobilis* (Fig. 5.11c), indicating a recent population expansion (Excoffier et al. 2009), though there is a hint that there could be a bimodal distribution. The pattern is very similar to that in the Eastern *G. nobilis* (Fig. 5.11d), with both distributions being intermediate between a clear unimodal distribution and a bimodal. SSD results were not significant for any population, failing to reject the null hypothesis of population expansion. The Raggedness index was similar between Eastern CR and Western COI and Eastern COI, whilst the statistic was considerably lower in Western CR, suggesting a weaker model fit.

5.3.4 Heat Shock Protein

Once trimmed for quality, the HSP70 alignment was 206bp long. BLASTn searches for the sequences showed close similarity on GenBank to HSP68 in *Tribolium castaneum* (Tenebrionidae, 75%, NM_001170729.1) and HSP70 in *Tenebrio molitor* (Tenebrionidae, 75%, JQ219849.1). No other Scarabaeidae has had HSP70 sequenced. Grouping the samples into haplogroups identified from the mitochondrial analysis (Western *G. nobilis*, Eastern *G. nobilis*, and *G. variabilis* for COI) showed similar levels of allele and nucleotide diversity in all three groups (Table 5.4), but *G. variabilis* showed lower levels of nucleotide diversity than the other clades: 0.025 in 24 alleles with 16 polymorphic sites, against 0.027 from 20 alleles with 21 polymorphic sites in the Eastern *G. nobilis* clade, and 0.031 from 142 alleles with 45 polymorphic sites in Western *G. nobilis*.

The alleles recovered by PHASE showed a chaotic double star pattern (Fig. 5.12), with no apparent consensus between COI identification and HSP70 alleles, including between *G. nobilis* and *G. variabilis*. The individuals were grouped according to their COI/CR haplogroup (Western *G. nobilis*, Eastern *G. nobilis*, Spanish *G. nobilis*, Italian *G. nobilis*, Western *G. variabilis* and Eastern *G. variabilis*), then tests of genetic differentiation (K_{ST} and K_{ST}*) were carried out between each pair of haplogroups. Only K_{ST}* reported significant genetic differences between some clades (Western *G. nobilis* and Eastern *G. nobilis*, Eastern *G. nobilis* and Spanish *G. nobilis*, and Eastern *G. nobilis* and Western *G. variabilis*) with K_{ST} approaching the significance cut-off (p=0.05) in all three instances (Table 5.5).

The Z-tests of selection showed all six clades (Western *G. nobilis*, Eastern *G. nobilis*, Spanish *G. nobilis*, Italian *G. nobilis*, Western *G. variabilis* and Eastern *G. variabilis*) to all differ significantly from the hypotheses of neutral evolution (Table 5.6) and purifying selection, whilst a null hypothesis of positive selection was not refuted.

5.3.5 *Orco*

Once trimmed for quality, the *Orco* alignment was 150bp long. BLASTn searches for the initial longer sequences revealed a close similarity to *Orco* from *Holotrichia* and *Anomala* (both Scarabaeidae), but with a large indel 66 base pairs long. Sequences on either side of this were 80-85% similar to *Holotrichia* and *Anomala*, but the indel did not match any other sequence on GenBank. However, contrary to expectations of this gene showing some

divergence between species within genera individuals of *Gnorimus variabilis, G. nobilis* and *G. subopacus* all had similar or the same alleles in this gene. One divergent allele cluster was more associated with *G. variabilis* in the network (Fig. 5.13), but still contained haplotypes from *G. nobilis* individuals.

The Z-tests of selection showed all five clades (Western *G. nobilis,* Eastern *G. nobilis,* Italian *G. nobilis,* Western *G. variabilis* and Eastern *G. variabilis*) to all differ significantly from the hypotheses of neutral evolution (Table 5.7) and positive selection, whilst a null hypothesis of purifying selection wasn't refuted.

The individuals were grouped according to their COI/CR haplogroup (Western *G. nobilis,* Eastern *G. nobilis,* Spanish *G. nobilis,* Italian *G. nobilis,* Western *G. variabilis* and Eastern *G. variabilis*) and tests of genetic differentiation (K_{ST} and K_{ST}*) carried out between each pair of haplogroups. In contrast to the HSP70 data, both statistics showed the same 'significant/non-significant' result for each pair of haplogroups. Western *G. variabilis* differed from every other group other than Eastern *G. variabilis,* Spanish *G. nobilis* differed from Eastern *G. variabilis,* Western *G. nobilis* and Italian *G. nobilis,* and Italian *G. nobilis* differed from Eastern *G. nobilis.*

5.4 Discussion

5.4.1 MtDNA diversity and phylogeographic patterns in European *Gnorimus*

Though mitochondrial DNA has traditionally been the marker of choice for phylogenetic and phylogeographic study (Hebert et al. 2003), modern studies based on gene sequences should use multiple markers to avoid inferring too much from single genes, where hybridisation, lack of reciprocal monophyly, incomplete lineage sorting or retention of ancestral polymorphism may complicate single-gene interpretations (Rand & Harrison 1989; Shaw 2002). Using the most commonly used and best understood mitochondrial gene (Cytochrome Oxidase 1) here shows a clear split between *Gnorimus nobilis* and *G. variabilis,* as expected from the morphological analysis (Chapter 4)), and supporting their description as distinct species. The COI data also identified differentiation among haplotypes within *G. nobilis* which form haplogroups with a geographical basis and appear to map to positions of expected classical glacial refugia: Iberia, southern France, southern Italy, the Balkans and Eastern

Europe (Russia - Caucasus) (Figures 5.2, 5.3, 5.4 & 5.5; Hewitt 1999). Most individuals from western-central to northern Europe possessed haplotypes belonging to a single common haplogroup, implying an expansion from a single glacial refugium, likely from southern France and/or the Balkans. Individuals from Spain and southern Italy possessed different haplotypes not found elsewhere in the western-central-northern European distribution, suggesting that refugial populations in these Mediterranean areas did not contribute to the post-glacial expansion into northern Europe and have remained isolated from the post-glacial distribution of *G. nobilis* across north-central Europe.

The distinct division between two geographically-based haplogroups in western-central Europe and eastern Europe (Russia and Ukraine) suggests an existing geographical differentiation of *G. nobilis* populations with little overlap, that most likely results from range expansions from a central (southern France and/or Balkans) and eastern (Caucasus / Caspian Sea) European refugium respectively. This does not fit closely with any of Hewitt's (2000; 2004) classic models of European species postglacial expansion (Chapter 1), nor does it tightly link to range expansions of some of the main host tree species (primarily form Iberia and Italy in the oaks (*Quercus* spp., Hewitt 1999; Petit et al. 2002.) and the Balkans in the beech (*Fagus sylvatica*; Magri et al. 2006)). The pattern is similar to that in the stag beetle (*Lucanus cervus*) in that it supports an east-west split, but *L. cervus* spread from a refugium in Iberia following oaks, not from a refugium around southern France (McKeown *pers. comm.*).

The presence of a genetically differentiable clade of *G. nobilis* in Russia / Ukraine supports the recognition of an Evolutionary Significant Unit (ESU; Moritz 1994) that may correspond to the description of the *G. n. bolshakovi* subspecies (all individuals identified as *G. n. bolshakovi* morphologically possessed "eastern" haplotypes). The identification of an individual from Greece / Macedonia that possesses an eastern clade haplotype when all other individuals from the Balkans possess western-central haplotypes may support the grouping of *G. n. macedonicus* with *G. n. bolshakovi*, and may indicate that either the eastern clade is spreading into the Balkans or that this individual represents a remnant population of the eastern clade that is being overrun by continuing expansion of the western clade eastwards. Further population genetic studies of the Balkans and Eastern Europe using more variable genetic markers such as microsatellites or genome-wide Single Nucleotide Polymorphisms (SNPs) will be needed to resolve these questions. Though the sample size for *G. variabilis* was considerably smaller, a similar split between samples from Ukraine and the rest of

Europe (Italy, France and the UK) indicates that both *Gnorimus* species may have undergone similar past population subdivisions (and subsequent expansions / range changes) across Europe during and after the last glacial maximum.

Tajima's *D* and Fu's *F* statistics differed significantly from neutrality in COI in the Western haplogroup, rejecting a hypothesis of constant population size, and together with the mismatch distribution, SSD's failure to reject population expansion, and a high raggedness index strongly indicates a historically recent population expansion. A recent population expansion was not indicated in the Eastern COI clade from the *D* or *Fs* values, though the SSD's failure to reject population expansion and a high raggedness index indicates a recent population expansion. The bimodal mismatch distribution indicates more long term demographic stability than the western clade and possibly both a small recent expansion and an additional division within the clade (Meraner et al. 2008).

This geographically-based division between western-central and eastern clades at COI, with divergent populations in Mediterranean refugial positions, was corroborated by the data from Control Region. Spanish, south Italian and western / eastern CR clades were recovered with strong support (Figures 5.7, 5.8, 5.9, & 5.10). More unique haplotypes were recovered by CR sequences than by COI sequences throughout central and northern Europe, including three unique to the UK, with one haplotype restricted to Kent, another to northern Worcestershire, and samples from southern Worcestershire possessing a different pair of haplotypes. Additional unique northern haplotypes were also recovered from Latvia. Such diversity at the species range edge adds weight to a possible expansion from a northern refugium as well as from more southern (France and/or Balkans) refugia (Figs. 5.14 & 5.15). That a recent population expansion occurred is highly likely: Fu's *F* statistic rejected a null hypothesis of no population expansion, Tajima's *D* nearly rejected a null hypothesis of no population expansion (P=0.059), SSD failed to reject a null hypothesis of population expansion, and within-group divergences for Western COI are considerably lower than those in any other *Gnorimus* clade (Table 5.2). The raggedness index was fairly low however (0.0597, compared to 0.1649 in Western COI). Haplotype diversity was considerably higher in Western CR than in Western COI (0.798 and 0.398 respectively), whilst it was similar for Eastern CR and Eastern COI (0.778 and 0.894 respectively), again suggesting that a historically recent population bottleneck and expansion has occurred in the western clade but not in the eastern clade (CR haplotype diversity is predicted to recover from a bottleneck more quickly than COI diversity).

Estimated times of divergence between the Eastern and Western clades were 1.24-2.44MY depending on the mutation rate, and 3.90-6.84MY between *G. nobilis* and *G. variabilis*. Despite a similar level of between species divergence to species in the *Osmoderma* species complex, Audisio et al. (2009) estimated the divergence between European *Osmoderma* to date to 8-9MY using a $2\%MY^{-1}$ rate (Nei 1987), slightly older than the estimates for *Gnorimus*, due in part to the different mutation rate used. Re-calculating the divergence between *G. nobilis* and *G. variabilis* using this mutation rate puts the divergence at 6.9-8.0MY, still slightly younger than the *Osmoderma* divergences, but still in the late Miocene. Population expansion times were similar for Eastern *G. nobilis* and *G. variabilis* in the Middle Pleistocene (408000-670000 years ago), whilst expansion times for Western *G. nobilis* were younger (225000-344000 years ago).

Gnorimus nobilis Control Region sequences show a similar molecular structure to that reported from many other insects: high A+T content (88%+), tandem repeats, and extremely large size (Zhang & Hewitt 1997; Lunt et al. 1998). Increased CR sequence variation compared to other mtDNA regions is thought to result from, among other mutational processes, slipped-strand mispairing during replication (Lunt et al. 1998) which has been postulated as a common occurrence in repeat units inside the first and last hypervariable regions (Fumagalli et al. 1996). Implications from point mutations suggest that CR is three times more variable than COI in some beetles (Mancini et al. 2008), thus making the gene considerably more suitable for phylogeographic studies than COI. Though the increased among-individual variation is useful in population studies, the region is difficult to work with in *G. nobilis*: single primer pairs often produce multiple CR products due to replication (Xu & Fonseca 2011), high A+T content makes primer design and product amplification unreliable (over 80 individual primers were used in this study), the extremely large size (approximately 6kb) due to a large number of repeats requiring high-fidelity *Taq*, and large blocks of mononucleotide repeats often caused sequencing to fail (some short A or T regions 5-10bp, but one long 20bp G region near the ND2 section of *G. nobilis* CR that was near impossible to sequence across and retain an acceptable level of sequence quality) (Cha et al. 2007). However, once a suitable region was found, PCR amplification and sequencing was straightforward, though it proved unreliable in older specimens (see Chapter 6).

That Control Region shows more variation than COI, but with largely the same phylogeographic patterns, complies with data on Lepidoptera (Vila & Björklund 2004) which have similar CR structures to Coleoptera, being AT rich (89-96%) and possessing satellite-

like repeat units in some species, though the region is rarely over 500bp long. One clade identified by CR that wasn't identified in COI is that English samples belong to a unique CR clade (Figs. 5.8 & 5.9), which may imply that said individuals spread to the UK from a northern refugium in France (Fig. 5.15). Trying to compare the CR data here to other studies that used CR in European terrestrial arthropod phylogeography is difficult due to the lack of similar studies. Vila et al. (2005) reported similar glacial refugia to those identified here (southern France and Iberia) in the butterfly *Erebia triaria*, but detected multiple Iberian refugia and no population expansion in this montane specialist. Meraner et al. (2008) discovered two CR clades in the coddling moth (*Cydia pomonella*) which appeared to have split in the mid-lower Pleistocene, but these haplotypes were found in the same populations, indicating secondary contact and hybridisation after recent range expansions. Mancini et al. (2008) reported divergence in CR between Iberian and Italian populations of the pollen beetle *Meligethes thalassophilus*, but didn't include samples from throughout Europe.

5.4.2 Nuclear gene diversity and allele conservation across *Gnorimus*

Data on population level diversity for Heat Shock Protein 70 is scarce for invertebrates. Hou et al. (2014) reported intraspecific sequence divergence to be between 1.4-4.4% within populations, and 4-6% between higher clades. However, no mention was made of heterozygous positions in this study. Most *Gnorimus* were heterozygous, with allele sharing between all three haplogroups used (Western and Eastern *G. nobilis* and *G. variabilis*). The allele network (Fig. 5.12) shows two primary allele clusters to the 'top' and 'bottom' of the network: most individuals are heterozygous possessing one allele from the 'top' cluster, and one from the 'bottom' cluster. HSP70 is the most conserved member of the HSP gene family (Baringou et al. 2016), so quite why *Gnorimus* has such extensive diversity in HSP70 but with no obvious differentiation between the different species is unknown. HSP70, as a generally expressed 'stress gene', can be expressed to help Diptera survive increased temperatures, hypoxia, and/or dehydration (Abaza 2015), or bacterial infection in *Copris* dung beetles (Hwang et al. 2008). A possible explanation is that the HSP70 has undergone a duplication event somewhere in the evolution of *Gnorimus*: other beetles have large regions of their genomes replicated (Wang et al. 2008), and heat shock proteins in general are known to have been replicated a number of times in insect evolution (Concha et al. 2014; Zhang et al. 2014). If such paralogues have been sequenced in this study then it may account for the

high diversity in HSP70 alleles discovered and the large number of apparently heterozygous individuals.

The Z test for signals of natural selection indicated that HSP was under positive selection in all the clades investigated here (Western *G. nobilis*, Eastern *G. nobilis*, Italian *G. nobilis*, Western *G. variabilis* and Eastern *G. variabilis*). Hughes & Nei (1989) discussed several reasons that might account for positive selection in genes, the main one being selection for advantageous mutations. However, as they also point out for the Major Histocompatibility Complex (MHC) in mammals, which also shows positive selection, selection for advantageous mutations can be ruled out because it cannot explain the high degree of polymorphism in the complex, nor its long persistence (in millions of years) within populations. In a similar vein, only two *Gnorimus* were homozygotes (97.8% heterozygosity), and that the same HSP70 alleles were found in both *G. nobilis* and *G. variabilis* (well separated species according to COI) indicated that these HSP70 alleles have been present in the wider 'European *Gnorimus*' population for at least a million years. Hughes & Nei (1989) accepted the hypothesis of overdominant selection as accounting for similar findings in MHC whereby "a particular class II MHC molecule preferentially binds to a particular foreign peptide, thus providing improved recognition of that peptide by helper T cells and enhanced immune response. A heterozygote for two different alleles at a locus will therefore have resistance to two different types of pathogens and consequently have a higher fitness than a homozygote for either allele" (Hughes & Nei 1989). As HSP70 mediates cell stress responses, it may be advantageous for individuals to possess two different alleles coding for very different proteins so that the two alleles can be upregulated under different stress conditions, giving the individual a higher chance of surviving, particularly in variable environments. The HSP70 pattern in *Gnorimus* could be explained via two hypotheses, with the limited knowledge available regarding *Gnorimus* biology and physiology: HSP70 allele variation helps mediate seasonal survivability in *Gnorimus* larvae; and HSP70 allele variation helps individuals survive as larvae and as sexual adults under different environmental conditions and evolutionary pressures.

Gnorimus larvae live and grow in hollows in trees, which experience seasonal variations in temperature. However, they are slightly insulated from the extremes: individual arthropods living in soil environments regulate HSP70 at a lower rate than those on the surface (Liefting & Ellers 2008). Additionally, *Gnorimus*-inhabited tree cavities are generally comparatively dry, with the larvae being found from the surface of the pabula to deep within the cavity

(Blake *pers. obvs*). These conditions, though not terribly extreme in temperature or humidity, may invoke responses in HSP as a general stress-tolerance gene. Both *Gnorimus* species are tolerant of high temperatures (Renault et al. 2005), but *G. variabilis* is tolerant year round to below-freezing temperatures (Vernon & Vannier 2001), whilst *G. nobilis* empties its gut over winter, relying on stored fat to survive throughout the winter and pupate in spring (Vernon & Vannier 2002). Only Vernon & Vannier (2001) have investigated the thermal tolerance of *G. variabilis* from multiple regions, finding populations further to the east in France to have considerably greater thermal tolerance than those to the West (at near-identical latitudes), corresponding with a climate less influenced by the Gulf Stream (Chapter 7). Thus if HSP70 is mediating stress responses in larvae due to seasonal fluctuations in temperature and humidity, including freezing resistance, then some variability in the subpopulations would be expected based on the different environmental conditions the larvae are living in. In addition, as populations of *G. nobilis* in the UK are restricted to fruit tree orchards, they may have different HSP70 alleles to allow them to cope with the reduction in insulation caused by the smaller diameter of the tree trunk and rot-hole. Though there is no apparent geographical segregation in the HSP70 network in our samples, if this hypothesis is correct then there will be a degree of population differentiation in the gene based on locality and the climatic factors there (so Mediterranean populations may be differentiated from those on the eastern European plains Chapter 7) that doesn't necessarily correspond to the assumed species divides: so Eastern *G. variabilis* will show the same level of differentiation to Mediterranean populations as Eastern *G. nobilis.*

As holometabolous insects, adult and larval *Gnorimus* will experience different selective pressures as larvae living throughout multiple years and seasons in a rot-hole, and as adults flying, mating and feeding in warm summer temperatures. Adults often bask in sunlight to warm themselves (Bates *pers. comm.*), especially when first emerging from the tree, and rapidly warm their bodies to over 30°C in an ambient temperature of 18.5°C (Bates *pers. comm.*). Being subjected to temperatures below freezing as a larva with no UV radiation, and temperatures above 30°C as an adult with UV radiation present, may suggest that the high level of heterozygosity in HSP70 is due to the completely different environments adults and larvae live in. If this is the case, then there should be very little genetic differentiation in HSP70 across the different populations found in different environmental regions, but possibly some differentiation related to the species, *G. variabilis* generally preferring warmer less extreme environments (Trizzino et al. 2013; Chapter 7).

The only significant values of genetic differentiation (K_{ST}*) recovered all involved the Eastern *G. nobilis* population, differing from the Western *G. nobilis*, Spanish *G. nobilis* and Western *G. variabilis* (Table 5.5). This result does not closely fit with either hypothesis for HSP70 diversity, but does indicate that the Eastern *G. nobilis* have adapted to a very different environment to the other populations, most likely due to their distribution in colder continental regions (Chapter 7), and again supporting the differentiation of the *G. n. bolshakovi* ESU. The lack of differentiation between the different populations largely adds support to the hypothesis of HSP70 generally mediating stress between the adult and larva: if differentiation was due to climate then it would be expected that the southern Mediterranean groups (Spain and Italy) would be different to the population from central and northern Europe.

This is the first study to utilise *Orco* as a gene in phylogeography. It was expected, based on other studies, that there would be clear differences in the gene sequences between *Gnorimus nobilis*, *G. variabilis* and *G. subopacus*. However, *G. subopacus* had the same alleles as the common haplotypes in *G. nobilis* and *G. variabilis* (Fig. 5.13). Considering the gene's potential role in mate detection (Shen et al. 2011), and the lack of significant host differences between the species (Mannerkoski et al. 2010b on *G. variabilis*, Mannerkoski et al. 2010a on *G. nobilis*, no data available on host preference for *G. subopacus*), this result was unexpected. However, more subtle host-finding cues could be used within the genus: Trizzino et al. (2013) demonstrated spatial niche partitioning between *G. variabilis* and *G. nobilis*, which may be due in part to preferences for different rot-hole causing fungi. It is unlikely that there has not been sufficient evolutionary time since the last common ancestor of all three species to allow for complete lineage sorting in this gene: given the high between clade distances in COI between *G. variabilis* and *G. nobilis* (Table 5.3; no *G. subopacus* were included in the COI study) and the violation of Palumbi et al. (2001)'s 'three-times rule' this strongly implies that selection is maintaining the retention of ancestral alleles.

The Z-tests of selection showed that the *Orco* gene is very likely to be under purifying selection in each of the clades investigated (Western *G. nobilis*, Eastern *G. nobilis*, Italian *G. nobilis*, Western *G. variabilis* and Eastern *G. variabilis*), which is likely to explain the retention / sharing of ancestral alleles between the three species. Other studies have found olfactory genes to be under purifying selection in arthropods (Vieira et al. 2007 on *Drosophila*). Using the genetic differentiation tests however shows that Western *G. variabilis* is significantly different from every *G. nobilis* group, but not from Eastern *G. variabilis*.

Eastern *G. variabilis* is only differentiated from Spanish *G. nobilis*, whilst Spanish *G. nobilis* are also differentiated from Western *G. nobilis* and Italian *G. nobilis*. Finally, Eastern *G. nobilis* are differentiated from Italian *G. nobilis* (Table 6). This suggests that there are strong evolutionary effects on this gene thanks both to the tests of purifying selection, and the population differentiation indicated by the K_{ST} tests.

5.5 Conclusions

Both COI and Control Region show a deep divergence within *Gnorimus nobilis* between the Western and Eastern populations, which corresponds with the original description of the Russian population as a separate subspecies. However, as this population shares its COI and CR haplotypes with individuals from Eastern Ukraine, then it is likely that *G. nobilis bolshakovi* isn't restricted to Russia as it was originally defined, but is instead found through to Eastern Ukraine. An earlier expansion time in Eastern *G. nobilis* than the Western clade could suggest that the population spread through Ukraine, and possibly further, but was swamped by a more recent expansion in the Western clade. This may account for the Eastern haplotype in Macedonia. Additional clades were also found in Spain and southern Italy which likely represent areas of glacial refugia from which the species didn't expand. Though the sampling strategy for *G. variabilis* was more limited, a similar (though less marked) divergence between Eastern and Western populations was identified.

Considerable nuclear gene sharing was also found between the different clades: Heat Shock Protein 70 showed a lack of structuring between populations and species, but instead strong evidence for positive selection as a driver of high genetic diversity was discovered. This contrasted with the gene *Orco*, which suggested purifying selection on in all populations tested, as well as genetic differentiation between some of the clades. Therefore whilst the patterns from HSP70 don't inform patterns of post-glacial phylogeography they may instead indicate the need for high genetic diversity for individual *Gnorimus* to allow individuals to respond to a variable 'stress' environment. *Orco* however suggests population divergence, which may affect the behaviour of each clade as *Orco* may mediate host cue or male/female pheromone detection.

Table 5.1. Primers used in the study

Gene / Region targeted	Name	Code
COI	COIF30	5’-GCAATTGGATTACTAGGATTTATTG-3’
	COIR546	5’-TTCATGTTGTGTATGCATCTGG-3’
Control Region	111fCR	5’-CCCCTATATCTGATTTTACTTT-3’
	610R CR	5’-GGGTTAATTGCTGAATCTT-3’
Heat Shock Protein 70	HSPF1	5’-CAGACACCGAACGTTTACTCG-3’
	HSPR1	5’-ATCGCGAACTGTGGTGCC-3’
Orco	Or83b118f	5’-CCAGGAAGGACCCTAATAACG-3’
	Or83b340R	5’-TCCAACAAATTGGAATGCAG-3’

Table 5.2. COI and CR mtDNA sequence variability measures for *Gnorimus* and tests of population expansion.

Subset	**Individuals**	**Haplotypes**	**Length (bp)**	**Haplotype diversity**	**Polymorphic sites**	**A.T:G.C ratio**	**Tau**	**Theta**	**Raggedness**	***D***	***F***	**SSD**
G. nobilis **Western COI**	97	9	318	0.3982	11	68.2:31.8	5.03	0.616	0.1649 *P*=0.760	-1.6692 ***P*=0.020**	-4.2573 ***P*=0.029**	0.001 *P*=0.730
G. nobilis **Eastern COI**	12	8	318	0.8939	13	67.5:32.5	9.82	3.251	0.1159 *P*=0.620	-0.3297 *P*=0.438	-1.700 *P*=0.150	0.077 *P*=0.160
G. v*ariabilis* **COI**	14	9	318	0.879	13	66.1:33.9	9.13	7.10	0.029 *P*=0.860	0.475 *P*=0.714	-1.685 *P*=0.184	0.024 *P*=0.60
G. nobilis **Western CR**	68	19	330	0.7976	21	88.7:11.3	1.20	1.472	0.0597 *P*=0.320	-1.38118 *P*=0.059	-8.97137 ***P*<0.001**	0.016 *P*=0.330
G. nobilis **Eastern CR**	9	4	330	0.7778	5	88.8:11.2	0.50	0.398	0.1728 *P*=0.340	-0.91004 *P*=0.244	-0.28567 *P*=0.318	0.032 *P*=0.210

Table 5.3: mean within group distances and between group distances in COI and CR. Within group distances are shown in the second column

			Western	Spanish	Italian	Eastern
COI	Western	0.0010				
	Spanish	0.0041	0.0061			
	Italian	0.0032	0.0180	0.0229		
	Eastern	0.0132	0.0486	0.0553	0.0549	
	Variabilis	0.0151	0.1913	0.1910	0.1866	0.2038
CR	Western	0.0073				
	Spanish	0.0068	0.0154			
	Italian	0.0	0.0238	0.0244		
	Eastern	0.0050	0.0872	0.0869	0.0852	

Table 5.4. HSP70 and *Orco* gene sequence diversity within *Gnorimus* haplogroups

Haplogroup	Gene	Length (bp)	Number of individuals	Number of alleles	Allele diversity	Nucleotide diversity	Polymorphic sites
Western	HSP70	206	71	142	0.975	0.031	45
Eastern	HSP70	206	10	20	0.963	0.027	21
Variabilis	HSP70	206	12	24	0.946	0.025	16
Western	Orco	150	58	116	0.733	0.0301	27
Eastern	Orco	150	5	10	0.622	0.0178	7
Variabilis	Orco	150	11	22	0.877	0.0522	18

.

Table 5.5. Genetic differentiation (K_{ST} and K_{ST}* - see text) estimated from HSP70 (A) and Orco (B) gene frequencies between *Gnorimus* mtDNA haplogroups (Significant departures from zero in bold).

(A) HSP70	Western G. nobilis		Eastern G. nobilis		Italian G. nobilis		Spanish G. nobilis		Western G. variabilis	
	K_{ST}	K_{ST}*	K_{ST}	K_{ST}*	K_{ST}	K_{ST}*	K_{ST}	K_{ST}*	K_{ST}	K_{ST}*
Eastern G. nobilis	0.0059 *p*=0.076	**0.0073** ***p*=0.019**	-	-						
Italian G. nobilis	-0.0031 *p*=0.711	-0.0019 *p*=0.788	-0.0208 *p*=0.752	-0.009 *p*=0.674	-	-				
Spanish G. nobilis	0.0005 *p*=0.325	0.0029 *p*=0.110	0.0253 *p*=0.059	**0.0344** ***p*=0.012**	-0.0117 *p*=0.586	-0.0089 *p*=0.652	-	-		
Western G. variabilis	0.0061 *p*=0.090	0.0033 *p*=0.099	0.0277 *p*=0.089	**0.0281** ***p*=0.036**	-0.0196 *p*=0.614	-0.0033 *p*=0.458	0.0041 *p*=0.331	0.0149 *p*=0.149	-	-
Eastern G. variabilis	0.0008 *p*=0.327	0.0001 *p*=0.396	0.0093 *p*=0.262	0.015 *p*=0.158	-0.071 *p*=0.934	-0.010 *p*=0.421	-0.006 *p*=0.523	0.0114 *p*=0.258	-0.0037 *p*=0.448	0.0115 *p*=0.234
(B) Orco	Western G. nobilis		Eastern G. nobilis		Italian G. nobilis		Spanish G. nobilis		Western G. variabilis	
	K_{ST}	K_{ST}*	K_{ST}	K_{ST}*	K_{ST}	K_{ST}*	K_{ST}	K_{ST}*	K_{ST}	K_{ST}*
Eastern G. nobilis	0.0025 *p*=0.267	-0.0005 *p*=0.386	-	-						
Italian G. nobilis	0.0135 *p*=0.110	0.0117 *p*=0.069	**0.1687** ***p*=0.034**	**0.1614** ***p*=0.014**	-	-				
Spanish G. nobilis	**0.0207** ***p*=0.043**	**0.0187** ***p*=0.047**	-0.0171 *p*=0.495	0.0033 *p*=0.290	**0.2446** ***p*=0.004**	**0.3031** ***p*=0.004**	-	-		
Western G. variabilis	**0.0464** ***p*=0.002**	**0.0376** ***p*=0.004**	**0.0995** ***p*=0.010**	**0.0704** ***p*=0.009**	**0.0948** ***p*=0.040**	**0.0470** ***p*=0.045**	**0.156** ***P*<0.001**	**0.158** ***P*<0.001**	-	-
Eastern G. variabilis	0.0039 *p*=0.287	0.0110 *p*=0.105	0.0384 *p*=0.183	0.053 *p*=0.128	0.0899 *p*=0.167	0.0671 *p*=0.226	**0.0774** ***p*=0.041**	**0.1414** ***p*=0.037**	-0.0417 *p*=0.931	-0.0285 *p*=0.928

Table 5.6. Test statistics (with significance) for deviation from signals of neutral evolution or natural selection acting on HSP70 gene frequencies in six *Gnorimus* mtDNA clades (Significant results in bold).

	Neutral evolution		Positive selection		Purifying selection	
Clade	Statistic	*P*	Statistic	*P*	Statistic	*P*
Western G. nobilis	-3.9963	**0.0001**	-4.0208	1	4.04119	**4.71E-05**
Eastern G. nobilis	-3.2500	**0.0014**	-3.2605	1	3.23116	**0.0007**
Spanish G. nobilis	-3.1159	**0.0024**	-3.0468	1	3.08463	**0.0009**
Italian G. nobilis	-2.2622	**0.0254**	-2.2759	1	2.36013	**0.0099**
Western G. variabilis	-2.5956	**0.0106**	-2.6866	1	2.70238	**0.0039**
Eastern G. variabilis	-2.4627	**0.0152**	-2.5463	1	2.60747	**0.0051**

Table 5.7. Test statistics (with significance *P*) for signals of neutral evolution or natural selection acting on *Orco* gene frequencies in six *Gnorimus* mtDNA clades (Significant results in bold).

	Neutral evolution		Positive selection		Purifying selection	
Clade	Statistic	*P*	Statistic	*P*	Statistic	*P*
Western G. nobilis	2.373	**0.0187**	2.385	**0.0094**	-2.330	1.000
Eastern G. nobilis	2.096	**0.0428**	2.048	**0.0224**	-1.996	1.000
Spanish G. nobilis	2.082	**0.0434**	2.035	**0.0230**	-1.972	1.000
Italian G. nobilis	1.982	**0.0497**	1.973	**0.0270**	-1.946	1.000
Western G. variabilis	2.011	**0.0310**	2.077	**0.0192**	-2.105	1.000
Eastern G. variabilis	2.060	**0.0352**	2.130	**0.0172**	-2.171	1.000

Figure 5.1. Sample map for *Gnorimus* used in this study, showing species/subspecies and sample size (see Table SX in Supplementary Materials for sample details). Base map from www.d-maps.com, in the public domain.

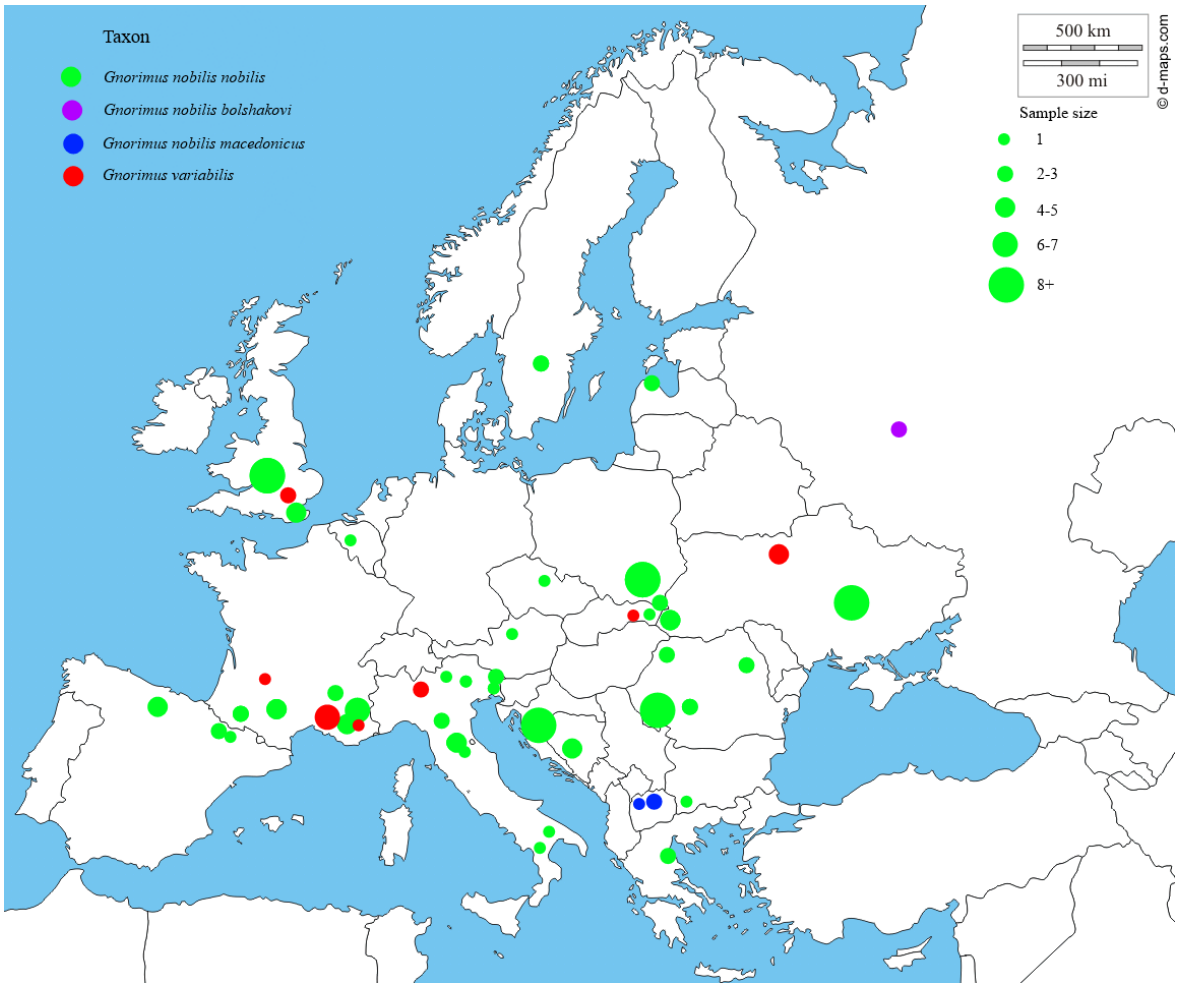


Figure 5.2. COI haplotype network for *Gnorimus:* branch lengths are proportional to number of mutational changes; circle size proportional to number of individuals displaying each haplotype. The distance between *G. nobilis* and *G. variabilis* has been resized for clarity.

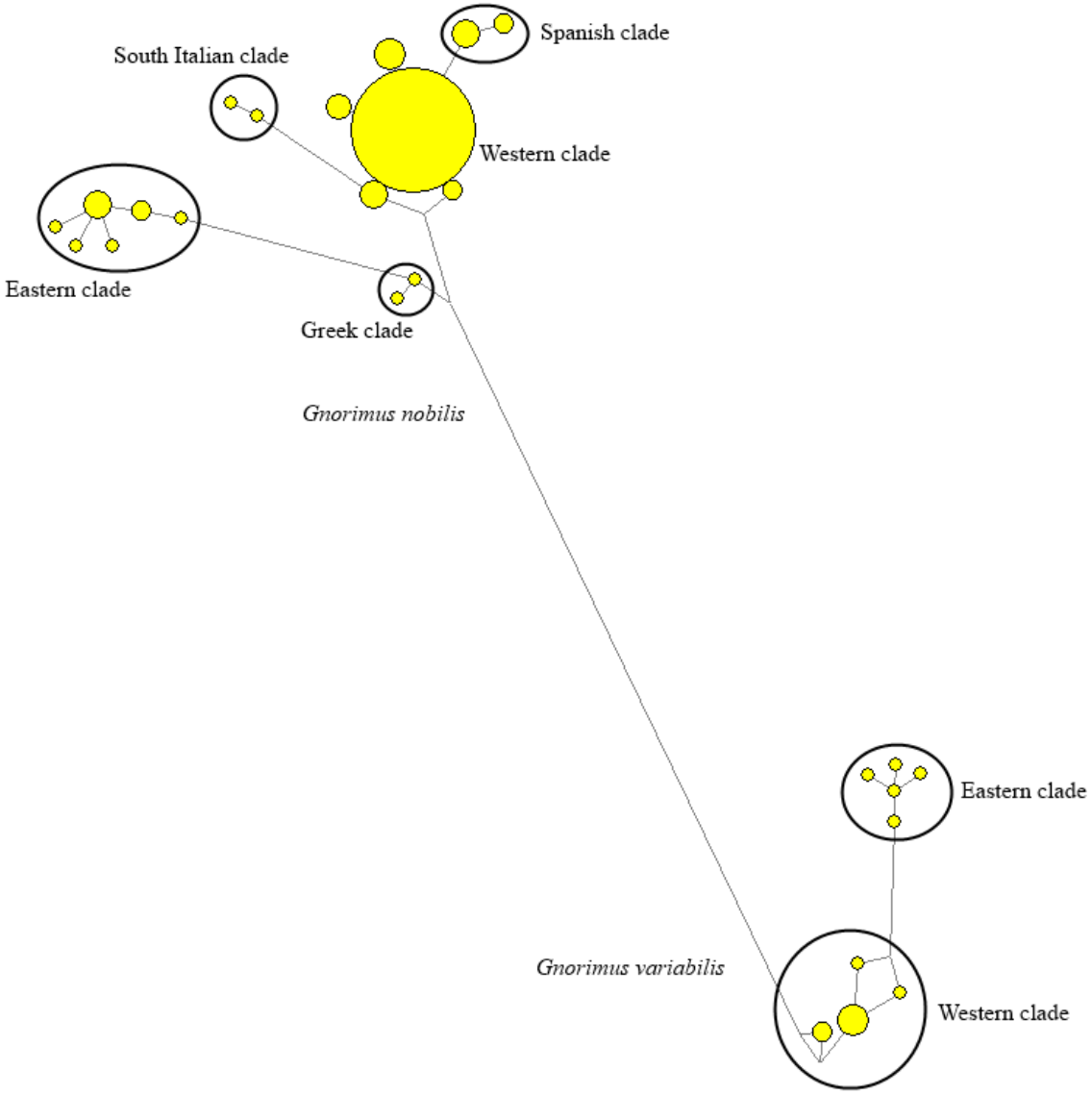


Figure 5.3. Maximum Likelihood tree of COI sequence variation in European *Gnorimus* (numbers at nodes indicate bootstrap support, 1000 replicates).

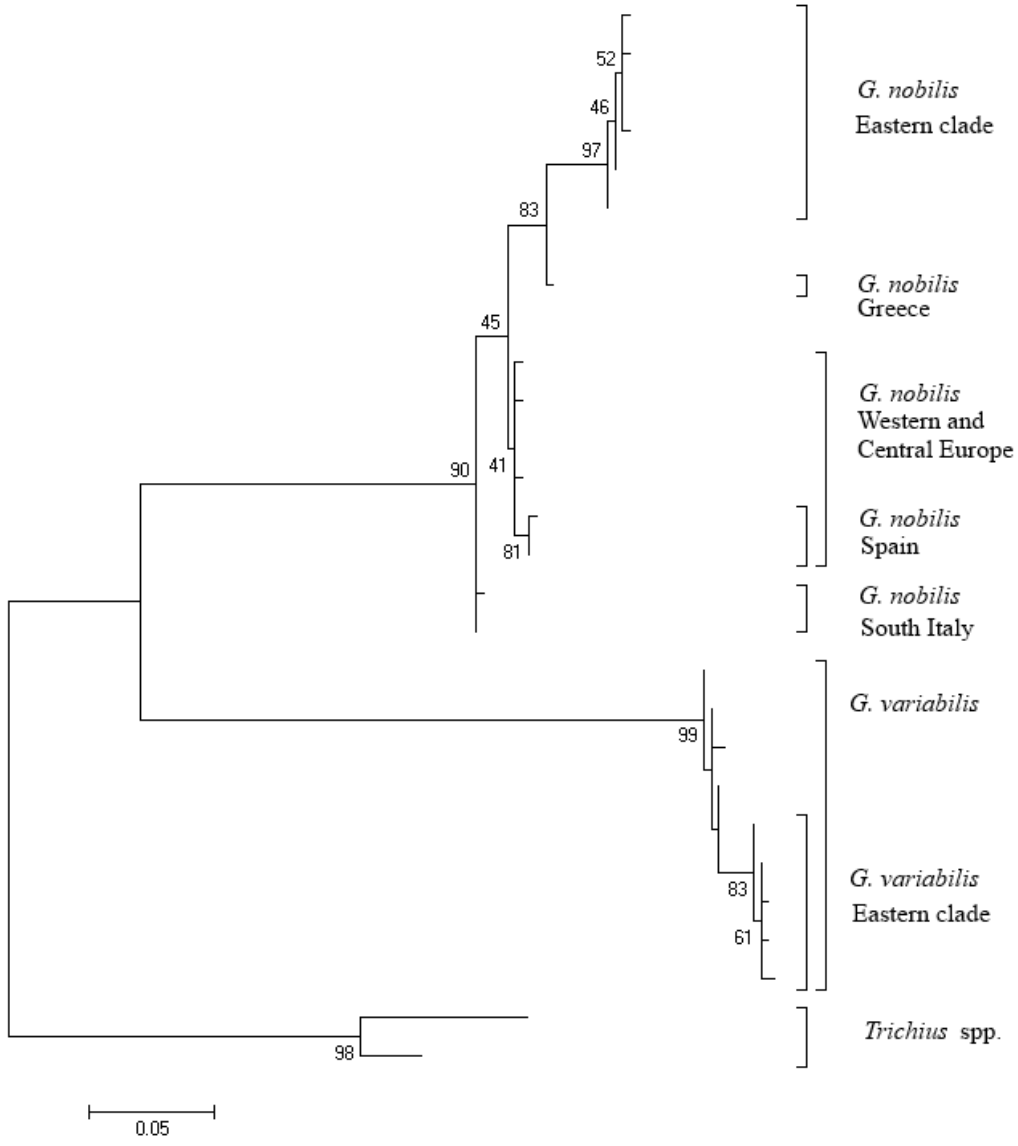


Figure 5.4. Neighbour Joining tree of COI sequence variation in European *Gnorimus* (numbers at nodes indicate bootstrap support, 1000 replicates).

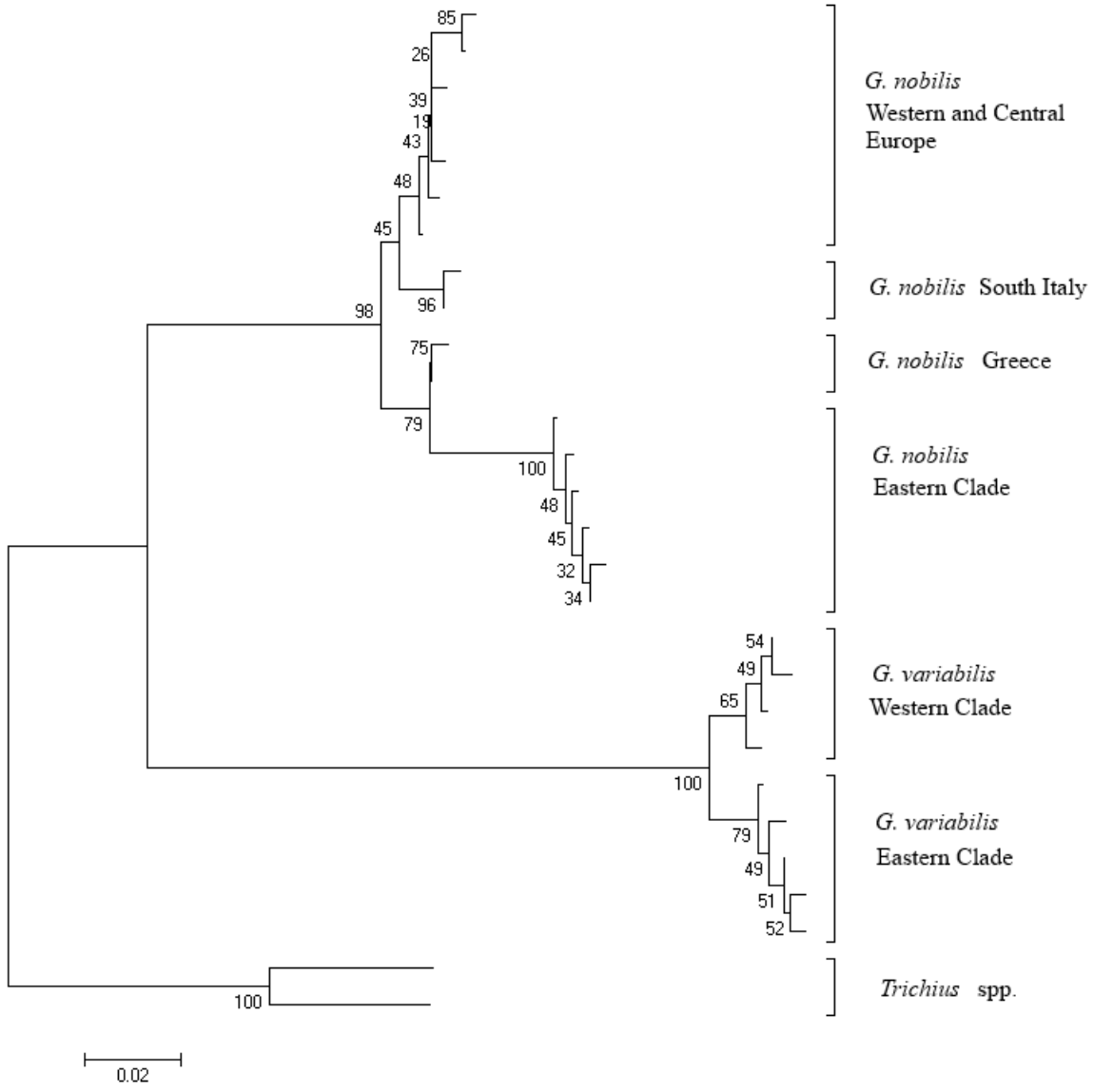


Figure 5.5. Bayesian Inference tree for COI sequence variation in European *Gnorimus* , built using MrBayes under the GTR+I+G model, run for 1,000,000 generations, with a burn in of 10% and with Markov chains sampled every 1,000 generations . All other settings were retained as defaults. Figtree used to display the consensus tree and displayed with posterior probabilities.

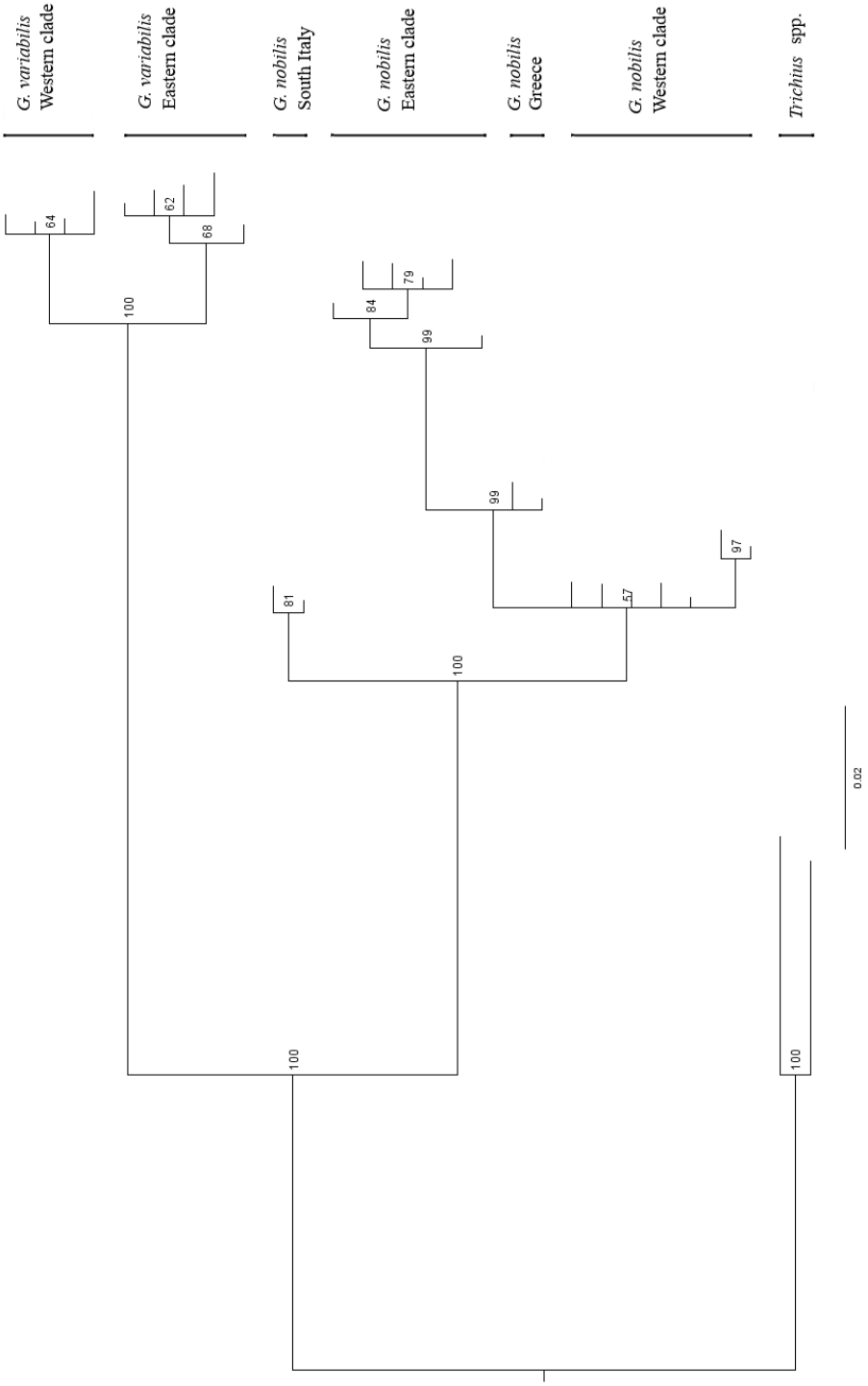


Figure 5.6: Distribution across Europe of simplified haplogroups from mitochondrial DNA (COI) for *Gnorimus nobilis* and *G. variabilis*.

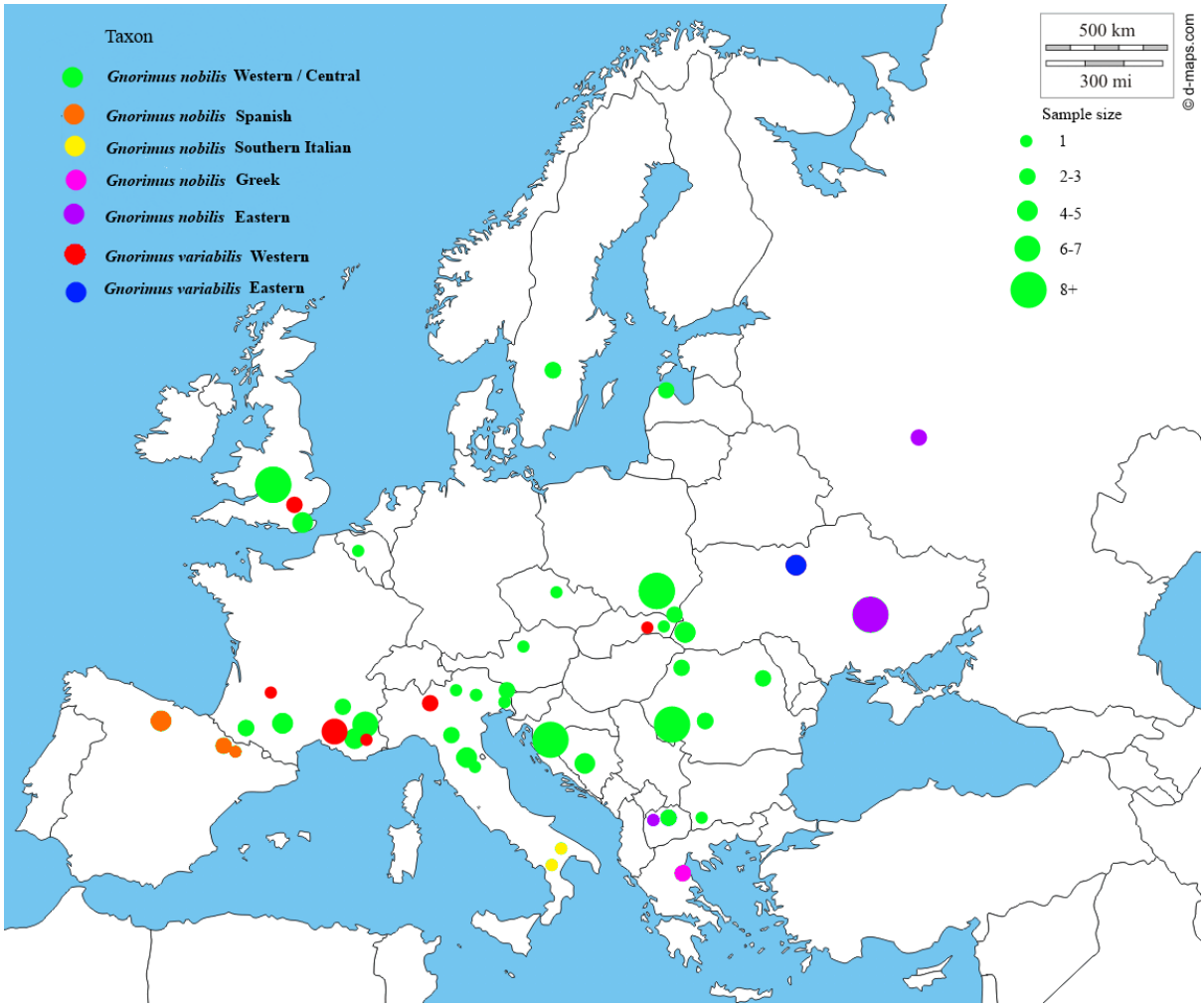


Figure 5.7. Haplotype network for *Gnorimus nobilis* Control Region: branch lengths are proportional to number of mutational changes; circle size proportional to number of individuals displaying each haplotype.

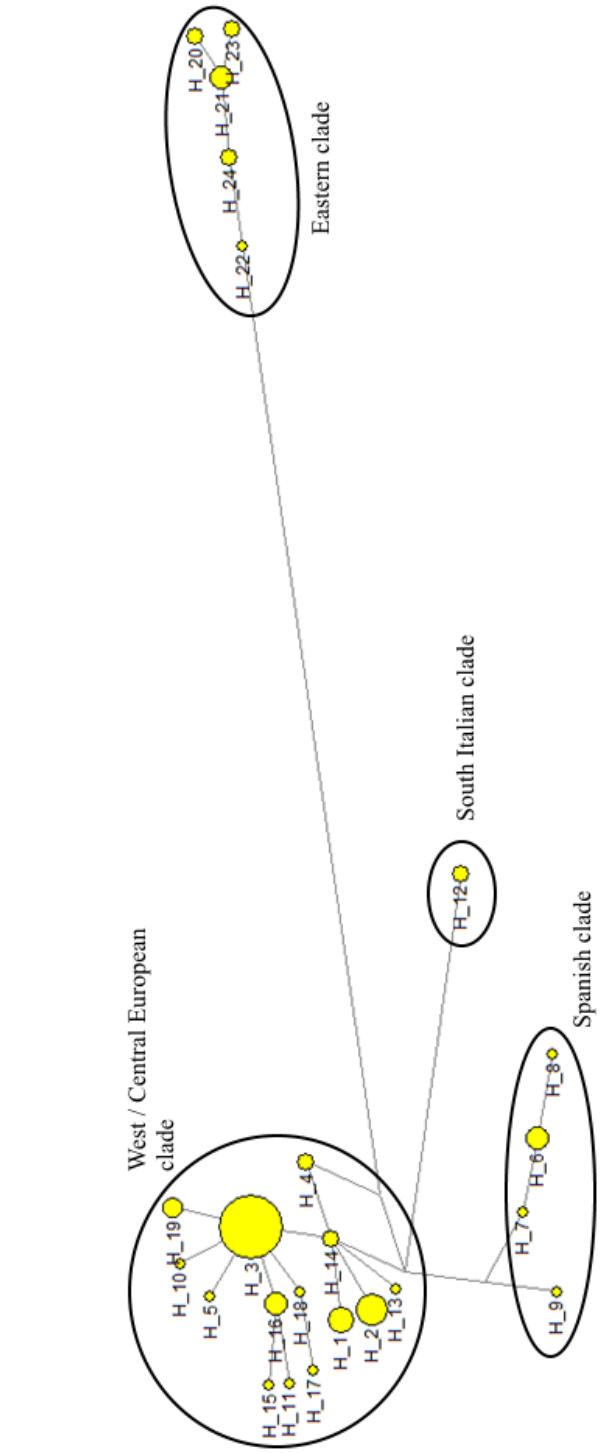


Figure 5.8. Maximum Likelihood tree of CR sequence variation in European *Gnorimus nobilis* (numbers at nodes indicate bootstrap support, 1000 replicates).

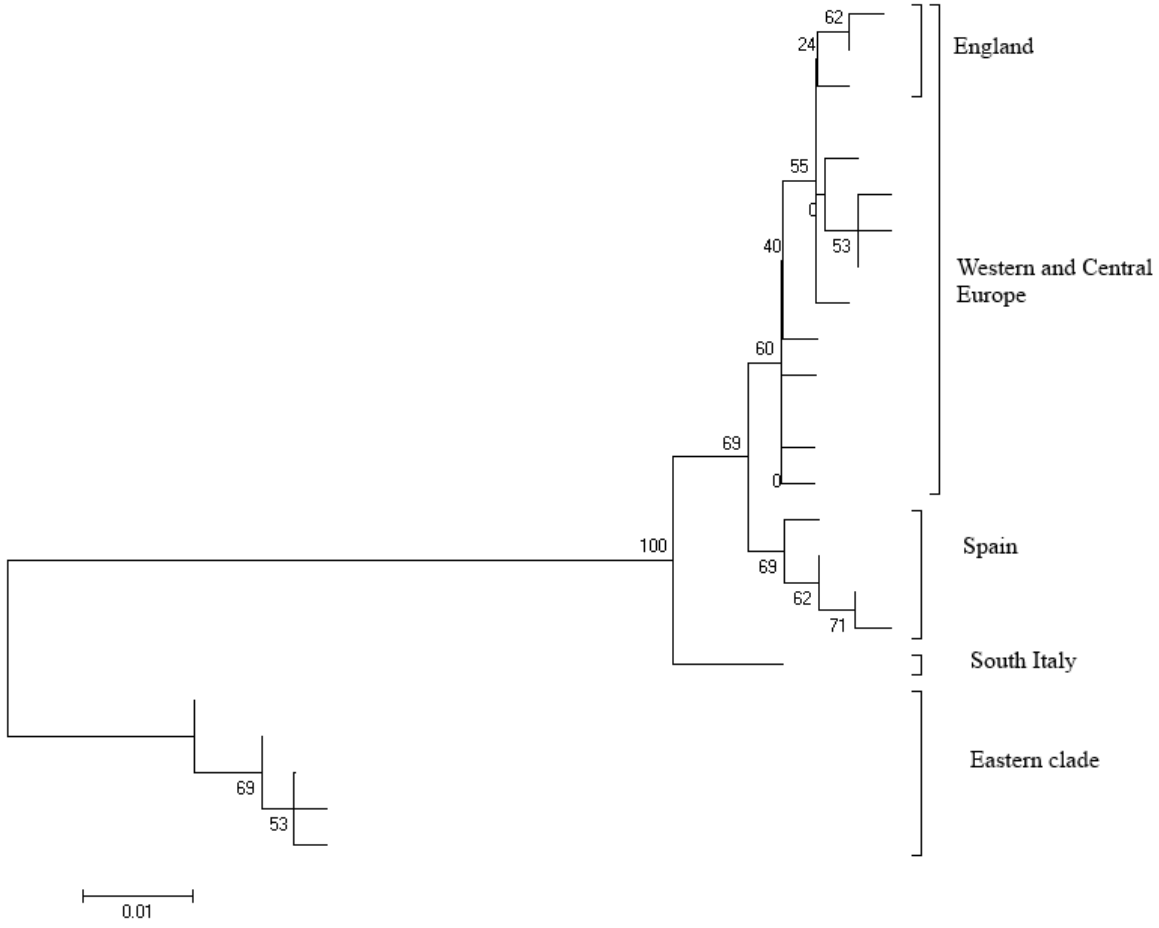


Figure 5.9. Neighbour Joining tree of CR sequence variation in European *Gnorimus nobilis* (numbers at nodes indicate bootstrap support, 1000 replicates).

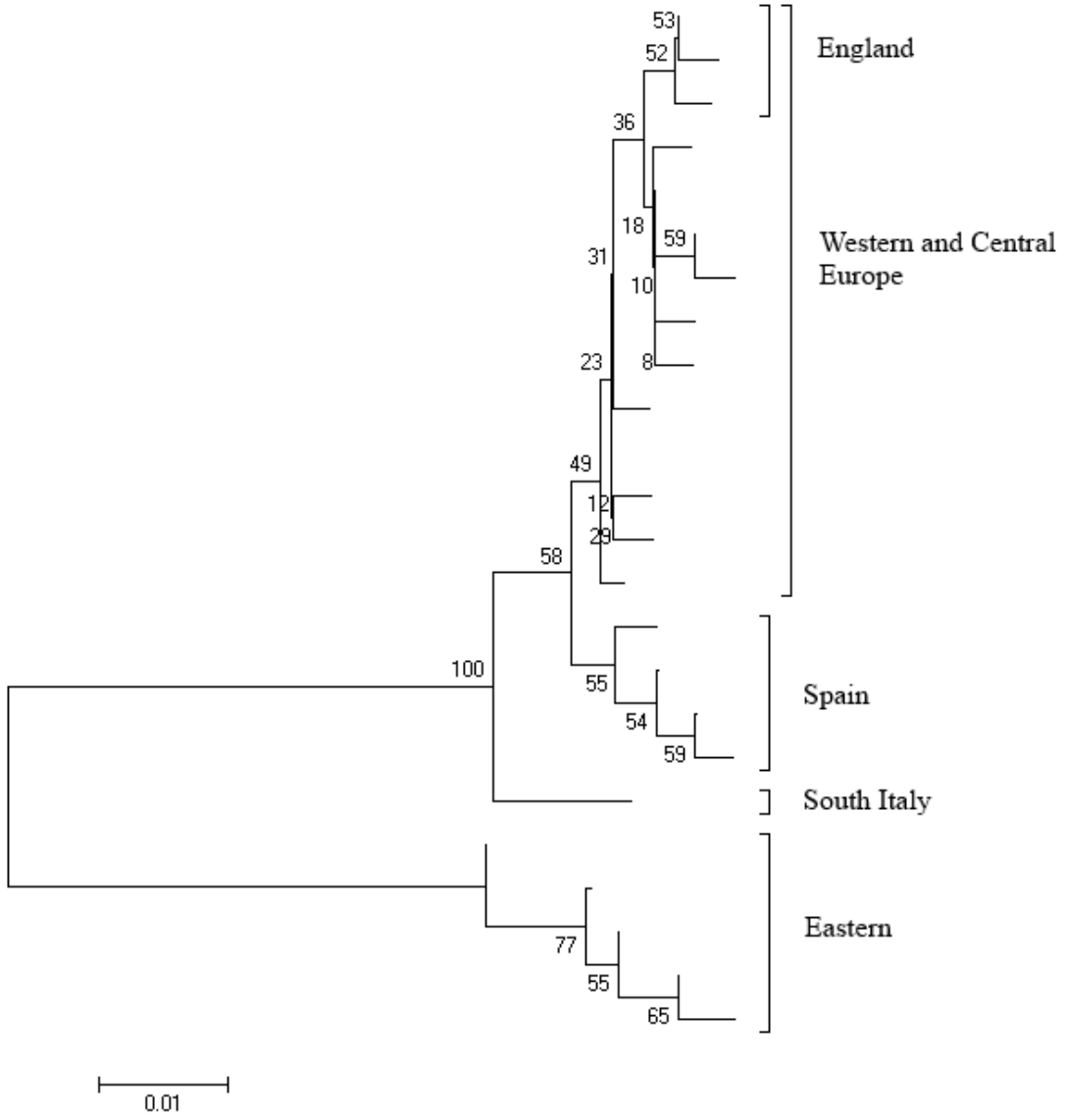
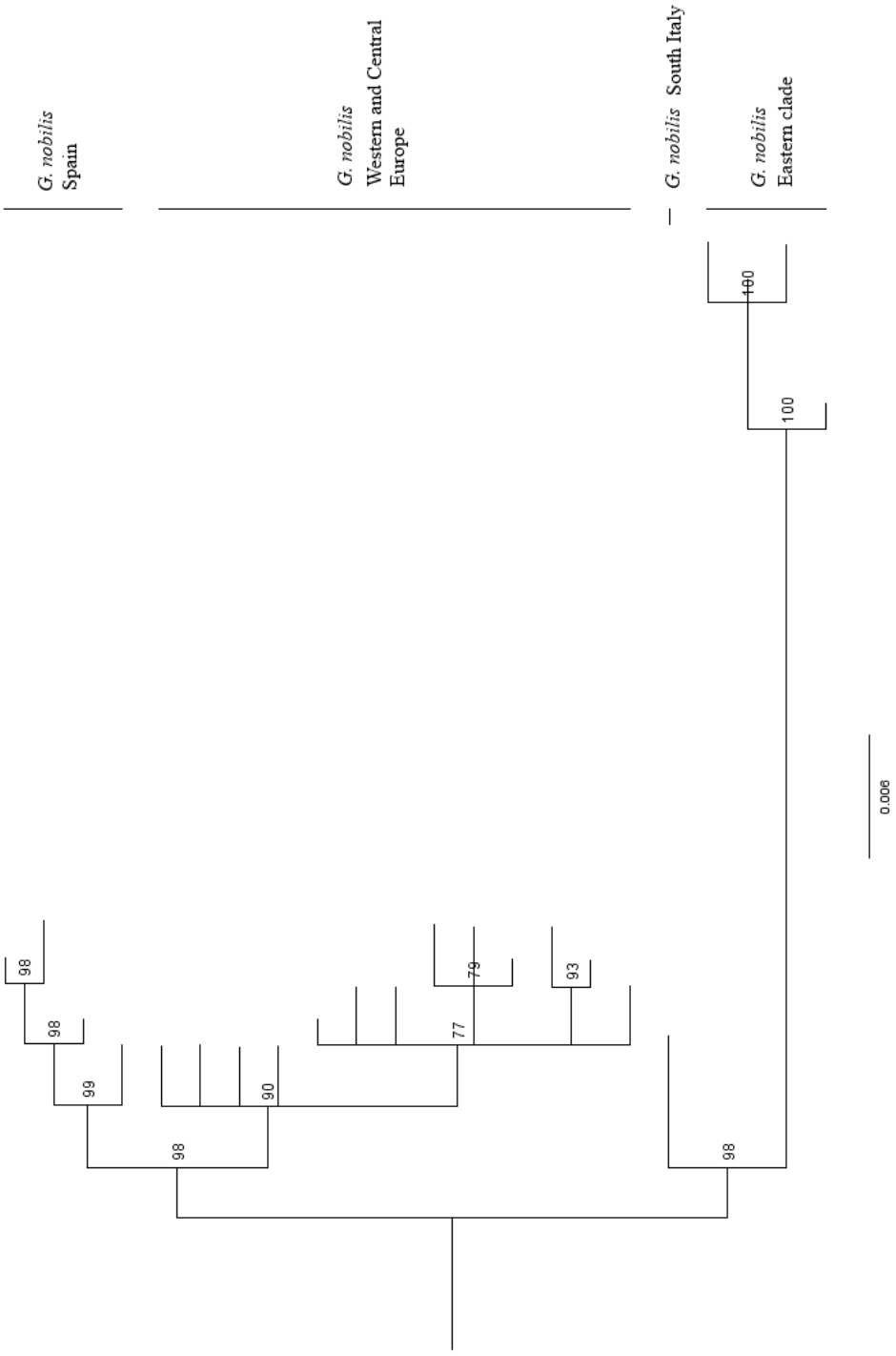
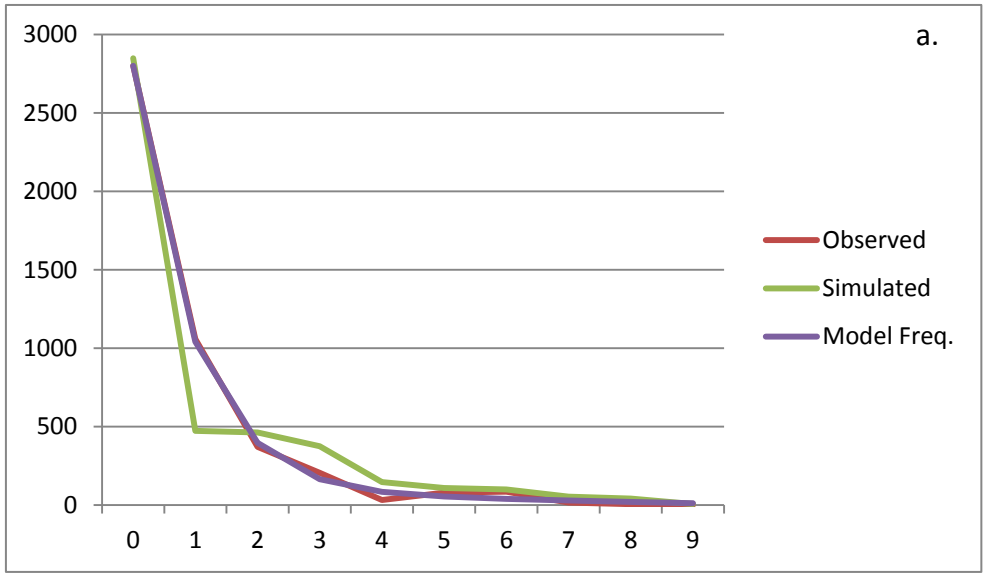
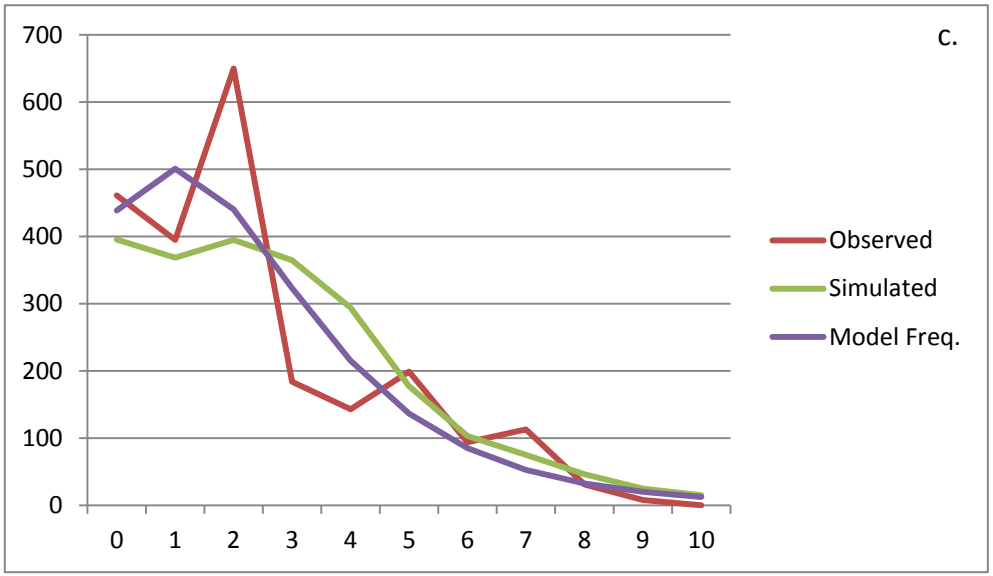
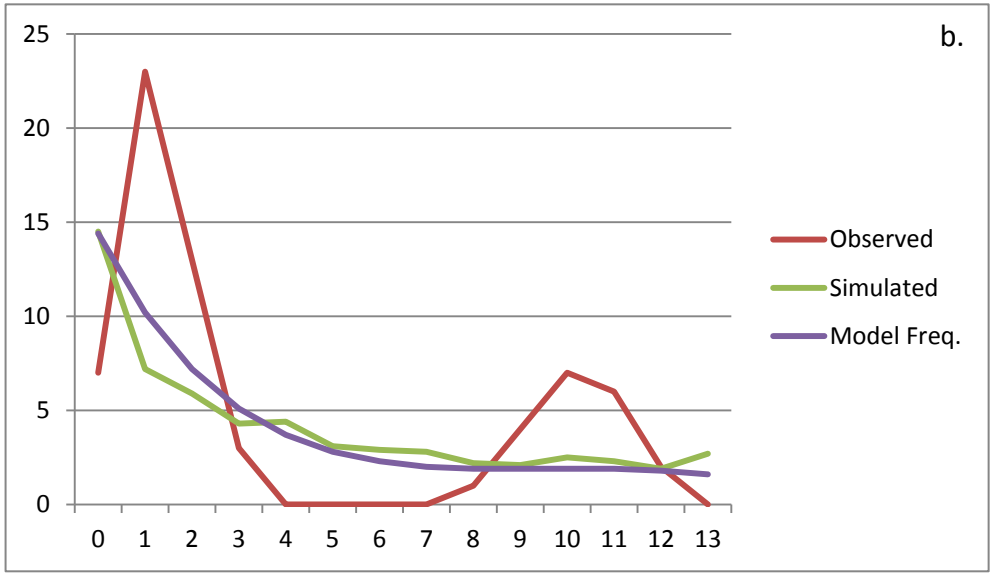


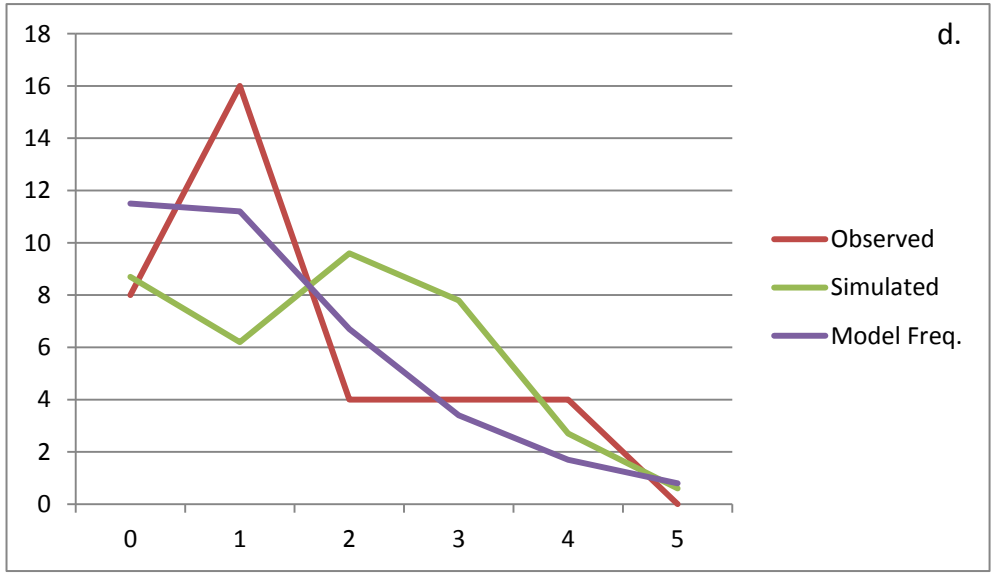
Figure 5.10. Bayesian Inference tree for European *Gnorimus nobilis* using Control Region. Built using MrBayes under the GTR+I+G model, run for 2,000,000 generations, with a burn in of 10% and with Markov chains sampled every 1,000 generations. All other settings were retained as defaults. Figtree used to display the consensus tree and displayed with posterior probabilities.




a.
3000
2500
2000
1500
1000
500
0
0 1 2 3 4 5 6 7 8 9
Observed
Simulated
Model Freq.


c.
700
600
500
400
300
200
100
0
0 1 2 3 4 5 6 7 8 9 10
Observed
Simulated
Model Freq.


b.
25
20
15
10
5
0
0 1 2 3 4 5 6 7 8 9 10 11 12 13
Observed
Simulated
Model Freq.


d.
18
16
14
12
10
8
6
4
2
0
0 1 2 3 4 5
Observed
Simulated
Model Freq.

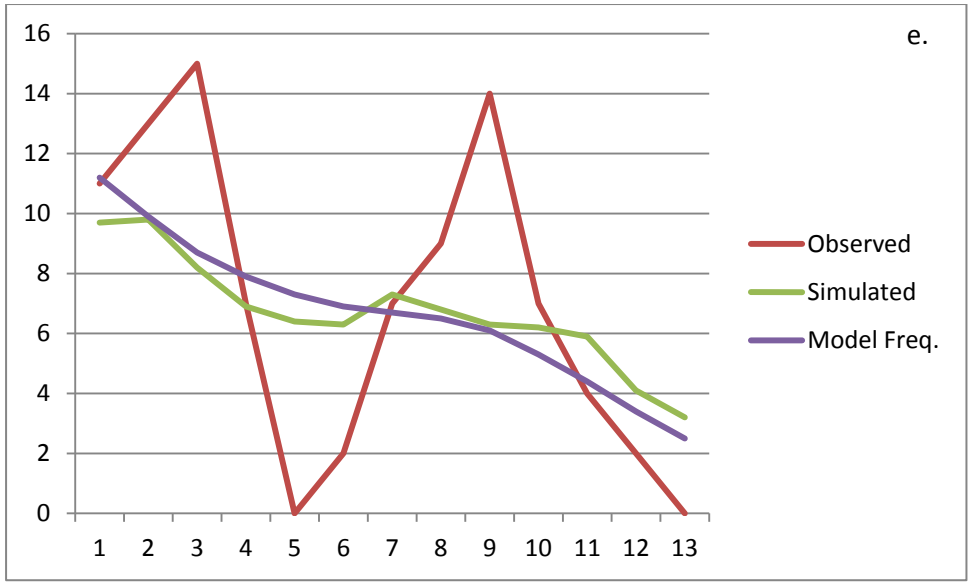


Figure 5.11. Observed and simulated MtDNA gene sequence mismatch distributions within *Gnorimus* clades: A) *G. nobilis* COI Western Clade; B) *G. nobilis* COI Eastern Clade; C) *G. nobilis* CR Western Clade; D) *G. nobilis* CR Eastern Clade; E) COI *Gnorimus variabilis*. Frequency appears on the Y axis in all graphs, and pairwise differences appear on the X axis.

Figure 5.12. Allele network for Heat Shock Protein 70 from data produced by PHASE, assuming no recombination. The colours of each allele represent the COI haplogroup to which the individual was assigned: yellow = Western *Gnorimus nobilis*, Green = Eastern *G. nobilis*, Blue = *G. variabilis*

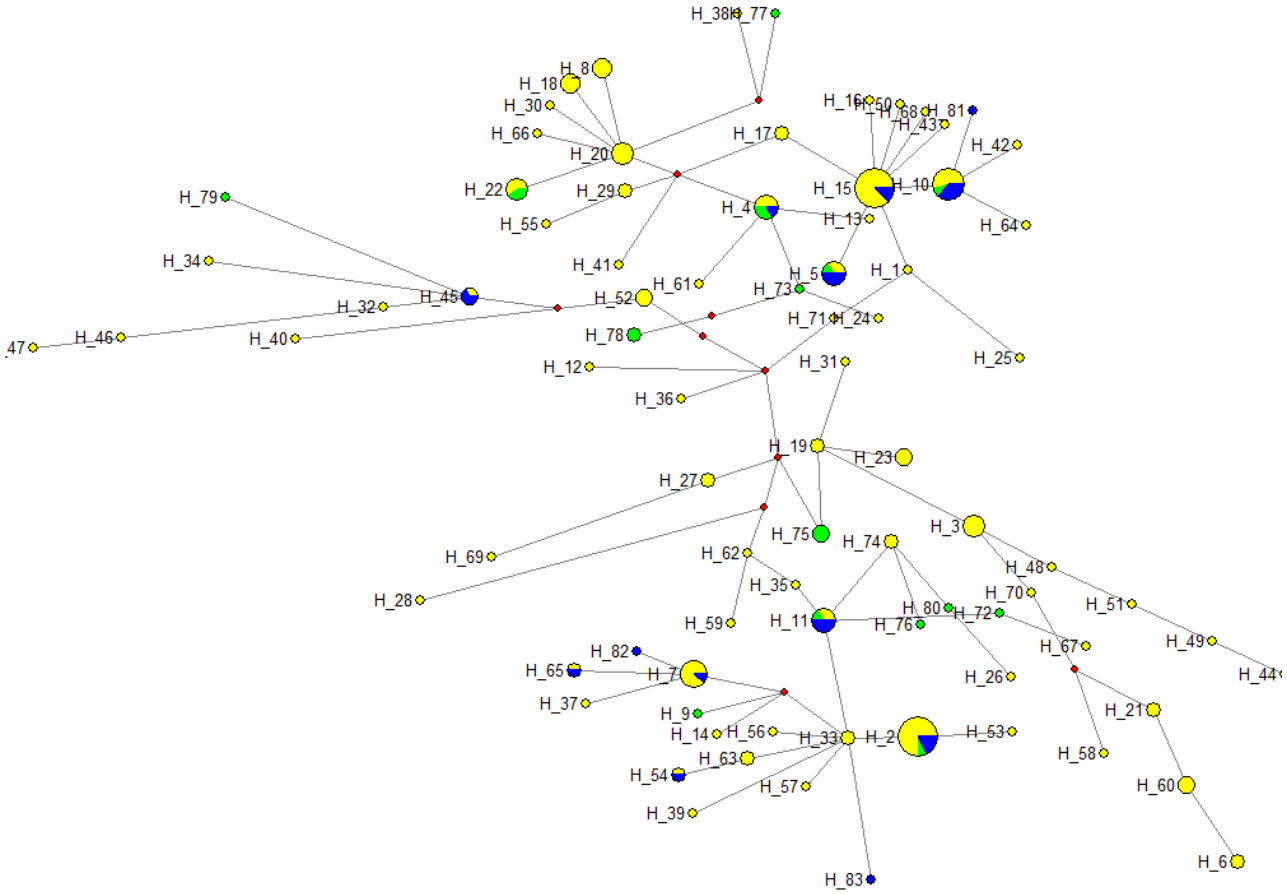


Figure 5.13. Allele network for *Orco* from data produced by PHASE, assuming no recombination. The colours of each allele represent the COI haplogroup to which the individual was assigned: Yellow = Western *Gnorimus nobilis*, Green = Eastern *G. nobilis*, Blue = *G. variabilis*, Orange = *Gnorimus subopacus*

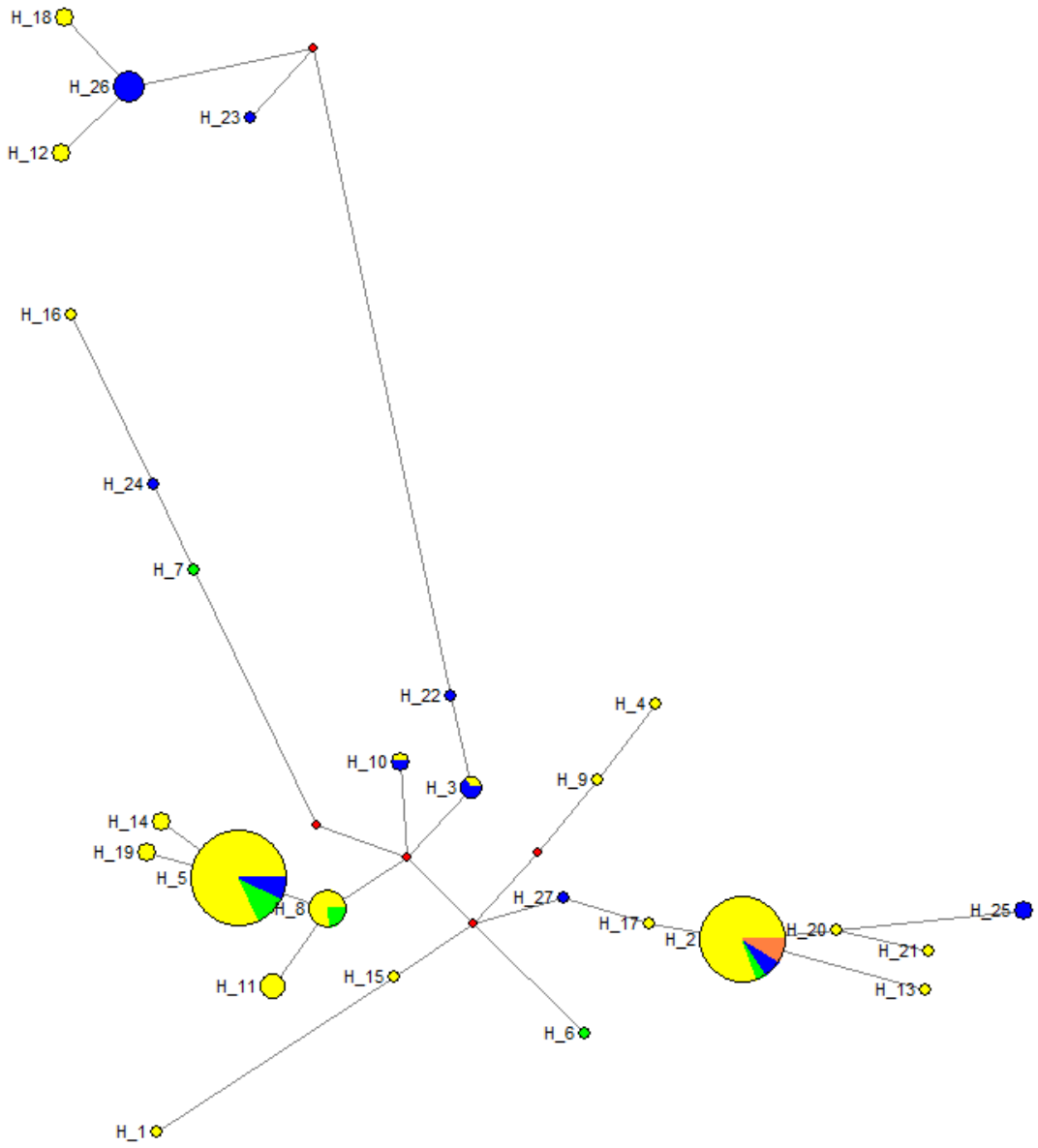
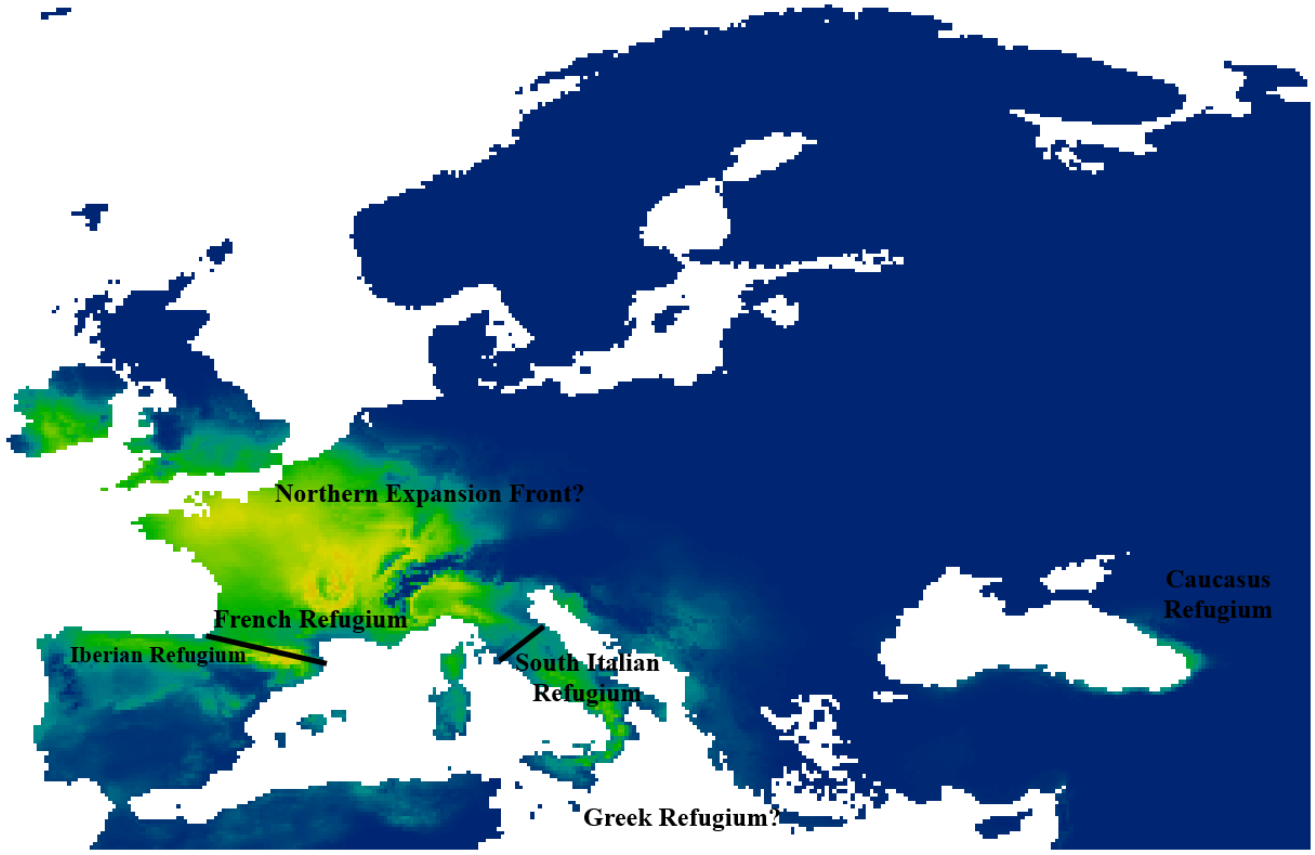
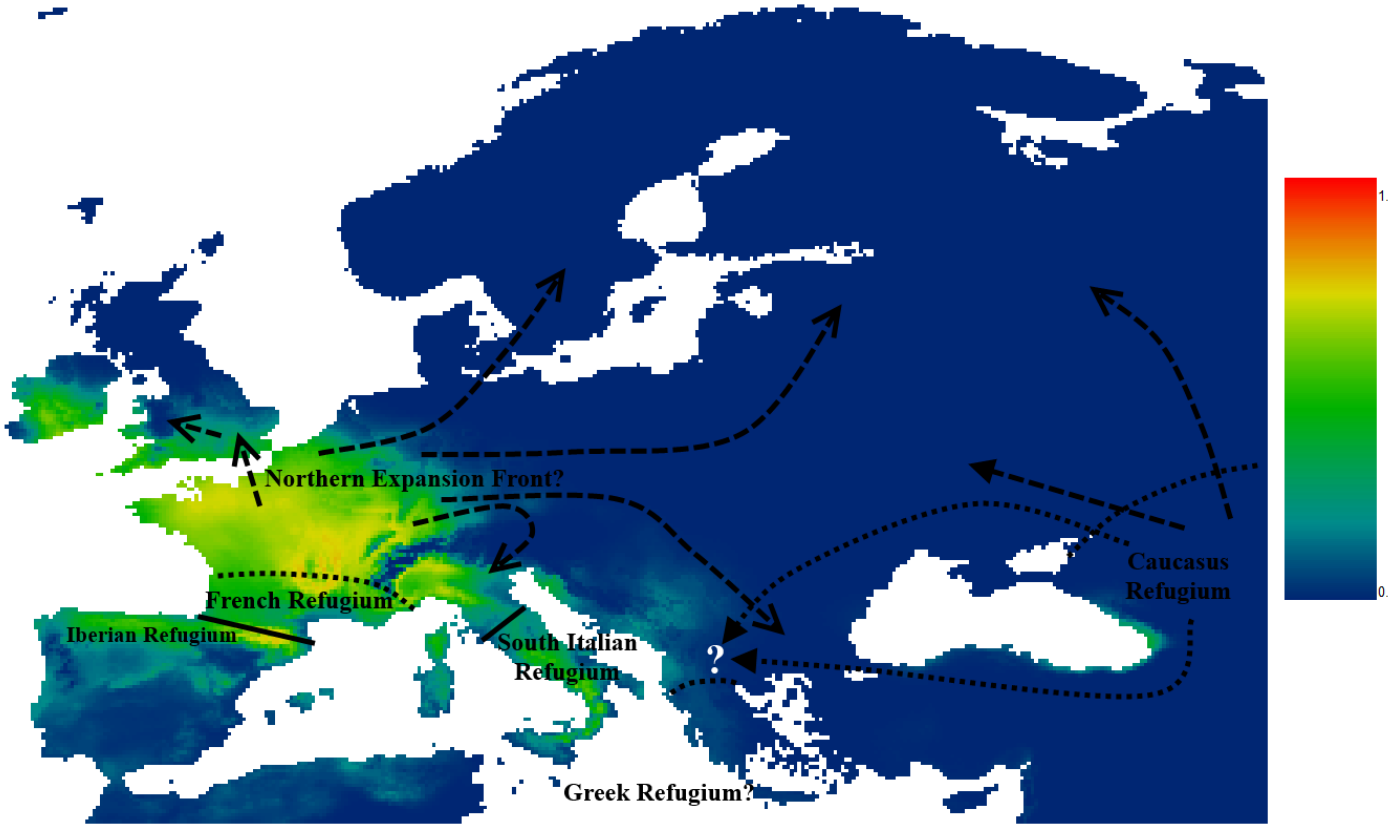


Figure 5.14.Hhypothesised refugial areas during Last Glacial Maximum (22KYA) for *Gnorimus nobilis*, based on phylogeography and species distribution modelling (Chapter 7). Black lines represent potential geographic barriers to populations



.

Figure 5.15. Hypothesised expansions from glacial refugia based on Fig. 5.14 above. Dotted lines represent possible limits to glacial refugia, dashed arrows represent likely directions of post-glacial expansions that are strongly predicted from genetic data, and dotted arrows represent possible directions of expansions that are weakly predicted from genetic data to explain the Eastern *Gnorimus nobilis* haplotype being found in Macedonia.



Chapter 6

DNA Sampling of rare and / or elusive species

6.1 Introduction

Population or conservation genetic studies of rare and/or elusive species require, like all genetic studies, samples containing intact DNA from the species in question. Whereas many studies can quickly and easily sample individuals from populations of common, diagnosable and perennial taxa, work on rare / elusive species is hampered by the small numbers of the species available, the usually patchy distribution of rare species, and the difficulty in finding elusive taxa (Hanski 1982; Magurran & Henderson 2003; Piggott & Taylor 2003; Mouillot et al. 2013). Conservation genetic studies on rare species also should avoid reducing the reproductive potential of the population (either by removing animals, or by removing tissue and reducing their fitness), and thus need to use well considered (and usually restrictive) minimal sampling strategies. In a similar manner studies on elusive taxa, regardless of rarity, can be difficult to achieve due to the difficulty in reliably sourcing and in an objective / randomised manner the individuals in question. Compared to standard studies, species that are both rare and elusive require alternative strategies to sample collection.

Non-invasive and non-lethal sampling techniques are extremely important for conservation genetics work on rare or elusive species. For example, use of non-living products from individuals (e.g. frass/faeces, hairs, shed exoskeletons) can provide DNA from which basic data such as presence and distribution of a species in an area (using DNA sequencing of universal mtDNA gene markers) or more detailed data on population demographics or structuring (using individual-specific markers such as microsatellites) can be determined. Examples of population data assessable by non-invasive sampling includes relatively 'simple' parameters such as targeted presence of a species in an area (with species-specific barcodes), up to estimates of population demography including effective population size, number of per-generation breeders and structuring and gene flow (Waits & Paetkau 2005; Schwartz et al. 2007; Pierson et al. 2015). Additional advances have also been made in environmental DNA sequencing (targeted to DNA molecules free in environmental samples), and there may be considerable overlap between the techniques used in eDNA sequencing and non-invasive sampling (Bohmann et al. 2014; Jones & Good 2016). In particular, studies on animal by-products don't have to find the study species itself; they just require that the by-product is easier to find and more abundant than the animal.

Targeting species of interest can be done in two main ways: by utilising environmental DNA (eDNA) and metabarcoding to determine community structure and identify rare species

presence (Jerde et al. 2011; Taberlet et al. 2012; Zhan et al. 2013; Thomsen & Willerslev 2015), or by specifically targeting a species and its by-products (Morin et al. 1993; Kohn & Wayne 1997; Taberlet & Luikart 1999; Valière et al. 2003; Mondol et al. 2009). There has been a recent push toward refining DNA extraction techniques to successfully extract DNA non-invasively or non-destructively from arthropods following the first paper to use non-destructive (samples from animal 'by products', like faeces) techniques to extract DNA from butterfly frass (the hard pelleted faeces produced mainly by larvae) and exuviae (Feinstein 2004), and earlier proposals that such techniques would be useful on larger animals (Taberlet & Luikart 1999; Piggott & Taylor 2003).

In contrast, non-lethal sampling removes material directly from the study organisms (a leg, haemolymph, etc.) but without killing them (Table 6.1). Papers that have used these techniques on arthropods include those targeted to identifying pest species (Lefort et al. 2012; Strangi et al. 2013), or to aid conservation programs (Monroe et al. 2010; Scriven et al. 2013), with non-invasive work on frass (hard faecal pellets) also being used in both instances. However, only Strangi et al. (2013) have extracted DNA from frass *in-situ*, working with 100mg of frass from trees infested with the longhorn beetle *Anoplophora chinensis*, whilst the techniques of other workers involved capturing individual insects, then collecting the frass for preservation (Lefort et al. 2012; Scriven et al. 2013). Though non-invasive in the sense that the insects aren't harmed by the treatment, such studies do still impose a time cost to the individual insect by removing them from their environment. In the case of pest species, this may not be a bad thing, but if the reproductive adults of a species of conservation concern were being retained (even for a short period) then this could lead to a small reproductive cost to the individual, and thus to the species as a whole. These techniques requiring live, physical animals may also be difficult to utilise in studies on rare or elusive taxa. Therefore it is important for conservation genetic programs to critically assess the best approach to sample collection, considering a trade-off between endangering the life of the animal or reducing its fitness, collecting and preserving high quality DNA, and the effort required to collect said samples.

Population sampling of saproxylic beetles such as *Gnorimus nobilis* presents no problems legally as it is not protected in the UK (Chapter 1; Smith 2003). However, the adults are hard to find in the wild, and considering the species is BAP listed, lethal or even invasive sampling should be avoided. Larvae are often found when an inhabited tree is cut down, but tests on the related Cetoniid *Pachnoda marginata* have shown that though larvae do survive leg

amputation and this yields enough DNA to use in PCRs to produce products of at least 330bp, larvae pupate 4-5 months (≈3000-3750 growing degree days at 25°C) later than control group siblings (Blake *pers. obvs.*). In the wild this would result in the reproductive imago emerging out of sync with other adults in the area, and it would not contribute reproductively to the next generation. So, non-lethal sampling of larval legs is not recommended.

Beetle adults are more robust than larvae, and non-lethal sampling by removing a leg should not cause any long-term viability problems, and has been trialled successfully in other beetles (e.g. Keller et al. 2005; Drees et al. 2008; Vila et al. 2009). The DNA extracted is generally of high quality, and can be used to provide COI amplification products of over 700bp. However, leg amputation is still invasive and so if a suitable alternative is available it should be considered. Beetle larvae produce copious quantities of frass (hard, pelleted faeces), which is used as an indicator for the presence of *Gnorimus* larvae within a veteran tree during survey work (Whitehead 2003), and may be expected to contain some beetle DNA. Research on captive colonies of *Pachnoda marginata* has shown that at a constant temperature of 25°C, the larvae produce 64 frass pellets per day on average (Blake *pers. obvs*.), and so the similar *Gnorimus* should be producing enough frass to represent a viable alternative source of DNA for conservation genetics studies. *Gnorimus* larvae are the only scarabs found in orchard habitats in the UK, and are easily identified (Shabalin & Bezborodov 2009); additionally the frass they produce is lozenge shaped, not rounded as in other Cetoniids (Blake *pers. obvs.*) so confusion between species is highly unlikely.

Frass is usually dry, rarely fragments during transport or storage, and can be stored dry in bags in the freezer. Other work has isolated DNA from frass (Table 6.1), though no frass tested was older than 7 days (Lefort et al. 2012). Seeber et al. (2010) tested DNA extracted from millipede frass up to five days old for PCR amplification with the universal Folmer COI primers (710bp; Folmer et al. 1994), finding a significant decrease in the number of positive PCRs as sample age increased. Collecting frass is expected to yield lots of low-quality genetic material, but for little fieldwork effort. Collecting a sample of frass from a tree could allow a genetic profile of the beetle population within the tree to be produced, though it is unknown how long frass stays intact in a tree. Because of this, the age of the frass in a tree could vary from a day to a year (or more).

In addition to recently collected non-invasive samples (frass), samples of *G. nobilis* can come from a number of other sources which have been collected without consideration of DNA preservation (such as being flash frozen, dried or placed into ethanol (Post et al. 1993; Moreau et al 2013)), and so methods for extraction of DNA from such non-optimal sample sources are needed. Most endangered populations and species have only become so within the last century, which coincides with the period of greatest focus on sampling species for museum collections (Wandeler et al. 2007), and so such collections represent a valuable source of samples as a historical time series. Studies using museum specimens have detected population bottlenecks and microsatellite alleles not present in modern populations (Harper et al. 2006), helping to date changes in population genetics. However older specimens typically have extremely degraded DNA which causes issues when barcoding (Van Houdt et al. 2010), with studies generally focussing on mitochondrial DNA (with a high copy number per cell) or short sections of nuclear DNA (Navascués et al. 2010).

Sometimes locally rare taxa may not be so on wider scales. *Gnorimus nobilis* and *G. variabilis* are both more common on European scales than they are in the UK (Smith 2003; Mannerkoski et al. 2010a., 2010b.), which means that finding individuals at European scales should be easier than within the UK, thus being useful for phylogeographic studies. As such, scientists and private collectors are also useful resources for specimen collection, and so represent a further source of samples that usually have not been collected with DNA preservation in mind (i.e. dried and mounted). Though the trade in beetles (large, showy scarabs in particular) is largely unregulated, there is no evidence that localised collection for trade has a major impact on locally common species. However, the world-wide trade on such species has the potential to pressure rare populations already threatened by loss and degradation of habitat, especially for those subjected to strong natural Allee effects (positive correlations between population density and individual fitness, e.g. stag beetles (Courchamp et al. 2006; Tournant et al. 2012; Verma 2016)) and thus the costs and benefits of using the private trade should be outlined for a particular species, and other options investigated, before starting a study (Muafor et al. 2012).

Sampling non-invasive material for *Gnorimus* is relatively straight forward and safe (Fig. 6.1) as the beetle is not dangerous to humans and there are no potential threats to humans living in the rot-holes in the UK. However, other taxa are considerably more dangerous, and direct contact should be avoided. Spiders (Araneae) are venomous arthropods, some of which are quite capable of killing a person (Finkelstein et al. 1976; King & Hardy 2013), but they also

produce copious quantities of conspicuous biological material in the form of silk webs that represent a potential source of DNA for genetic studies. Silk is a fibrous biopolymer which is a plesiomorphic character within Araneae (Vollrath 1999) and has been widely adapted for various purposes within the group (Vollrath 2000, Blackledge et al. 2009, Blackledge 2012). Assessing how useful this material is for genetic studies will be useful to help barcode via non-invasive means some of the 44,000 known spider species. Though poorly researched from a conservation genetics standpoint, of the species that have been assessed for the IUCN Red List (162 species; IUCN 2015), 132 species (81.5%) are listed as Vulnerable or worse (considerably above the average for terrestrial Arthropoda where 24.9% of the 5018 assessed species are Vulnerable or worse (IUCN 2015)). As predators usually specialising on Arthropods, loss of spiders could result in ecosystem destabilisation and the wild fluctuation of prey numbers, including crop pests (Yen 1995; Cardoso et al. 2004; Crespo et al. 2014; Hendrixson et al. 2015).

The aim of this chapter was to investigate potential sources of DNA from non-invasive methods of sample collection from endangered or elusive arthropods, and the molecular genetic markers that can be applied to said DNA source material: from frass as a DNA source for scarab beetles, and from web silk for spiders.

6.2 Methods

6.2.1 DNA extraction from beetle frass

Three DNA extraction methods were assessed: a Chelex based method following McKeown & Shaw (2008); DNeasy Blood and Tissue Kit (Qiagen); and a CTAB-phenol-chloroform-isoamyl alcohol method (Winnepenninckx *et al*. 1993). Each method was tested with 8 pieces of frass from three different sites. Frass from third instar *G. nobilis* larvae weighs on average 2.2mg (SD=0.51) with a range of 1.1-3.2mg (n=30, dry weights from a single locality). All frass was washed in distilled water, dried by blotting on lab roll, then crushed before digestion.

For the Chelex extraction, pieces of frass were incubated with 50µl of 5% chelating resin (Chelex®, Bio-Rad) and 10 µl of Proteinase K (10 mgμl^{-1}) at 55°C for 3 h, before being boiled (at 100°C) for 10 min. Samples were then centrifuged at 13,000RPM for 10 min, and

the supernatant removed and stored at –20°C. For the DNeasy Blood and Tissue kit, the 'Purification of Total DNA from Animal Tissue' protocol (see Qiagen instructions) was followed including repeating the elution.

For the modified CTAB-Phenol-Chloroform-Isoamyl alcohol protocol, frass was initially placed into a 1.5ml Eppendorf tube, 350µl of CTAB buffer and 10µl of 5µg/ml Proteinase K added, followed by 10s on a vortex. This was incubated overnight (18-22 hours) at 37°C. 350µl of equilibrated 6.7/8.0 pH Phenol-Chloroform-Isoamyl alcohol was added into each tube before being shaken by hand for 10 minutes. The tubes were then centrifuged for 10 minutes at 13,000RPM, after which the top layer was pipetted into a new tube, and the waste discarded. 990µl of 100% Ethanol was added to the new tube, and the solution was incubated at -20°C for at least two hours. This was then centrifuged at 13,000RPM for 10 minutes; the liquid solution poured off, with small remaining volumes removed using a pipette. A 70% ethanol wash was then completed by adding 990µl of 70% ethanol to each tube, then incubating the solution at -20°C for at least two hours. This was then centrifuged again at 13,000RPM for 10 minutes; the liquid solution poured off, with small remaining volumes removed using a pipette DNA pellets were then dried in open tubes in a fume cupboard for 20 minutes, after which 50µl of distilled water was added. DNA was then left overnight at 4°C to go into solution.

For all three methods all extractions were then assessed by running 5µl DNA stock on 1% agarose gels using GelRed (Biotium) as a staining agent in a concentration of 1µl per 100ml of agarose solution. This staining agent was used throughout the work.

6.2.2 DNA extraction from spider web

Two spider species were chosen for study which utilise different web types. *Psalmopoeus cambridgei* (Theraphosidae) produces a vertical sheet web in enclosed spaces in trees, which is then covered in loose material surrounding the web structure; primarily detritus and leaves (Bushell *pers. comm*.). *Pholcus phalangioides* (Pholcidae) builds 'space webs' which are used as prey-detection structures from which the spider hunts prey (Jackson & Brassington 1987). The *Psalmopoeus* was a captive specimen, but the *Pholcus* samples were all collected from inside a house in Wales (52.4113N, -3.9897W), a natural wild habitat for this species. Webbing was cleaned of large particles of detritus (including exoskeletons from prey or the web holder), but the majority of the fine detritus (small pieces of prey, faeces from the spider,

local substrates, etc.) remained stuck to the web. Webbing was then cut into pieces (n=6 for *Psalmopoeus*, n=8 for *Pholcus*) to give individual sample weights of 2.3-8.5mg. Only the CTAB-Phenol-Chloroform-Isoamyl alcohol method was tested for the webbing, following the protocol above. All eluted DNA was assessed by running 5µl DNA stock on 1% agarose gels, and by quantification by running the DNA in a Nanodrop 2000 (ThermoScientific).

6.2.3 DNA extraction from adult beetle specimens

The aim of this test was to assess the facility to extract and amplify/sequence DNA from adult beetles from historic collections, i.e. that had been collected and preserved (usually air dried) under "non-optimal" conditions for DNA studies (Supplementary tables 1 & 7).

For the *Gnorimus* and *Trichius* specimens, phenol-chloroform-isoamyl alcohol protocols were followed (see Chapter 5 and 3 respectively). Specimens were collected from a variety of sources, including other researchers, commercial insect suppliers, and museums. Sample ages ranged from one year to 58 years old. No *Trichius* was preserved specifically for DNA-extraction purposes, whilst 12 *Gnorimus* were preserved in 100% ethanol immediately after being killed to preserve their DNA to act as positive controls. These positive controls were not over two years old when sequenced. All DNA was assessed by running 5µl DNA stock on 1% agarose gels.

6.2.4 Microsatellite development

Microsatellite loci were also identified using the same RAD-seq and enrichment protocols as outlined in Chapter 2. A total of nine primer pairs were designed and tested on the samples from Chapter 5. All loci produced PCR product of expected sizes, however, with the exception of Locus NC-SSR-CT1 (Supplementary Material) no allelic variation was detected. Locus NC-SSSR-CT1 also revealed extremely low levels of variation among samples from Chapter 5 and so was not included in that analysis. However, this locus was used to assess the PCR amplification of microsatellite among frass. PCRs were performed following the same reaction mic and thermoprofile described in Chapter 2 but with the modification that cycle number was increased to 55.

6.2.5 PCR and sequencing protocols

To test the success and effectiveness of the three methods of extraction for recovering DNA that can be used in further procedures, eluted fractions were tested by PCR with species- or genus-specific DNA primers designed to produce specific amplicon lengths for the various target species. Successful amplicons were assessed for their ability to yield clean species-specific DNA sequences through standard sequencing protocols.

Frass extractions were diluted in a series from stock to 1/10, 1/100 and 1/200 with ddH_2O, and used in PCRs with *Gnorimus nobilis*-specific primers from Chapter 5, plus two internal primers , COIF213 (5'-TGTCTTCCTATTTACAGTGGG-3') and COIR249 (5'-ACGTAATGGAAATGAGCAACT-3'). COIF30 was used with COIR249 to amplify a 219bp section of Cytochrome Oxidase 1 (COI), whilst COIF213 was used with COIR546 to amplify a 333bp section of COI which slightly overlapped the other primer pair, allowing for the full 516bp section to be sequenced by concatenating the sequences. PCR mixes were: 20µl PCRs consisting of 10µl of Biomix (Bioline, London, UK), 1µl of each primer at 10µM, 5µl of dH_2O, and 3µl of diluted DNA. PCR conditions were 95°C/3 minutes, 55X (95°C/30s, 52°C/30s, 72°C/30s), 72°C/3 minutes.

PCR amplification and genotyping of *G. nobilis*-specific microsatellite DNA loci was also used to test extractions from frass material. Additional frass samples were extracted using the CTAB-Phenol-Chloroform-Isoamyl alcohol method from 12 localities, with four pieces of frass used in each (48 in total). Locus NC-SSR-CT1 was used.

Species-specific primers were designed for *Psalmopoeus cambridgei*, whilst within-genus primers were designed for *Pholcus phalangioides* (Table 6.2). The widely used universal invertebrate COI Folmer primers (710bp, Folmer et al. 1994) were also tested for both species. Spider web extractions were diluted to between 1/20 and 1/50. PCRs were performed in 20µl volumes consisting of 10µl of Biomix, 1µl of each primer (10µM), 3µl of DNA diluted from stock to 1/50, and 5µl of ddH_2O. Thermocycler conditions for all reactions were: denaturation at 95°C for 3 minutes, then 45 cycles of 95°C for 30s, 50°C (45°C for the Folmer primers) for 45s and 72°C for 45s, final extension at 72°C for 3 minutes.

For the adult *Gnorimus* specimens, species- or genus-specific primers were used for COI (512bp), Control Region (330bp) (both mitochondrial genes), *Orco* (150bp), Heat Shock Protein 70 (206bp) and a microsatellite locus (120bp) (three nuclear genes) (Chapter 5).

Using these primers, all *G. nobilis* and *G. variabilis* specimens were screened using 20µl PCRs consisting of 10µl of Biomix (Bioline, London, UK), 1µl of each primer at 10µM, 5µl of dH_2O, and 3µl of diluted DNA. PCR conditions were:

95°C/3 minutes, 55X (95°C/30s, 52°C/45s, 72°C/45s), 72°C/3 minutes for COI

95°C/3 minutes, 55X (95°C/30s, 54°C/30s, 72°C/30s), 72°C/5 minutes for Control Region

95°C/3 minutes, 45X (95°C/30s, 57°C/30s, 72°C/30s), 72°C/5 minutes for HSP70

95°C/3 minutes, 45X (95°C/30s, 52°C/30s, 72°C/30s), 72°C/5 minutes for *Orco*

95°C/3 minutes, 55X (95°C/30s, 52°C/30s, 72°C/30s), 72°C/3 minutes for the microsatellite

Because of issues with amplification from Control Region in *G. variabilis*, likely due to mutations in the primer binding sites, this species was not included in the CR dataset, nor was it included in the microsatellite dataset.

For the *Trichius* specimens, COI (461bp) with genus-specific primers was used, alongside *Wingless* (149bp) (a nuclear gene) – see Chapter 3. PCR conditions were:

95°C/3 minutes, 55X (95°C/30s, 55°C/45s, 72°C/45s), 72°C/3 minutes for COI

95°C/3 minutes, 55X (95°C/30s, 55°C/45s, 72°C/45s), 72°C/3 minutes for *Wingless*

Amplicons from all extraction stocks and primer combinations were checked on a 2% agarose gel, then cleaned with SureClean Plus (Bioline) following the manufacturer's protocol (but increasing the initial centrifugation step to 20 minutes), and sequenced with AB BigDye technology. Sequences of all genes were checked and edited in Chromas Lite (Version 2.1; 2012; Technelysium Pty Ltd), then checked for identity using the BLAST algorithm against the GenBank database. Mixed peaks were edited to include degenerate base codes if these were present (Chapter 3)

6.3 Results

6.3.1 Microsatellite development

25 microsatellites were tested from the RAD-seq library (representing approximately 10% of the available putative microsatellites), plus 45 from the enriched library. Some loci failed to amplify, others produced a mixture of different product sizes, and other putative loci were monomorphic. Approximate success rates (i.e. a putative microsatellite flanked by enough suitable base pairs for primer design) for sequenced clones in the enriched genomic library were one in every five sequences possessing a testable microsatellite, but of the 70 loci subjected to PCR testing, none were useable as microsatellites.

6.3.2 Frass extraction and sequencing

Success rates of extraction and PCR amplification are reported in Table 6.3. The Chelex method produced very low (almost zero) amplification success, whereas both DNeasy and CTAB-chloroform/isoamylalcohol extraction methods showed high success rates (90%+) at higher dilutions for the two primer sets. All sequenced amplifications of COI were aligned to reference material and confirmed as *Gnorimus nobilis* (compared to known adult specimens). BLAST results also confirmed the sequences as the sister group to *G. variabilis* (the only *Gnorimus* species on GenBank at the time).

DNA extracted from frass from all localities tested also was amplified reliably using the nuclear microsatellite locus, and confirmed that same pattern of a lack of variation as found for fresh samples of British adults and larvae tested in (Chapter 5), with only three of the 44 successfully amplifying samples showing alleles that were not 120bp homozygotes (all four were 118 / 120bp heterozygotes). The 118 allele was not recovered from any British adults or larvae, but had been found in Romanian, Spanish and Ukrainian individuals (n=4). Successful genotyping rates seemed to be affected by the sample locality: nine of the 12 sites had all four pieces of frass amplify successfully, two sites had one sample fail and one site had two samples fail.

6.3.3 Spider web extraction and sequencing

The concentration of the extracted DNA was 15.6-23.3ng/μL for the *Psalmopoeus* webbing, and 1.1-7.4ng/μL for the *Pholcus* webbing. Sequences returned were clean of mixed peaks, and identified to the correct species using the BLAST algorithm against the GenBank database. Samples from both species amplified and were sequenced with both their species- or genera-specific primers, as well as the longer 710bp Folmer primers (Fig. 6.2). All six *Psalmopoeus* web samples successfully underwent PCR and were sequenced, as did seven of the eight *Pholcus* samples. The lightest (1.1ng) *Pholcus* sample consistently failed to produce amplicons.

6.3.4 Adult *Gnorimus* specimen extraction and sequencing

All sequences were checked for identity using the BLAST algorithm against the GenBank database. Sequences were determined to have been successful if they were clearly readable and aligned with the other material from the gene in question. Extraction + amplification + sequencing success rates for the different age classes are shown in Fig. 6.3.

For *Gnorimus* patterns were variable across all five gene regions tested, with relatively high success in certain genes and samples less than 10 years old (Fig. 6.3; COI = 84.5%, and the microsatellite locus = 97.9%) and lower in other genes with 10 year old material (Control Region = 65.3%, *Orco* = 47.3%, HSP70 = 63.0%). With older material (from 10 years and above, up to 58 years old), the same genes produced relatively similar, but sometimes considerably lower, success rates with successful sequences (COI = 87.9%, Control Region = 18.8%, *Orco* = 39.4%, HSP70 = 24.2%, and the microsatellite locus = 95.7%). Success rates were generally highest for COI and the microsatellite, and lowest for CR, but both CR and HSP70 showed a strong decline in success rates between the 1-10 year and 10 year plus samples. Neither COI nor the microsatellite showed significant decreases in amplification success in the older samples.

For *Trichius* adults success rates were uniformly high for the two gene regions tested and for samples of all ages (Fig.6.3.f and Fig.6.3.g).

6.4 Discussion

Extracting DNA from non-invasive sources has the potential to allow for conservation genetic studies to be completed on a study species, but without needing to find and capture the species in question in the wild. Though used for many years on megafauna, such work on terrestrial arthropods has only been completed in the last 12 years, and is still limited to a handful of studies (Table 6.1). However, these studies have worked on a diverse range of insects (Coleoptera, Diptera, Hymenoptera, Lepidoptera, Odonata) as well as other arthropods (Araneae, Diplopoda), suggesting that the techniques can be widely applied across terrestrial arthropods.

DNA extraction from faeces/frass is the only non-invasive sample that has been studied across a wide range of terrestrial arthropods, with exuviae and pupal cases also being commonly utilised. However, only one previous study has used frass collected from wild populations of terrestrial arthropods (Strangi et al. 2013), showing a clear decline in the quantity of PCR product amplified from older samples (up to 3 years old, from a 345bp DNA region), but that the species-specific primers still amplified this old product. Table 6.2 shows that both a Phenol-Chloroform extraction and the commercially available kit DNeasy can produce DNA from which it is possible to amplify COI reliably from *Gnorimus nobilis* frass using species-specific primers, with extremely high success rates with well diluted DNA. As the age of these frass samples is unknown, it is potentially a representative wild sample. Sample weights were also considerably under those used by Strangi et al. (2013) at 2.2mg against ≈100mg. A slight decline in the reliability of the larger primer pair was also detected (Table 6.2), suggesting that the frass was old enough for the quality of DNA present in each sample to have declined (Piggot 2005; Deagle et al. 2006), consistent with other studies on degraded DNA (Junqueira et al. 2002; Hajibabaei et al. 2006).

In addition, species-specific microsatellites were successfully amplified from 91.7% of the frass samples. Though most individuals were homozygous for a common allele (n=40, allele size=120), this pattern is consistent with allele 120 being by far the most common across Europe and in the UK, and so does not indicate the possibility of large allele drop out resulting from preferential amplification of small target regions in poor quality DNA (Banks et al. 1999; Butler et al. 2003). A larva from The Cherrys site showed the same genotype as all four frass samples from the same tree, but one individual from the Tiddesley Wood orchard site (frass from two trees was used in the extraction) was a heterozygote (120/122).

This is unusual as the two adults genotyped from the same orchard are both homozygous for allele 110. This could be explained by large levels of subpopulation structuring in the orchards, with individual trees holding very distinct inbred breeding groups of *Gnorimus*, but as the source tree from which the adults came is unknown, it is possible that they came from a different tree (i.e. family) to the frass sample. These successful and reliable amplifications from frass indicate that it represents a potentially useful sampling target for future studies into beetle landscape genetics.

Although both mitochondrial (COI) and nuclear (microsatellite) regions were equally successfully amplified from DNA extracted from frass in this study, multiple copies of mitochondrial genomes are present in every animal cell, thus there is a greater chance of undegraded target mtDNA than nDNA being present. Deagle et al. (2006) suggested a model for DNA damage from strand breaks or chemical modification where:

$$f(x)=\lambda e^{-\lambda x}$$

In which x is the distribution of undamaged fragment sizes, λ represents the probability of a nucleotide being damaged, and $e^{-\lambda x}$ is the complement of the cumulative exponential distribution of an amplicon of xbp in size. As product size increases, there is an exponential decline in amplifiable copies, the rate of decline being mediated by λ, which is in turn affected by the DNA environment. Each mitochondrion possesses multiple copies of their own mitochondrial genome, and each arthropod cell typically has hundreds of mitochondria present, thus mitochondrial genomes outnumber genomic DNA by thousands of times per cell (Minelli et al. 2013). Therefore, future studies on non-invasive wild samples should focus on reliably amplifying short sections of mtDNA, preferably from short sections which show from previous studies some within-population variation, rather than investigating the longest amplicons possible.

In addition to the successful amplification and sequencing of COI from beetle frass, DNA from spiders was also successfully recovered from web samples from two species in different settings, one from a captive Theraphosid, and one from a wild Pholcid. The present study adds to a previous study (Xu et al. 2015), by demonstrating successful DNA extraction using a different method, amplification and sequencing of a longer COI region, and application of the technique to a wild spider. The technique should be useful to metabarcode spiders from ecosystems from web samples, as well as potentially being useful in population genetic studies. As small microsatellite loci can be amplified in beetle frass, it seems likely that

nuclear targets of a similar, or larger, size can also be amplified in spider webbing. As prey can be molecularly identified in webbing (Xu et al. 2015), this should lead the way toward being able to apply metabarcoding techniques to spider webbing involving high resolution analysis of prey species targeted by a particular individual or the population (Taberlet et al. 2012; De Barba et al. 2014; Krehenwinkel et al. 2016).

The results from the amplification of DNA extractions from adult beetle samples from "non-optimal" sources (e.g. old dried specimens in museums) show that the mtDNA COI region and the small nuclear microsatellite locus display little change in amplification success with specimen age (Fig. 6.3 a), and success rates for these genes may be more to do with the preservation method of the sample rather than age. This is consistent with some other work on COI from old specimens (Junqueira et al. 2002 on a mixture of mtDNA markers on Diptera up to 70 years old), but is in contrast to Miller et al. (2013) who reported a strong decline with age in successful sequencing of COI in spiders preserved in ethanol: samples 15 years old only successfully amplified and sequenced approximately 50% of the time, whilst 45 year old specimens only succeeded 20% of the time. Though the sample size reported here is smaller, COI was sequenced successfully for all three 58 year old samples of *Gnorimus*, and was similarly successful in *Trichius* up to 44 years old. *Trichius* have somewhat smaller legs than *Gnorimus*, presumably with less DNA present per leg sample, so the finding of roughly similar COI sequencing success rate between the two genera is a reliable indicator that CTAB-Phenol-Chloroform is a useful DNA extraction method, regardless of specimen size, though it may also be due to the shorter COI length targeted for *Trichius* (461bp vs 516bp).

Control Region, a long A+T rich region with large tandem repeats in *Gnorimus* (Chapter 5), amplified considerably less reliably than COI. *G. variabilis* was not tested in this experiment: tests indicated that the gene only amplified infrequently in this species, and may be due to mutations in the primer binding sites rather than issues with specimen quality, therefore these samples were removed from the analysis. No adult specimen over 15 years old was successfully sequenced for this gene, which may confirm that A+T rich templates degrade faster than those with more 'normal' A+T ratios (Kool 2001). The shorter length of the CR region targeted (330bp) compared to COI and the lower success rates are testament to its difficulty as a DNA template.

Nuclear genes have been little investigated in studies looking at museum specimens. Raupach et al. (2010) noted that nuclear genes could be sequenced in their dataset from specimens up to 12 years old, but provided no data on sequencing attempted on older specimens, and its success rate. None of the nuclear gene targets here were over 206bp, but the four genes (HSP70, *Orco*, *Wingless* and the microsatellite) showed slightly different patterns of sequencing success: HSP70 worked very well on younger specimens (under 10 years old), and was sequenced in one of the oldest 58 year old samples, but largely performed poorly with samples over 10 years old. *Orco* was less reliable overall, but was sequenced in more specimens over 10 years old. In contrast to these markers, which performed far more poorly than COI, *Wingless* performed as well as COI in *Trichius*, amplifying in every sample above 15 years old, up to the maximum 44 years old. The 120bp microsatellite performed extremely strongly regardless of the sample age, with higher success rates than the mitochondrial COI. This is likely due to its very short size (approximately a quarter of the length of the COI section targeted).

Watts et al. (2007) used a suite of microsatellite markers (from 125-283bp in length) on damselfly specimens from museums, reporting a steep drop off from acceptable success rates (60%+ from samples 30-50 years old) to 0% in samples more than 50 years old. Thomsen et al. (2009) also reported mixed success using the mitochondrial COI and 16s genes, with primers targeting sections 77-204bp in length, but did manage to sequence genes from material 190 years old. Strange et al. (2009) also reported declines in amplification success on museum bumblebee samples for microsatellites, but still achieved 50% + success rates with samples up to 50 years old. Nakahama & Isagi (2016) reported a strong effect of microsatellite product size and sample age on the successful amplification of these loci from museum butterflies, with shorter products (under 120bp) clearly performing more reliably than longer products (140bp+) on samples older than 30 years. Finally, Krehenwinkel & Pekár (2015) reviewed a number of effects on sample sequencing success by standardising the study species (the wasp spider *Argiope bruennichi*), and DNA extraction method, and finding increases in successful sequencing to correlate with sample body size, mitochondrial rather than nuclear markers and shorter marker length, plus additional factors that are interesting, but difficult to compare across the other studies, including strong effects of the museum collection itself, and the distance between the sample locality and the museum it was housed in. Though none of these studies are directly comparable due to the different species studied with different DNA extraction methods, they all point to arthropod DNA being

reliably amplifiable from museum tissue from ages up to and above 50 years provided that the correct genetic markers are chosen.

Nuclear genes have not received the research attention that they should have for arthropods, despite a number of studies suggesting that they can be useful markers. Whilst nuclear genes are certainly harder to amplify and sequence than 'typical' mitochondrial barcoding genes, small products can be sequenced with some reliability in older specimens, though this depends on the gene. Why there is such a stark difference between the success of *Wingless* and *Orco* (149bp and 150bp respectively) is unknown, but the success of the microsatellite locus in amplifying so reliably is likely due to its small size. As both beetle frass and spider webbing have proven to be useful samples for COI amplification, and microsatellites for beetle frass, the next step for population and conservation genetics is to utilise such samples for full multi-microsatellite marker studies to see if such non-invasive methods are useful for investigating population demographics alongside barcoding and mtDNA analysis.

.

Table 6.1. The tissue sources and taxa that have been investigated for non-lethal and non-invasive sampling in terrestrial arthropods. An '*' denotes a study completed on wild-sampled material, as opposed to under laboratory conditions or where an animal has been taken into captivity (even only briefly).

Source	Taxa	References
Frass and faeces Non-invasive	Araneae, Coleoptera, Diplopoda, Hymenoptera, Lepidoptera, Odonata	Feinstein 2004, Monroe et al. 2010, Seeber et al 2010, Lefort et al. 2012, Scriven et al. 2013, Strangi et al. 2013*, Sint et al. 2015
Exuviae and pupal cases Non-invasive	Coleoptera, Diptera, Lepidoptera, Odonata	Feinstein 2004, Watts et al. 2005*, Dhananjeyan et al. 2010, Lefort et al. 2012, Richter et al. 2012*
Wing clippings Non-lethal	Lepidoptera, Odonata	Vila et al. 2009*, Monroe et al. 2010*
Leg clipping Non-lethal	Lepidoptera, Odonata	Vila et al. 2009, Monroe et al. 2010
Defensive secretions Non-invasive	Coleoptera	Donald et al. 2012
Spider webs Non-invasive	Araneae	Xu et al. 2015

Table 6.2**:** Details of novel primers used in the study on spider webbing. The number in the Psal primer names denotes the position from the 5' end of a *Psalmopoeus cambridgei* cytochrome oxidase 1 sequence from GenBank (accession number JQ412455.1), whilst in the Phol primers the number in the primer name denotes the position from the 5' end of a complete mitochondrial genome of *Pholcus phalangioides* from GenBank (accession number JQ407804.1). 'F' and 'R' at the end of the primer name refer to whether the primer is a forward or reverse respectively.

Taxon and targeted gene	Primer name	Primer sequence
Psalmopoeus cambridgei CO1	Psal-333F	5'-GGGGCCGGGTGAACTATTA-3'
Psalmopoeus cambridgei CO1	Psal-530R	5'-TACAGACCACAAACGCG-3'
Pholcus spp. CO1	Phol-415F	5'-GGGGTTTCTATGGATTTTGC-3'
Pholcus spp. CO1	Phol-459F	5'-GGCTTCTTCTATTATAGGGGC-3'
Pholcus spp. CO1	Phol-633R	5'-GTCAGTCAACAATATGGTAATAGC-3'
Pholcus spp. CO1	Phol-694R	5'-CAGCCGTAATTAAAACAGACC-3'

Table 6.3: PCR amplification success (number of reactions out of 24 tests displaying DNA products) using 3 different primer sets on frass-derived DNA recovered using different extraction techniques and stock dilutions.

	DNA region tested	
Extraction source and dilution	1st CO1 (219bp)	2nd CO1 (333bp)
Chelex Stock	0	0
Chelex 1/10	0	0
Chelex 1/100	1	0
Chelex 1/200	1	1
DNeasy Stock	6	0
DNeasy 1/10	24	13
DNeasy 1/100	23	23
DNeasy 1/200	24	19
Phenol-Chloroform Stock	0	0
Phenol-Chloroform 1/10	9	0
Phenol-Chloroform 1/100	22	13
Phenol-Chloroform 1/200	24	23

Figure 6.1. *Gnorimus nobilis* frass in a typical position in an apple tree. Here the species has been present in the hole for long enough for frass to reach a point where it spills out of the rot-hole

Figure 6.2: representative PCR success of *Pholcus phalangioides* webbing with HyperLadder II (Bioline) on a 3.5% agarose gel in TBE. Sample 1: Phol-415F:Phol633R; sample 2 Phol-459F:Phol633R; samples 3 & 4- Folmer primers on two different samples of *P. phalangioides* webbing. The 300bp and 700bp ladder markers are labelled

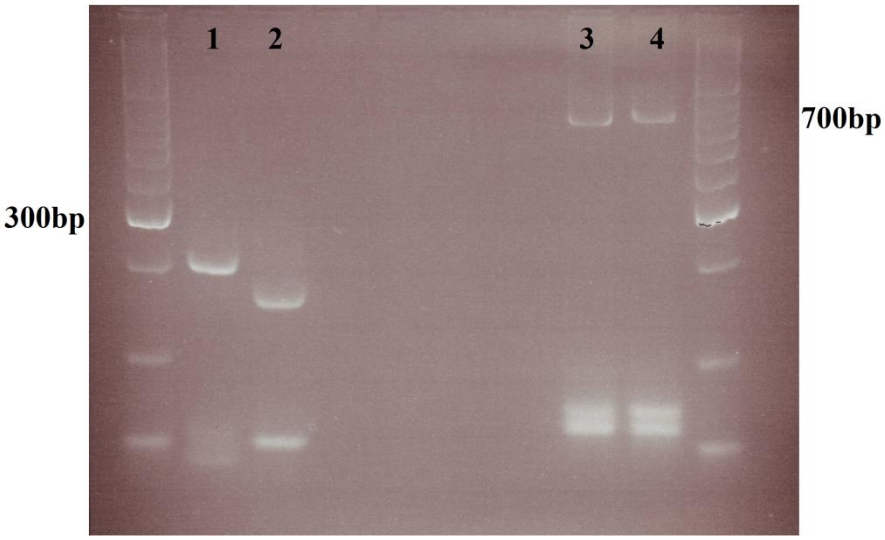
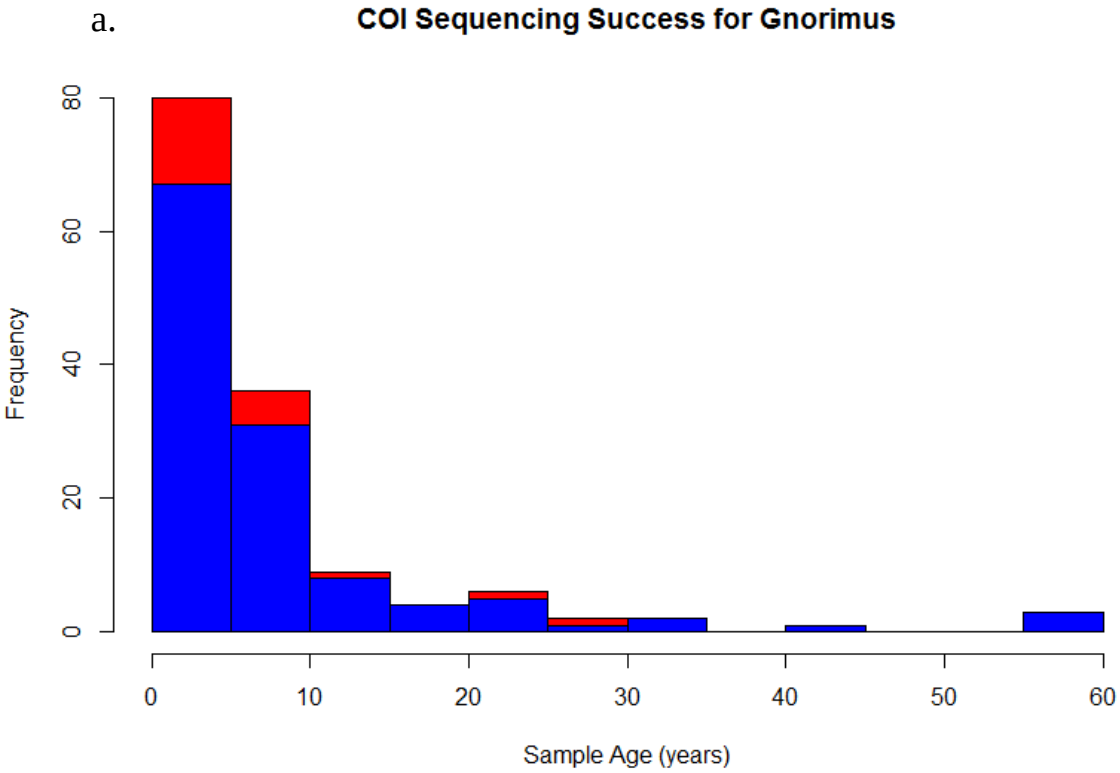
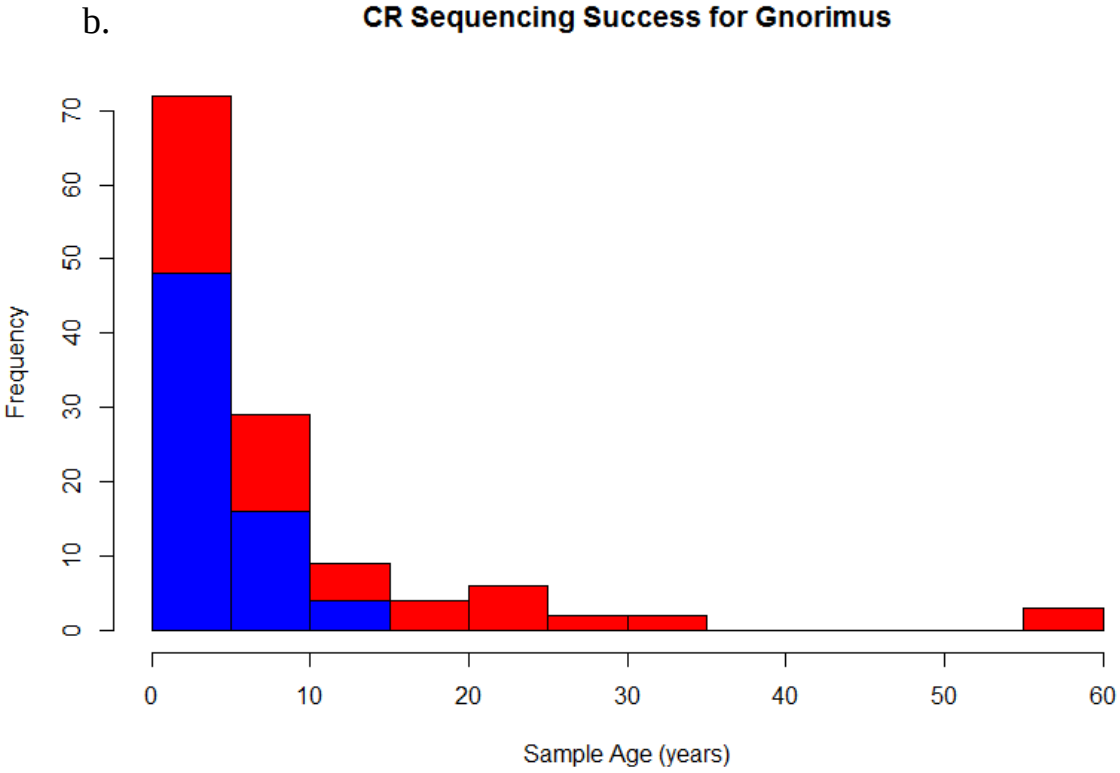


Figure 6.3. Histograms showing the success rate (blue) of production of clean DNA sequences from attempts using *Gnorimus* or *Trichius* adult beetle samples of different ages (red) amplified for a range of DNA targets. Not all samples were used with every marker, hence the different sample sizes

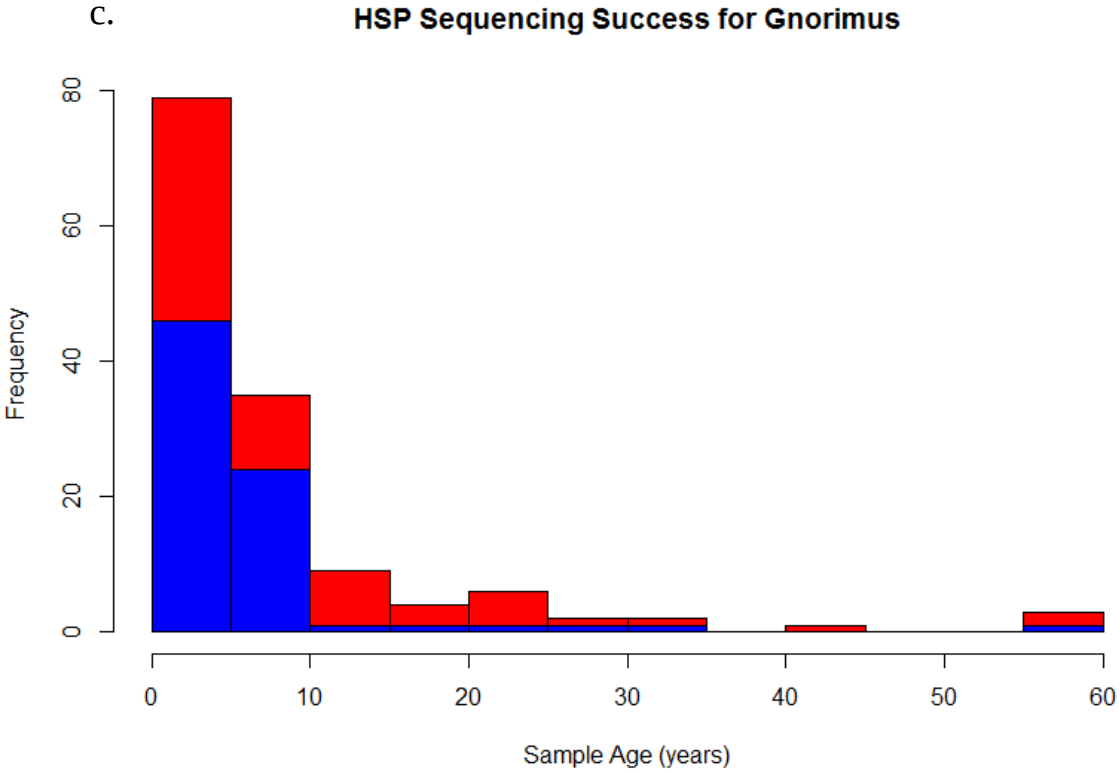
a.


COI Sequencing Success for Gnorimus
Frequency
0
20
40
60
80
0
10
20
30
40
50
60
Sample Age (years)

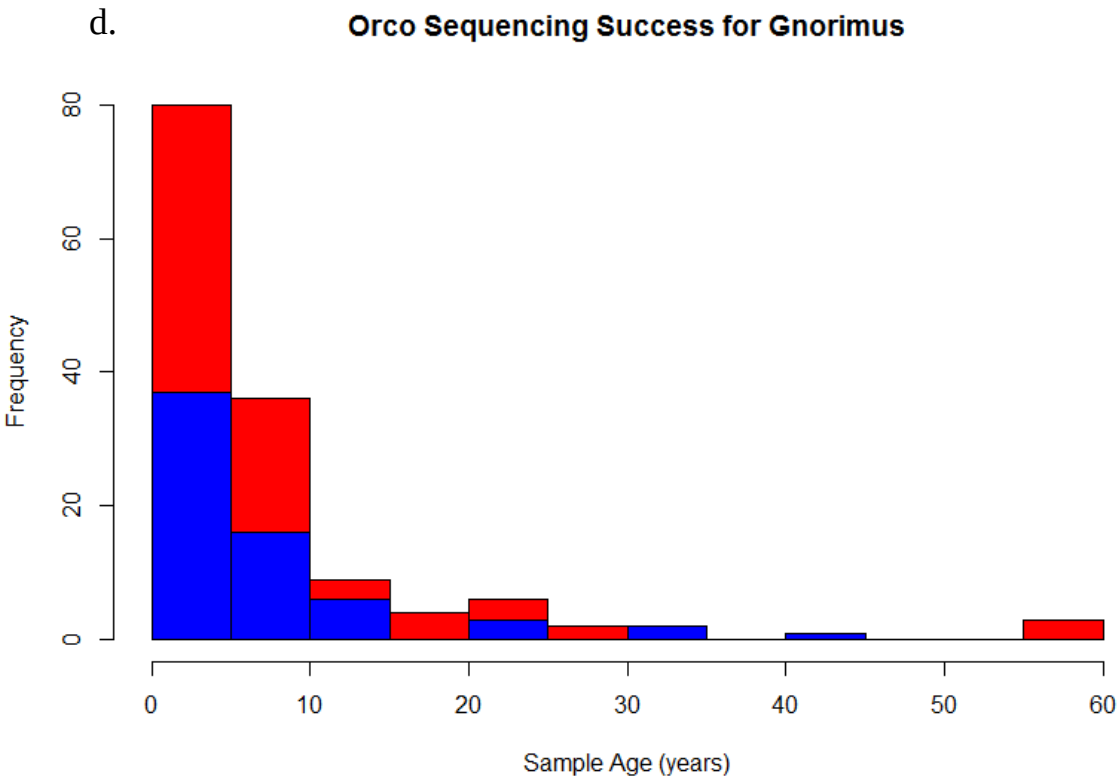
b.


CR Sequencing Success for Gnorimus
Frequency
0
10
20
30
40
50
60
70
0
10
20
30
40
50
60
Sample Age (years)

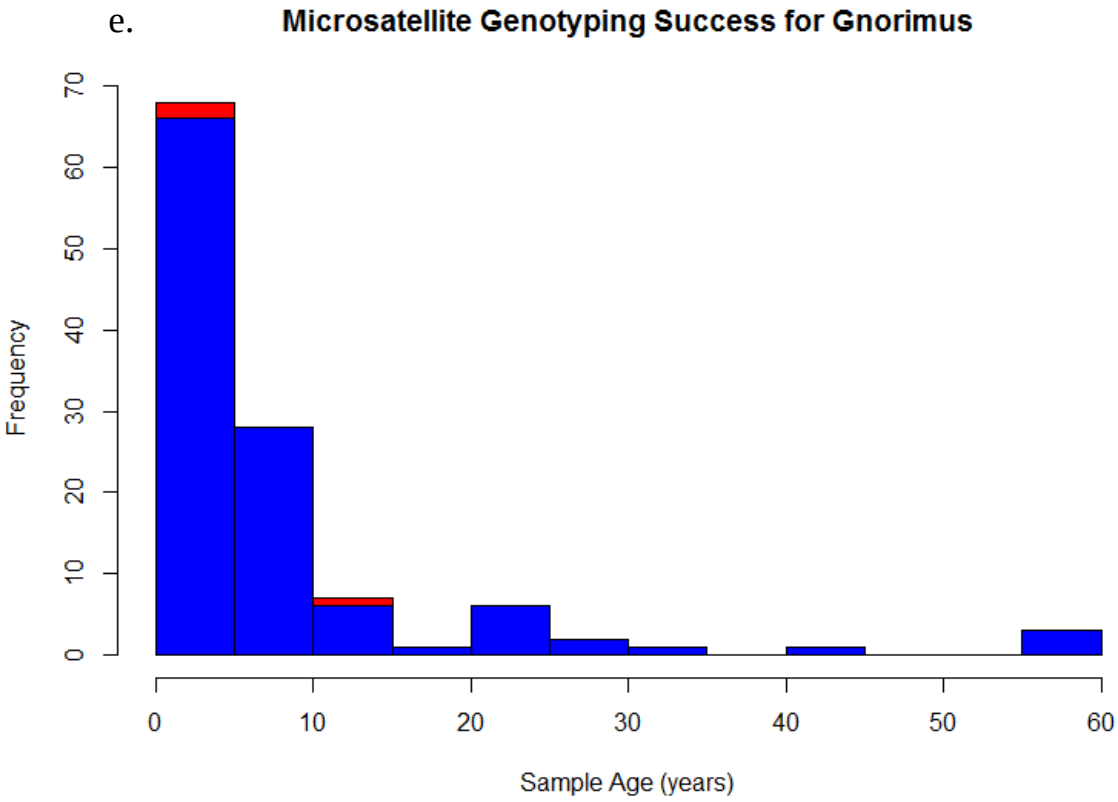
c.


HSP Sequencing Success for Gnorimus
Frequency
0
20
40
60
80
0
10
20
30
40
50
60
Sample Age (years)

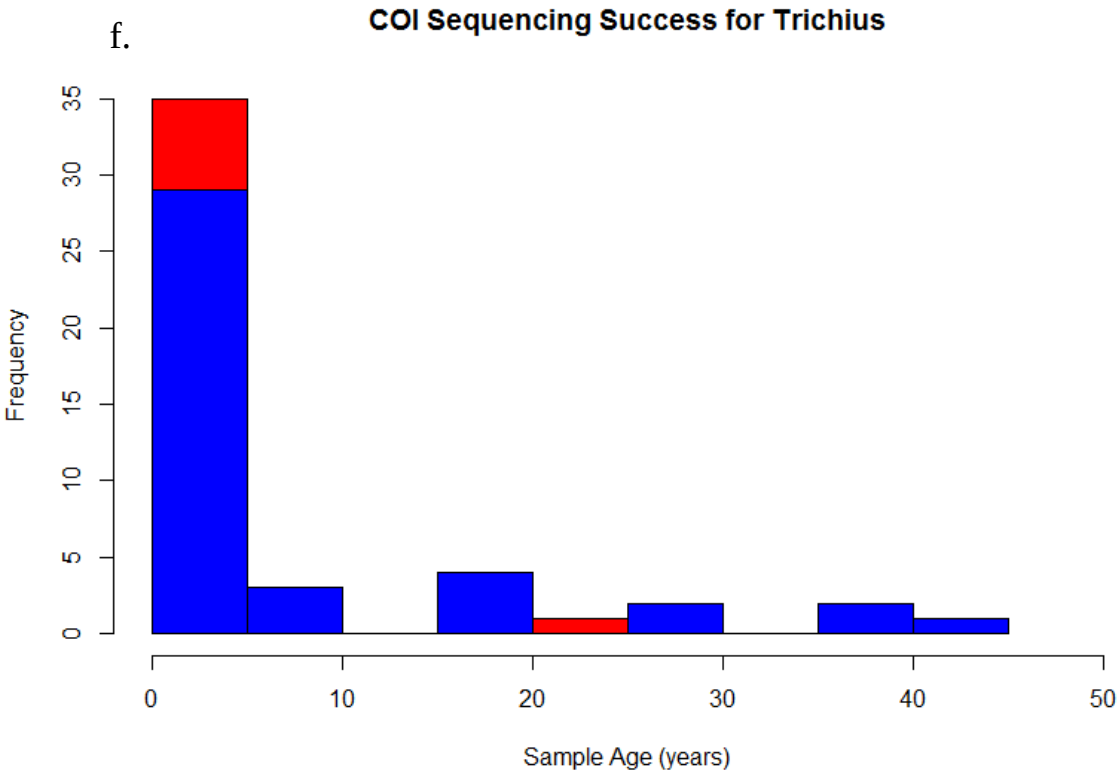
d.


Orco Sequencing Success for Gnorimus
Frequency
0
20
40
60
80
0
10
20
30
40
50
60
Sample Age (years)

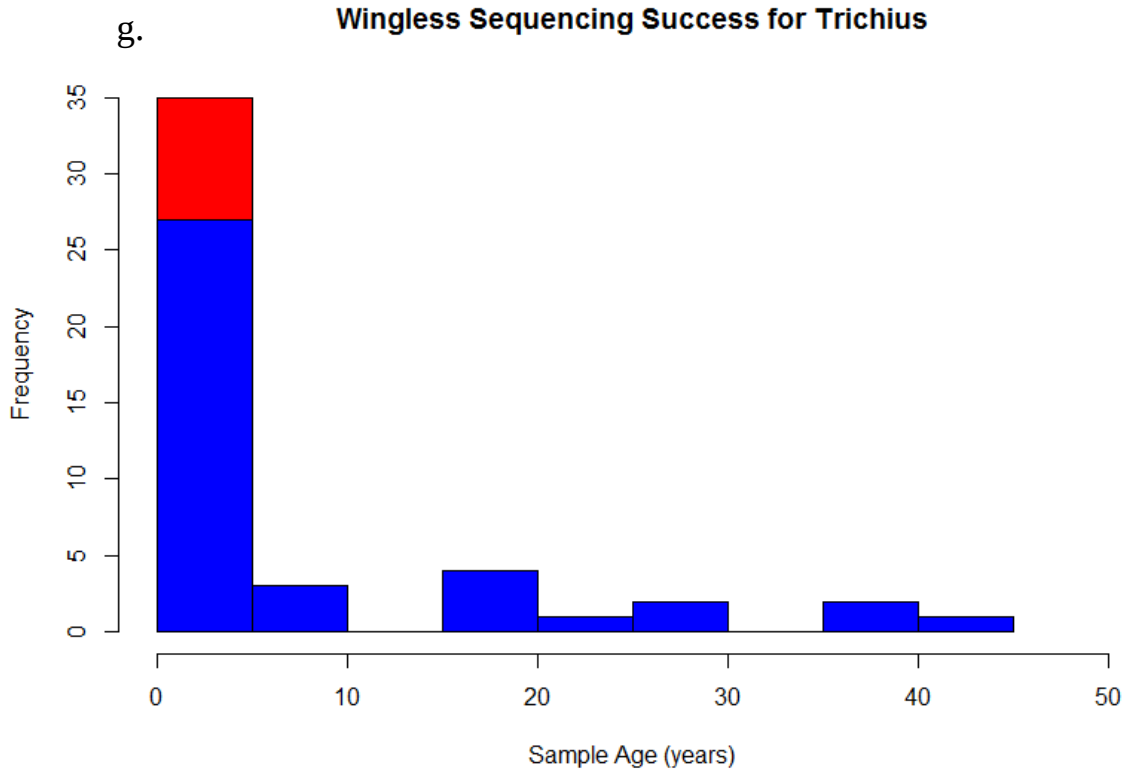
e.


Microsatellite Genotyping Success for Gnorimus
Frequency
0
10
20
30
40
50
60
70
0
10
20
30
40
50
60
Sample Age (years)

f.


COI Sequencing Success for Trichius
Frequency
0
5
10
15
20
25
30
35
0
10
20
30
40
50
Sample Age (years)

g.


Wingless Sequencing Success for Trichius
Frequency
0
5
10
15
20
25
30
35
0
10
20
30
40
50
Sample Age (years)

Chapter 7

Species Distribution Modelling of Saproxylic Scarabs

7.1 Introduction

Despite being large, charismatic insects found worldwide, saproxylic scarabs (Scarabaeoidea) have received surprisingly little research interest, even within Europe where the conservation status of many species has been assessed (Nieto & Alexander 2010; Horák et al. 2012). Though not early stage ecosystem engineers (like the European *Cerambyx cerdo*, Cerambycidae (Buse et al. 2008)), saproxylic scarabs perform decomposition roles that no other species can by using their large body size, powerful mandibles and manipulation of fungi to engineer ecosystems in decomposing wood (Tanahashi et al. 2010; Mouillot et al. 2013; Sánchez-Galván et al. 2014; Micó et al. 2015; Landvik et al. 2016; Tanahashi & Hawes 2016).

As inhabitants of forest ecosystems, saproxylic habitats currently exist in a fragmented state across much of Europe, with very little undisturbed ancient habitat remaining (Rukke 2000; Linder et al. 2010). Loss of habitat and degradation is the main threat facing saproxylic beetles (Alexander 2003 b.), which are being additionally stressed by climate change. However, the extent to which climate change will affect saproxylic scarab distributions is unknown. One approach to assessing the projected impact of climate change is to use Species Distribution Modelling (SDM) approaches by predicting the distribution of a species in geographic space, then projecting the species-specific factor responses onto a geographic model for the time period in question (Elith et al. 2011) to aid conservation planning (Loiselle et al. 2003; Pearson & Dawson 2003; Franklin 2010).

As well as being able to make predictions for the future, SDM approaches can also help make inferences about species distributions in the past, in particular predicting the locations of glacial refugia (Provan & Bennett 2008; Svenning et al. 2008). This latter application of SDMs becomes particularly powerful when combined with phylogeographic approaches, allowing cross-discipline hypothesis testing (Knowles et al. 2007; Wielstra et al. 2013). In one of the most commonly applied and accessible approaches to developing SDMs, climatic data (such as the Bioclim bioclimatic variables dataset; Hijmans et al. 2005) are used alongside species presence data as inputs into a presence-only algorithm (such as Maxent, a machine-learning method which minimises relative entropy between the probability density of the species presence points and the wider study landscape; Phillips et al. 2004; Philips et al. 2006; Elith et al. 2011) to produce a robust model of species distribution with a statistical basis.

Despite being widely utilised, SDMs have been criticised for being applied to weak datasets (Jiménez-Valverde et al. 2008) or being applied where recent historical processes have affected species distributions (Hortal et al. 2012), leading to coincidental congruence between distributions caused entirely by historical processes (such as human-mediated habitat loss and degradation) and those arising from genuine contemporary climatic factors. Though such concerns are certainly valid, and must be considered in any SDM workflow, that historical processes (such as post-glacial recolonisation of terrestrial environments from glacial refugia, Hewitt 1999) have affected distributions of European terrestrial species is without doubt. Additionally, modelling conducted at wide study scales using range-restricted species can often yield artificially inflated Area Under the Curve (AUC) values (used to measure the predictive performance of SDMs; Jiménez-Valverde et al. 2008), increasing as the study area is increased relative to the true geographic range of the organism (AUC being relative to the occurrence area in question). This can lead to over simplifications of model performance and a focus on achieving high AUC values for publication (Lobo et al. 2007), as this statistic is often described as an indicator of model performance and for cross-dataset analyses (Fielding & Bell 1997; Hortal et al. 2012).

Additional difficulties can arise from situations where the species used in the modelling is a habitat specialist, but the habitat itself is fragmented or locally patchy. In the case of saproxylic beetles feeding on decayed trees for at least part of their life cycle (Alexander 2008a), their chosen habitat is usually specific, such as heartwood rots in certain tree species caused by certain fungi. Models predicting changes in areas of suitable climate for species under climatic change scenarios will lead to an over simplification of the species' niche. Regardless of the suitability of the climate, if the habitat is not suitable, the species will be unable to colonise said areas.

Approaches using null climate or presence data to test the power of SDMs have been shown to be useful in assessing a variety of taxa generally considered to be 'difficult' to model (Beale et al. 2008; Chapman 2010; Hijmans 2012; Algar et al. 2013). Typically a Principle Components Analysis is performed on the input climatic data to reduce the number of predictor variables used in the SDMs (reducing computing time) and to eliminate multicollinearity between individual predictors, whilst retaining the majority of the variance contained within the climatic data (Baldwin 2009; Williams et al. 2015). Then, SDMs can be assessed against two different null model predictions: via null climate data which replicates the spatial pattern of each observed climate variable and the relationships between each

variable, and null presence data which replicates the spatial pattern in the observed species distribution. Where a significant association between predictor variables and species presences is found (when the median AUC value of the observed SDM is significantly higher ($p \leq 0.05$) than those of the null models), models can then be projected via hind- and fore-casting to produce the best prediction possible based on the input climatic variables and species presence points (Williams et al. 2015).

Using multiple species under a single SDM framework allows for robust hypotheses to be drawn up from cross-species comparisons (Svenning et al. 2008). Here, a range of species from Scarabaeoidea have been incorporated into the modelling framework, including species specialising in certain habitats (climatic specialists, such as those associated with Mediterranean or boreal climates) or food sources (food specialists in rot-holes versus generalists feeding on a range of rotten wood, plus those species flexible enough to feed on other decayed biological material).

It is predicted that there will be little difference in the glacial refugia utilised by food-source specialists and generalists, but much stronger responses to hind and forecasting by those climatic specialists. Whereas forecasting climate change may prove to result in much of the Mediterranean region becoming unsuitable for some Northern European generalists, it may lead to expansions of Mediterranean specialists as suitable climates are found further north (Williams et al. 2015), replacing the generalist species in environments they are currently found in. Hindcasting the models to compare the glacial refugia of this mixture of species will also allow for refugial hotspots to be identified and compared to insights from population genetics and phylogeography (Forester et al. 2013). These multi-disciplinary approaches help add weight to the hind-casts, allowing the model approach to be better evaluated.

7.2. Methods

7.2.1 Species records

Saproxylic scarab beetles (Scarabaeoidea, Coleoptera) were chosen as the study taxa due to the long standing interest in their distributions, their conservation value (Nieto & Alexander 2010), and their status as ecosystem engineers (Micó et al. 2015; Tanahashi & Hawes 2016). A wide mixture of Scarabaeoidea species were initially screened for the study, representing a

range of different feeding preferences, conservation listings and specialisations. Species distribution data were collected from the Global Biodiversity Information Facility (GBIF) website (www.gbif.org) using the RGBIF package (Chamberlain et al. 2016) in R (R Development Core Team 2008). Data were defined as being suitable for the study if they were geotagged and from 1950-2016 to match the models for current climate as closely as possible. Due to biases in data collection from such public resources (Maldonado et al. 2015) this was supplemented by literature searches for the chosen species. As well as specialist obligate saproxylics, facultative saproxylics (which also feed on a variety of other plant-derived substances such as compost and leaf hummus) were also investigated for comparison. Species selected to be taken forward had to have geotagged locality data from across their approximate range. Conservation listings were taken from Nieto & Alexander (2010) and the IUCN Red List (2016). Biological data came from the BugsCEP program (Buckland & Buckland 2006; Buckland 2007) and from the wider literature (Supplementary Material).

7.2.2 Model Generation

Bioclimatic data were collected for recent observed conditions (1950-2000) using the general circulation model MIROC-ESM (Watanabe et al. 2011) downloaded via the WorldClim website (Hijmans et al. 2005). MIROC-ESM is an earth system model which couples observed climatic data from weather stations with biogeographical components, which can then be projected to simulate past and future climate. Projections using the MIROC-ESM model were collected for: the Mid-Holocene (6KYA), the Last Glacial Maximum (22KYA), the year 2050 under Representative Concentration Pathway (RCP) 4.5, and the year 2070 under RCP4.5. RCP4.5 represents a ‘stabilisation without overshoot’ model of global radiative forcing (Moss et al. 2010) and is considered to be a moderate model of future environmental change. All datasets were used at a resolution of 10 arc minutes to reduce computing time.

Climate variables from the observed data were subjected to a Principle Components Analysis (PCA) within the study area (-12.0 to 47.0 degrees East/West, and 34.0 to 72.0 degrees South/North), representing the Western Palearctic. Scores for the first three PCA values were used, whilst PC scores for the future and past climate scenarios were derived by standardising future data using means and standard deviations from observed modern climate data, before

subsequently applying the eigenvectors from the PCA of the observed modern data (Williams et al. 2015).

For each species, a null distribution was generated by randomly placing the number of points (species observations) on the area used in the projections (above), then iteratively moving a single presence point at a time, and retaining the change if it reduced the deviation from the spatial pattern in said species' observed presence data (Williams et al. 2015). The spatial pattern for each species was assessed via an empirical semi-variograph, a density histograph and the area of the convex hull encompassing all presence points (Beale et al. 2008; Williams et al 2015). Deviation from the observed data was calculated for a total of 50000 iterations (if it did not terminate before reaching this number) for each of the three measures as the sum of squares of the difference between the observed and generated measure, the product then being taken to give the overall deviation of a distribution generated from the pattern in the observed data.

For each species 200 null distributions were generated, from which 99 were selected for each species by rejecting any null presence distribution which had a correlation coefficient above $1x10^{-4}$ with the observed presence distribution, by rejecting any null distribution for which the sum of squares deviation for the semi-variogram was greater than $1x10^{-5}$, and finally by selecting those 99 null distributions that minimised the total cross-correlation between the resulting set (Williams et al. 2015). This set of 99 null distributions capture as many qualitatively different distributions as possible, whilst also adequately conforming to the spatial pattern of the observed data.

Maxent was then used to compute the SDM for each species with the input layers from above. For Maxent, outputs reported include the area under the Receiver Operating Characteristic Curve (the Area Under the Curve, AUC), widely regarded as the most important signal of a robust model (Fielding & Bell 1997) where values of 1 demonstrate perfect predictive performance, and 0.5 represents random chance. Values above 0.75 are typically reported as representing models with useable predictive power.

Combined with the 99 null presence models, the observed species distribution was included to produce a population of 100 models. A Kruskal-Wallis test was then used to detect whether the median AUC of any of these 100 models was significantly different to any other. If a significant difference was detected, pairwise t-tests (with Bonferonni corrections computed in MATLAB (MathWorks 2012)) were used to compute the number of null models

that had significantly lower median AUCs than the observed model. The association between climate and distribution as identified by the SDM is significant under this test if the AUC of the observed model is significantly greater than that of 95 of the null models, giving a one-tailed significance cut off of P=0.05 (Williams et al. 2015).

The internal fitting parameters that are selected by default in Maxent are designed to capture the different responses of species to their environment. The five different types are: Linear, where continuous variables should be close to observed values, Quadratic, where the variance of continuous variables should be close to observed values, Product, where the covariance of two continuous variables should be close to observed values, Threshold, where the proportion of the model that has values above a threshold for a continuous variable sould be close to the observed population, and Hinge, whereby a linear feature is truncated at a threshold (Elith et al. 2011). The program internally decides which features are available to the model depending on the number of species presence points: the more presence points in the dataset, the more features are available to the model and the more complex and fitted the final SDM can become (Philips & Dudik 2008, Williams et al. 2015). However, this approach can overfit models, especially when the number of species presence records is low (Raes & ter Steege 2007; Warren & Seifert 2011). Therefore, every possible nested Maxent model was tested, and the most parsimonious was chosen for comparison to the null models. The final nested model chosen for each species was the one which returned the lowest Akaike Information Criterion (AICc) value the most times over 1000 split-data (60% training, 40% testing) replicates (Warren at al. 2010).

7.3. Results & Discussion

7.3.1 Selection of species

From the combined search methods 24,840 records of saproxylic scarabs were collected from Europe. From the initial pool of 25 species, 13 were selected as being suitable for the study (Table 7.1), representing a diverse range of saproxylic ecologies, plus species flexible enough to feed on decayed wood and other decayed biological material. Sample sizes varied from 39 for each of *Propomacrus bimucronatus sensu lato* and *Protaetia mirifica* to 6380 for *Lucanus cervus* (Table 7.2).

7.3.2 Key Environmental Factors Shaping Species Distributions

Together the first three principle components explained 82.75% of the variation contained within the 19 Bioclimatic variables (Table 7.3). The individual factors that each PC correlates with are outlined in Table 7.4. Broadly, each variable explains:

PC1: seasonality and extreme seasonal variations. High scores are indicative of regions with large seasonal temperature and precipitation fluctuations in continental regions, with much of western Europe being currently very moderate in this regard likely due to influence of the Gulf Stream (Fig. 7.1a). This extreme seasonality is predicted to become more extreme under the climate change scenario, with Spain, Italy and Eastern Europe becoming associated with high PC1 values (Fig. 7.1b, c). In contrast, seasonality is predicted to have been limited under both the 22kya and 6kya projections, only showing high PC1 values in the north east, and extremely low values on west facing shores. Much of Europe appears to have been very moderate in this regard, with lower PC1 values across much of eastern Europe in particular than today (Fig. 7.1 d and e), implying a milder climate.

PC2: high temperatures, with high summer temperatures and low precipitation. High scores are indicative of hot dry regions in the southern Mediterranean, whilst low scores are indicative of high southern mountains (the Alps, Pyrenees, and Carpathians) and mountainous areas of Scotland and Norway. A trend toward moderation can be seen in the forward projections (Fig.7.2 b and c) with much of Spain, Turkey and the Balkans showing lower PC2 values, though there is predicted to be very little change to areas with very low PC2 values. Much of northern Europe is predicted to have had very low PC2 values during the last glaciation (as expected from current knowledge of glacial extent) with a general push south in high PC2 values (Fig. 7.2 d). 6kya (Fig. 7.2 e) is intermediate between the two, with both northern and southern areas demonstrating more moderate PC2 values.

PC3: precipitation seasonality, particularly in regions with extreme wet/cold quarters. High scores in the modern dataset are found in Atlantic Norway and Portugal, Western Greece and southern Turkey (Fig. 7.3 a). These currently high PC3 southern areas are predicted to become less extreme in the 2050 and 2070 analyses (Fig. 7.3 b & c), whereas Atlantic Norway is predicted to become more extreme, along with Atlantic Scotland. Low scores for this PC are found in northern France, Germany and England, and are predicted to spread eastwards through Europe as climate change advances. There are very few differences in PC3

throughout Europe between the current values and the values for 6KYA (Fig. 7.3 d), whereas for 22KYA areas are generally more extreme (Fig. 7.3 e).

7.3.3 Scarab SDMs

In general, most species demonstrated broad climatic niches, with much of western and central Europe predicted as being suitable. Generally the climate becomes less suitable for these species toward the extreme north (the Scottish highlands and mountainous Scandinavia) and the east (Russia in particular). Such species also responded strongly to PCAs 1 and 3. For many of the widely distributed species the 6kya projections show large areas of highly suitable climate across central Europe, in many cases there are more areas with high species distribution probabilities (represented by yellow to red in the Figures) in the 6kya projections than in the current climate scenario due to this period being milder than current (Andersen et al. 2006). Most species are also predicted to have glacial refugia in Iberia and Italy (as expected from standard glacial refugia theory (Hewitt 2000)), but many species also show predicted distributions indicating northern refugia (Provan & Bennett 2008), though naturally, especially for the obligate saproxylics, these predictions hinge on the presence of host trees. Projecting these niches onto future climate change scenarios suggests that these widely distributed species should move further north east as the climate shifts, and that there may be large decreases in suitability in the south of the range, particularly in southern France and Italy. In contrast, species with Mediterranean distributions are predicted to do well as the climate shifts, largely being predicted to find suitable climate directly northwards, but without suffering southern losses of suitability.

None of the Maxent models that retuned the lowest AICc value were identified as the most complex model possible (as would be the case from the default Maxent settings). No scarab used a model including the Hinge features (Table 7.2). The AUC values for each species are listed in Table 7.2, along with the 'pass' ($p<0.05$) or 'fail' ($p>0.05$) state of the AUC against the null models. Two species failed this test, *Osmoderma barnabita* and *Trichius fasciatus*, resulting in the remaining 11 species passing against the null model, showing significant associations between predictor variables and species presences. That both of these species failed to pass the null test may be due to their distribution being constrained by non-climatic factors, which could be accounted for in a number of ways, including habitat fragmentation or issues with accurate species taxonomy. Both species have complicated taxonomic histories

(see the section on *Osmoderma* below, and Chapter 3 on *Trichius* species boundaries), which may go some way toward explaining their not passing the null model test. Though both species were included in the hind- and –forecasting models, their results should be taken *cum grano salis* as there is no significant climate signal from their distribution data, though they are included for the sake of completeness and for broad comparisons. They are not considered in future discussions unless explicitly mentioned.

All AUC values were above 0.75 and therefore should show strong predictive performance (Fielding & Bell 1997), with seven of the species both passing the null test and having an AUC value above 0.85 (Table 2).

Though Beck et al. (2014) have warned against extrapolating too much from GBIF insect records due to a large number of biases inherent in the data (East-West sampling biases in particular), the null modelling approach will help to increase the robustness of the models produced, partially due to the pass/fail nature of the tests. Models from species that fail the null test should be treated as having inherent biases in the dataset and are not necessarily representative of the 'true' SDM for that species.

7.3.4 Species-by-species results

Cetonia aurata. The species responds most strongly to PCA1, but PCAs 2 & 3 are also important predictors of presence (Fig. 7.4). The species is known to be found in southern Greece and Portugal (Ahrens et al. 2013) but as no reliable high resolution localities from these regions could be found, they were not included in the study. Therefore the estimated niche of *C. aurata* is an underestimate of its true flexibility. That most of the British Isles is suitable for the species also suggests that its currently restricted distribution there is due to a scarcity of suitable larval habitat. Large areas of Sweden, Norway, Finland and northern Russia are predicted to become increasingly suitable for the species as long as there are suitable areas of vegetation matter, including organic compost (Ødegaard & Tømmerås 2000; Horák et al. 2013), though as France becomes increasingly unsuitable for the species Iberian/Pyrenean populations may become isolated from the rest of Europe. As a generalist species feeding on organic matter, the species isn't tied to ancient woodland and should largely do well in the face of climate change, despite southern reductions in suitability. Because PCA 1 is an important variable for the species, much of continental Europe is predicted to have been suitable for the species 6kya, facilitating a rapid spread from the

Iberian, Italian and Balkan southern refugia, and from northern refugia in France and Germany as indicated by the projection for 22kya. The oldest known fossil of the species in Britain is from Dumfriesshire (9640ya; Bishop & Coope 1977), which corresponds well with Britain being suitable for the species in the hindcast for both 22kya and 6kya. Isolation of Iberian populations and those in southern Italy may have happened in the last glacial maximum leading to the distinct haplotypes in these regions in the barcoding gene COI as identified by Ahrens et al. (2013). Additionally, that much of Germany, France, Bosnia and Herzegovina, and Serbia were apparently suitable for *C. aurata* indicates a post-glacial spread from these regions, as indicated by Ahrens et al. (2013). Genetic data and SDMs have therefore both predicted that northern refugia were more important than those in the south for post-glacial recolonisation of *C. aurata*.

Dorcus parallelipipedus: this extremely widespread species suffers from its own familiarity to entomologists as the distribution records for this species used in this study are a massive under representation of its true range (Fig. 7.5, Alexander et al. 2010). As such the species appears to be restricted to suitable habitat in west and southern Europe, with unsuitable climates in the north east, Russia and Estonia in particular, despite being found in both of these countries in reality. This species represents therefore a victim of the sampling biases inherent in certain public databases (Beck et al. 2014). However, the species is predicted to respond in a broadly similar manner to *Cetonia aurata*, with suitable climates shifting to the north east, and the potential for habitat fragmentation in Iberia, Italy and Turkey. PCA3 is the strongest predictor of occurrence in the species. As the species is usually found in areas with very low PCA3 values, and not at all in areas with high PCA3 values, this appears to have led to a similar scenario as for *C. aurata* in the predictions for 6kya, in that much of Europe is predicted to have been highly suitable for the species. Unlike *C. aurata* however, all *Dorcus* spp. are obligate saproxylics, and therefore require their host tree species to shift their ranges before *Dorcus* can. As in *C. aurata*, Iberia, Italy, France and Germany are predicted to have been refugia for the species, along with additional restricted areas of suitable climate in the British Isles, Switzerland, Sardinia, Corsica, and the Mediterranean Balkans.

Gnorimus nobilis: the species responds most strongly to PCA1, though it also shows a strong response to low values of PCA3 (Fig. 7.6). That the species finds suitable climates throughout the British Isles is contrary to the result of Whitehead (2003) who predicted that the climate was not suitable in the north of the British Isles for this species. Instead, this suggests some different scenarios: that given suitable habitat (which is currently lacking

throughout the British Isles, the species should be more widespread, and/or that British populations may be locally adapted to the British climate and therefore don't make use of suitable habitat outside of their current climatic envelopes. The species has been present in Britain for over 9,500 years (Osborne 1974), matching the hindcasting predictions. Generally the species avoids extremes, and is predicted to show a similar response to the above species as climate change continues, with suitable climates opening up to the north east. This is likely to isolate populations in Iberia, southern Russia and may lead to elevational shifts in the Balkan mountain ranges (Thomas et al. 2006). Similar refugia to those predicted for *Cetonia aurata* are also predicted for *G. nobilis*, and correspond to genetic data (Chapter 5).

Gnorimus variabilis: As predicted by Trizzino et al. (2013), *G. variabilis* occupies a different climatic niche to *G. nobilis*, being most responsive to PCA3, but not at low values, having a broad tolerance to PCA1, and a very narrow tolerance to PCA2 (Fig. 7.7). Though these responses appear to be broadly similar to those of *G. nobilis* at the wide scale, *G. variabilis* is more suited to lowland climates and isn't predicted to be suited to the mountain ranges (such as the Alps, Pyrenees, Apennines and the Carpathians), though under the climate change scenario these ranges become increasingly suitable for the species. The British Isles, Germany, southern Sweden and Russia are predicted to become more suitable for *G. variabilis* as the climate warms, whilst southern areas become increasingly less suitable. Hindcasting the SDM indicates that France, Italy and Iberia are the key areas for predicted glacial refugia, whilst additional areas in northern France and Iberia become suitable, though other suitable areas in the Balkans may be unconnected to the primary predicted distribution centre.

Lucanus cervus: *Lucanus cervus* shows a very similar predicted presence to *Cetonia aurata*, though it seems less well adapted to northern latitudes and northern Russia (Fig. 7.8). The model shows clear relationships with climate in mountain ranges in the south, with lowlands (Belgium, England) being more important in the north. As with *C. aurata*, PCA1 contributes the most to the model. Forecasting the model, the northern European plain (Belarus, Lithuania, Latvia, Estonia and North West Russia) becomes increasingly suited to *Lucanus cervus*, whilst other regions suited to the species become increasingly isolated (Portugal, the southern Caucasus). As predicted by Cox et al. (2013), hindcasting 22kya indicates that suitable areas of climate were found in Iberia, Italy, the southern Balkans, and around Syria and Lebanon, which correspond to genetically distinct units (McKeown *pers. comm.*), along with a possible refugia in France. Using the 6kya model as a midpoint between glacial retreat

and the modern day, this suggests that the climate in the mid-Balkans and central Europe was suited to *L. cervus*, representing a possible recolonisation pathway, whereas although much of Iberia became suitable for *L. cervus*, the regions to the north were less so, and therefore the species may not have spread from Iberia across Europe, but instead from the Balkans.

Osmoderma barnabita: this species failed the null model, therefore the results from the species will only be discussed briefly with few comparisons (Fig. 7.9). As with *C. aurata* and *L. cervus* the species is predicted to move increasingly toward the northern European plains, and into higher mountainous regions than previously (the Carpathians and Dinaric Alps especially). Glacial refugia are predicted throughout France, Italy and northern Iberia, with more fringe areas in the Balkans and the south Caucasus. This is at odds with its currenteastern distribution, especially as the 6kya projection suggests much of France and Italy was unsuited to the species, but that the Balkans and Turkey were, which may represent the true glacial refugia for this species.

Osmoderma eremita: this species isn't found in the British Isles (though there is a subfossil from Essex approximately 337kya (Roe et al. 2009), but this is equally likely to be from *O. barnabita*), but this appears not to be due to climatic reasons as much of England and Scotland is apparently suited to it, and instead may be due to the low dispersal capability of the species (Ranius & Hedin, 2001), or cryptic extinction (Fig. 7.10). Certain regions toward the south Balkans and north east Europe are also predicted to be suitable, though here the species is replaced by its relative *O. barnabita*. As the climate warms the species is predicted to move increasingly eastward, which may bring it into contact with *O. barnabita*. Exactly what effect this will have on the two species, and whether they will successfully hybridise is unknown. Hybrids between the two have not been identified, though as they were only recognised as truly separate species relatively recently (Audisio et al 2009; Landvik et al. 2013) this is not a surprise. When collecting data for *Osmoderma*, sample positions were compared by eye to the predicted distribution maps in Audisio et al. (2009) and the species name was changed accordingly. Samples from areas where the two species could overlap (central Germany in particular) were omitted. Areas where the two species should co-exist will prove extremely interesting as the two species become increasingly in contact, and research looking to identify putative hybrids and gene flow between the two species induced by climate change should be extremely fruitful. In general, few areas with suitable climate for *O. eremitia* are predicted to be lost under climate change (though those lost will be from Iberia and Italy, potentially areas with high genetic diversity), and the species may not fare

too badly so long as there are good numbers of veteran trees to use. The species is predicted to have a very similar pattern of climatically-suitable sites to *Gnorimus variabilis*, with France, the Pyrenees and Italy being suitable. This is another species which likely found thr 6KYA climate to be better suited to it than the modern climate.

Oryctes nasicornis: the species clearly has a very broad climatic niche, strongly responding to both PCA1 and 3 (Fig. 7.11). As most of Europe is apparently suited to it, its absence from the British Isles may be down to poor dispersal, or an unrecorded extinction due to habitat loss, though again its wide range of food substrates makes this unlikely. The species is not known to be a poor disperser however: extrapolating from Dubois et al. (2009) indicated the species has a maximum dispersal distance of 11km in a single flight, and although this is only a third of the distance needed to cross the Strait of Dover, with lower sea levels in in the early stages of glacial retreat and a strong wind it is not inconceivable that the species could have dispersed to Britain. *Gnorimus nobilis* is a considerably poorer flyer (Whitehead 2003; Bates *pers. comm.*) and yet has been in England for over 9,000 years (Osborne 1974). Only very small areas are predicted to become less climatically suitable for *O. nasicornis*, but additional areas in northern Russia and Finland are predicted to become suitable. The three typical areas for European glacial refugia (Iberia, Italy and the Balkans) are all predicted to have been suitable for this species, along with much of France and Germany, during the last glacial maximum. Additionally, much of the British Isles, including Ireland, has also been predicted to have been suitable 22kya, which makes the species absence from there more enigmatic.

Propomacrus bimucronatus: a rare species thought extinct for much of the 1900s (Young 1989) confined to the southern Balkans through to the Middle East, it responds extremely strongly to PCA2, and is predicted to find large areas of new climatically suitable habitat opening up as climate change advances (Fig. 7.12). Young (1989) predicted the species would be found in at least "Albania, Bulgaria, Armenian and Georgian Russia, throughout Greece, northern Iraq, Lebanon and throughout Yugoslavia", and the discovery of the species in southern Bulgaria (Bekchiev & Zlatkov 2010) gives weight to some of his predictions. However, the climatic envelope of the species indicates that Georgia, Armenia and Serbia are not suited to this species, though the southern reaches of the Dinaric Alps may be suitable. In addition, Young's prediction that the species may find mountainous Italy climatically suitable appears to be slightly inaccurate: southern Italy, west of the Apennines, Sardinia and Sicily, as well as southern Iberia, seem to be suitable. Though the species is extremely large and charismatic, its cryptic nature means that it is conceivable that the species has a wider range

than expected. Hindcasting also suggests that although much of its current range was still suitable 22kya, by 6kya the Greek and Cypriot/Turkish climatic regions were split, implying some isolation between the two populations, which may add weight to the hypothesis that the species is made up of two separate evolutionary units (Hadjiconstantis et al. 2015). The exact relationship of the two *Propomacrus* species to each other is poorly understood, but *P. cypriacus* may be a subspecies of *P. bimucronatus*, though this data has not appeared in the peer-reviewed literature (Hadjiconstantis et al. 2015). Due to the similarity of their habitat and close relationship they are here grouped as one species and referred to as *Propomacrus bimucronatus sensu lato* for the rest of the text.

Protaetia mirifica: this Mediterranean species shows an extremely strong response to PCA1, being split into four distinct regions: Iberia, southern and western France, Italy and the southern Balkans (Fig. 7.13). Future projections indicate suitable climates being found in Sweden, Estonia and Finland, areas far north of the current distribution. There are large oaks in these regions used as habitat by *Osmoderma*, so it's not inconceivable that *P. mirifica* will find these regions suitable, though it is likely to be a poor disperser and may not be able to move into this suitable habitat over the next 50 years. However, from the projections its current distribution is not expected to become unsuitable for it, so it isn't likely to be at risk of climate-shift induced extinction. Hindcasting shows the four populations centres are likely to have been separated for thousands of years, and thus may represent distinct evolutionary units. As a charismatic, vulnerable species (Nieto et al. 2010) population genetic work should shed interesting light on its phylogeographic patterns.

Protaetia speciosissima: this species displays a very similar climatic envelope to *Lucanus cervus*, associating with mountains in the south of the range, and flatter land in the north (Fig. 7.14). In an almost identical manner to *L. cervus* it is expected to find suitable climate in the northern European plain as climate change continues, whilst southern populations are at risk of extinction. As with *Oryctes* and *Osmoderma eremita*, the British Isles seem climatically suited to the species, so its absence there is unusual. Hindcasting shows likely glacial refugia in north east Iberia, France, Italy and certain areas in the Balkans. *Protaetia speciosissima* is listed as "*P. aeruginosa*" in older literature and in some databases, but is here referred to as *P. speciosissima* following Krell et al. (2012).

Trichius fasciatus: the species failed the null model, and will therefore only be discussed briefly (Fig. 7.15). The true distribution of this species, and indeed exactly which genetic unit

of *Trichius* has been identified by the databases GBIF uses, is impossible to ascertain due to the complex species boundaries between species in the genus (see Chapter 3). *T. fasciatus* appears to be a species adapted to cold regions, whereas *T. gallicus* and *T. sexualis* are more common in the south. These latter two species would be expected to follow similar patterns to *Cetonia aurata* and *Protaetia speciosissima*, whereas, as is indicated by the model, *T. fasciatus* is suited to mountain ranges in the south, and much of northern Europe. As with *Dorcus parallelipipedus*, it may be a victim of its own common nature, and may not be recorded from many areas, despite its distribution there. Southern populations may run out of mountain to live on if climate change continues, whilst the species may have found glacial refuge further north than most other saproxylic species, utilising France, Germany, Belgium, the Netherlands, Switzerland, Norway and the British Isles in the North, and the Apennines, Alps, Croatia and Portugal in the south. The model predictions line up with the presence of a *Trichius* species in Britain (Kent, south east England) 12kya (Coope 1998), implying that it survived in southern England throughout the last glaciation. However, its failure to differ to the null model limits the conclusions that can be drawn.

Valgus hemipterus: another poorly recorded species, *V. hemipterus* avoids 'extreme' habitats with high PCA3 scores, and is generally found associated with lowland areas without large climatic fluctuations (Fig. 7.16). As with many other saproxylic scarabs, it is expected to spread to the northern European plains, though the climate envelope for southern populations isn't predicted to shift as dramatically for this species as for others. It is also predicted to find suitable climate in Turkey, Iraq and Syria as time advances. Six major areas for putative glacial refugia are also indicated: Iberia, northern Africa, France, Italy, the southern Balkans and Syria and Iraq. This adaptable species has been found in previous interglacials in the British Isles, and has been found in sediments from 3.6-3.8kya in Britain (Campbell & Robinson 2007), showing that the species is now extinct in Britain, likely due to habitat loss.

In general, many specialists and generalists show a similar predicted response to climate change: localised extinctions in southern areas surrounding the Mediterranean, expansions vertically up mountain ranges, and expansions north-eastwards into the northern European plain (*Cetonia aurata, Dorcus parallelipipedus, Gnorimus nobilis, G. variabilis, Lucanus cervus, Protaetia speciosissima* and *Valgus hemipterus*), a common pattern for European fauna (Hickling et al. 2006; Thomas et al. 2006, Lindner et al. 2010). These species are a mixture of habitat generalists (*C. aurata, D. parallelipipedus, L, cervus*, and *V. hemipterus;* Table 7.1) and specialists (*G. nobilis, G. variabilis* and *P. speciosissima*, all specialists in

veteran tree rot-holes), as initially predicted. Habitat for the specialists takes many hundreds of years to form (Warren & Key 1991) as the host trees must become large and old enough to support fungi which produce cavities in the wood formed by heartwood rot (Sebek et al. 2013). This lag time between climate change and the development of suitable habitat is significant even for insects where habitat changes and formation can be measured on a decadal scale (Warren et al. 2001; Menéndez et al. 2006), and for species where habitats take hundreds of years to form this lag could be catastrophic.

Iberia, Italy and southern France are predicted to have been climatically suitable for *C. aurata, D. parallelipipedus, G. nobilis, G. variabilis, L. cervus, Os. barnabita, Os. eremita, Oy. nasicornis, P. mirifica, P. speciosissima* and *V. hemipterus*, with northern refugia more important for *T. fasciatus* and Balkan refugia more important for *P. bimucronatus*. Additional refugia in the Balkans may also have been important for *C. aurata, G. nobilis, L. cervus, Os. barnabita, Oy. nasicornis, P. mirifica* and *V. hemipterus,* with putative northern refugia (northern France and Germany, and the British Isles) predicted for *C. aurata, G. nobilis, G. variabilis, Os. barnabita, Os. Eremita, Oy. nasicornis, P. speciosissima, T. fasciatus* and *V. hemipterus*. Thus there is little separation between the glacial refugia predicted to have been used by habitat generalists and specialists. A clear distinction in the responses of Mediterranean specialists and those from temperate climates can be made, with climatic generalists strongly predicted to have used northern refugia.

Hindcasting to predict glacial refugia reveals that patterns predicted from population genetics generally match well with our findings from SDMs, and in line with other studies combining SDMs and phylogeography (Forester et al. 2013; Dalmaris et al. 2015):

- *Cetonia aurata* – Ahrens et al. 2013. Southern refugia in Iberia, southern Italy and Sicily, the Balkans and Lebanon. The similarity of COI haplotypes in the area between France, Scandinavia and the Balkans suggests that either the species recolonised northern areas as the glaciers retreated from French and Balkan refugia, or that in fact much of continental Europe remained suitable for it, with one large, widespread population being found in France, southern Germany and into the Balkans. Haplotypes unique to Iberia and southern Italy don't appear to have contributed to the species' expansion. The hindcast for 22kya strongly supports this, with mountainous regions of Iberia and Italy being predicted as suitable, along with much of France and Germany, and a split population in the Balkans (over Serbia,

Croatia and Bosnia and Herzegovina). A separate region in Lebanon is also strongly predicted to have been suitable.

- *Lucanus cervus* – Cox et al. 2013; McKeown *pers. comm.* Refugia in Iberia and/or Italy likely contributed to the post-glacial expansion of the species, with possible secondary introgression from a Caucasus or Russian refugium. Separate refugia in the Balkans and Turkey didn't contribute haplotypes of COI to the northern populations. Our data strongly support refugia in Iberia, Italy and the western Balkans, but also predict a northern refugium in France, southern Germany and Switzerland. The lack of genetic differentiation between these regions may be due to a population of the species spreading from Iberia through France and Switzerland into Italy, stopping genetic differentiation from taking place in these refugia. Additional isolated refugia are predicted in the Caucasus and Lebanon which may match with restricted haplotypes identified in Cox et al. (2013) (Chapter 2)
- *Gnorimus nobilis* – Chapter 5. Results from COI indicate refugia in Iberia, southern France, southern Italy, Greece and a separate population/subspecies in Ukraine and Russia. The finding of a COI haplotype from the eastern population in Macedonia indicates that the eastern population spread from the Caucasus through to the Balkans, but was genetically swamped from an expansion from the west, the implication being that the eastern population is adapted to a cold eastern continental climate. The results from 22KYA strongly support refugia in Iberia, southern France and throughout Italy, but also suggest that northern France, Germany, southern England and Ireland were suitable for the species, indicating a northern refugium. Additional areas in the Balkans were also weakly supported, but only a very small region in the Caucasus is supported as being suitable. This may support the hypothesis that the eastern population is adapted to a different climate to the western and thus isn't correctly modelled in our approach (locality data for eastern haplogroup populations were included, but represent under 1% of the locality samples).
- *Trichius fasciatus* – Chapter 3. As the model failed the null test, the results can't be tightly compared to results from genetic data. Little genetic differentiation was found in populations from France to central Russia in *T. fasciatus*, suggesting that the species was not restricted to southern refugia in Europe. The 22KYA results show that the species likely found much of Europe suitable with an uninterrupted band of

suitable climate from Ireland to Serbia and Romania, corroborating the genetic results.

As predictions from the SDM approach have been to some extent supported by population genetics and phylogeographic patterns, broad hypotheses based on the SDM patterns indicated may be drawn out for the other climatic generalist species used in the present study. If beetle species persisted in multiple unconnected refugia throughout the last glacial maximum (LGM), or have had climatically mediated population contraction-expansions in the past, then the phylogeny of the species should show deep divisions in gene trees utilised in phylogeography (COI being the best example in animals), resulting from variable patterns of isolation and drift in frequencies among different haplotype clades between the various unconnected glacial refugia. Subsequent population expansions and spread after the LGM will result in three different predicted phylogeographic patterns across northern Europe depending on how many and which refugia colonising populations originate from:

i. If species populations have been restricted to multiple southern refugia during the LGM and have subsequently spread from more than one of these refugia, then much of their current range should be populated by mixtures of different haplotype clusters (clades) originating from the various refugia, and distinct differentiation between this diverse population and any relict populations which have not expanded beyond their respective refugia (as suggested for *Lucanus cervus* and *Cetonia aurata*).
ii. Alternatively, if the populations have been restricted to multiple southern refugia during the LGM, but have subsequently spread from one refugium only, then much of their current range should be populated by very few common and related haplotypes (i.e. little deep genetic differentiation), but with distinct differentiation between the relict populations which have not expanded beyond their respective refugia (as suggested for *Gnorimus nobilis*).
iii. Finally, for species that are less climatically restrained, there should be very few phylogeographic differences between present populations due to the past species distribution being largely panmictic (or at least potentially connected) across much of Europe throughout the last glacial maximum (as suggested for *Trichius fasciatus*).

Predictions for the other northern European species are as follows:

- Expansion from multiple refugia: *Dorcus parallelipipedus*, *Protaetia speciosissima*, *Valgus hemipterus* and *Gnorimus variabilis* from Iberia, France, east from Italy, and the Balkans.
- Expansion from a single refugium: *Osmoderma barnabita* from the Balkans (though see below), *Osmoderma eremita* from France and Germany
- Limited recent expansion due to panmixia: *Oryctes nasicornis*, though as this species has a large number of subspecies panmixia may only apply to the populations across northern Europe

The Mediterranean specialists predicted to expand northwards (*Protaetia mirifica* and *Propomacrus bimucronatus*) may be able to move into hollow trees that are no longer suitable for other species, so long as these are within their dispersal limits. Unlike every other species studied, both Mediterranean species appear to have had more areas of suitable climate 22KYA than 6KYA, the mild climate 6KYA being apparently unsuited to them. The species should therefore have deep genetic divergence between refugia, a hypothesis implied by Hadjiconstantis et al. (2015) on *Propomacrus* with deep splits between Turkish and Greek populations.

Because the species boundaries between the two *Osmoderma* studied are poorly understood (Audisio et al. 2009; Landvik et al. 2013), climate change may bring some complicated scenarios to bear. *Osmoderma eremita* is predicted to expand into the current range of *O. barnabita*, which may facilitate hybridisation between the two species. Exactly what effect this will have on both species is unknown, but should be focussed on by molecular studies within the predicted expansion areas. Both species are predicted to expand into the northern Europe plains, though the concerns about the availability of large hollow trees are as applicable for them as they are for *Gnorimus*. In contrast to southern species, the cold-adapted *Trichius fasciatus* is at serious risk of its distribution being fragmented by climate changes into distinct populations separated by unsuitable habitat, though there is still potential niche space for it to expand north eastwards. However, as it also failed the null model such conclusions are tentative and should be revisited with further study. The predictions for *Oryctes nasicornis* suggest that it is such a climatically adaptable species that it may be little affected by climate change, being likely to suffer no major southern losses of climatically suitable habitat whilst expanding toward the north east of the study area.

The non-intuitive observation that a large number of beetle species are predicted to have found much of Europe more climatically suitable 6KYA than today (*Cetonia aurata*, *Dorcus parallelipipedus, Gnorimus nobilis, Osmoderma eremita, Oryctes nasicornis, Trichius fasciatus* and *Valgus hemipterus*) matches with knowledge about the climate in this period being considerably warmer and less variable than modern climates (Peterken 1993; Andersen et al. 2006; Seppä et al. 2009). The period 8-4KYA has been linked to rapid post-LGM population and range expansions in deciduous trees (Peterken 1993, Birks & Willis 2008a; Seppä et al. 2015) and the associated forest fauna (Hofreiter & Stewart 2009; Massilani et al. 2016). Examples are known from other taxa where allopatric species inhabiting the same ecological niche, but with different climatic tolerances, have varied in historical abundance related to shifting climates: red deer (*Cervus elaphus*) have been more abundant in Norway during warm periods compared to moose (*Alces alces*) which, as ungulates adapted to deep snow, are more common in cold periods (Rosvold et al. 2013). Additionally, northern area extinctions of species that spread poleward in the immediate post-LGM warm period have occurred as the climate cooled again (Sommer et al. 2011). Similar processes could have applied with rot-hole dwelling scarabs.

Why both *Trichius fasciatus* and *Osmoderma barnabita* failed to demonstrate a climatic signal in their distributions that was significantly stronger than those predicted from null species distributions is puzzling. *Osmoderma barnabita* had a reasonable sample size (n=353), whilst *T. fasciatus* had the second highest number of samples (n=5976), with both species showing samples from across Europe. As noted by Williams et al. (2015) discussing parasites, host and habitat distribution may be as or more important for predicting tick distributions. As both beetle species are found in rot-holes, it may be the distribution of veteran host trees (birch for *T. fasciatus*, oak and beech for *O. barnabita*) that are more important predictors of distribution than is climate for both species. AUC values for both species were also high (0.8299 and 0.82595 respectively) indicating strong model performance.

Improvements to the modelling approach detailed here could involve utilising ensemble modelling with multiple general circulation models and taking forward the mean projection for hind- and forecasting (Forester et al. 2013), or including models of the habitat. As the specialist saproxylics modelled here are dependent on trees and fungi, producing SDMs for known host species will allow for changes in saproxylic habitat to be inferred, potentially adding an additional layer for use in the Maxent program alongside the principle components.

That veteran trees are some of the most important habitats for biodiversity is without doubt (Vera 2000; Gibbons et al. 2008; Bouget et al. 2014). However, little has been done to quantify exactly where hotspots of ancient tree presence are across the western Palearctic. This field should become increasingly important as the climate shifts, opening new areas of habitat toward the northern European plains for the veteran tree/rot-hole specialist beetles. But it is unknown if these areas have enough suitable ancient trees for the beetles to utilise. Though none of these specialist species are particularly common even in forests where they are present, they perform decomposition roles that no other species can, and as such are vitally important ecosystem engineers (Mouillot et al. 2013; Sánchez-Galván et al. 2014; Micó et al. 2015; Landvik et al. 2016). That specialists respond less well to climate change scenarios than generalists is well known (Slayter et al. 2013), and as such the three 'expansionist' generalists identified above (*Cetonia aurata, Lucanus cervus* and *Oryctes nasicornis* should do well under the climate modelled (despite some southern extinctions), providing there is suitable dead wood for the two saproxylics, and leaf humus and detritus for *C. aurata*.

7.4 Conclusions

Despite the fungi and/or tree species that saproxylics depend on being cosmopolitan, saproxylic beetle distribution is tightly linked to the presence of these species at microclimatic scales rather than gross climate. Due to this, and the wide climatic niches inhabited by most species used in this dataset (with the exception of Mediterranean specialists), future research time producing SDMs for saproxylic Coleoptera should be best focussed on producing and testing models robustly, hind- and fore-casting changes, and focussing on multi-species interactions (Svenning et al. 2014). These can then be used in combination with genetic methods to better understand such research topics as population connectivity, predicting glacial refugia, and identifying leading range edges and predicting areas into which species will expand.

Evidence that saproxylics have already begun to expand their ranges is strong: see Seidl et al. (2011) & de la Giroday et al. (2012) on bark beetles (Scolytinae), Buse et al. (2013) on the Jewel beetle *Coraebus florentinus* (Buprestidae), Ge et al. (2014) on *Anoplophora chinensis* (Cerambycidae), and Köhler (2014) on rare German saproxylics (also see Hickling et al. 2006 for a general review across taxa). Much research focus has investigated saproxylics

contributing to tree decline, but distinctions need to be made between ‘pests’ with different ecologies which produce different types of dead wood suitable for different saproxylics. The European Capricorn beetles (*Cerambyx* spp., Cerambycidae) for instance attack oaks, speeding up senescence and the development of rot-holes/tree hollows (Buse et al 2008 a.), opening up habitat for a large number of rare beetles associated with tree hollows (Buse et al. 2008 b.) by producing a succession of decay (Müller et al. 2014). Expansions of such species are likely to be good for hollow-dwelling Cetoniinae (Micó et al. 2015), which are important secondary colonisers after an initial attack, affecting resource availability for smaller species by enlarging cavities and allowing their own faeces to build up (Buse et al. 2008; Schenke 2010; Sánchez-Galván et al. 2014; Micó et al. 2015). However, expansions of species which kill trees before they can senesce (Scolytinae in particular) may be good for generalist species which don’t need ancient or hollow trees (*Dorcus* and *Lucanus* for instance).

Expansions of different early-colonising saproxylics will be a mixed blessing for saproxylic scarabs. Whilst generalist species will likely do well under climate change scenarios as new climatically suitable areas open up, specialists may run into trouble if the ancient trees they rely on die faster than they can be replaced. With ancient trees (and their habitat) taking 100-400 years to grow, lags between the development of suitable habitat and changes in climate may squeeze hollow-dwelling saproxylics into ever dwindling areas of climatically suitable ancient trees. Whilst forestry practitioners may see expansions of wood-boring species to be a cause for concern as they encourage die back of large trees (and profits), they may be the only thing facilitating the climatic-driven expansion of rot-hole dwelling scarabs.

Table 7.1. The Scarabaeoidea used in the study. The taxonomic hierarchy for each species is given, along with a basic description of its niche and conservation status

Species and taxonomic hierarchy	Larval ecology	Conservation Status	References
Dorcus parallelipipedus Lucanidae, Lucaninae, Dorcini	Flexible on decaying deciduous trees, particularly with white-rot fungi. Feeds on heartwood rot and fallen logs and branches. Facultative in heartwood rot. Obligate saproxylic. Life cycle length usually two years.	Least Concern	Alexander et al. 2010
Lucanus cervus Lucanidae, Lucaninae, Lucanini	Flexible on deciduous trees attacked by white-rot fungi, but most commonly oak (*Quercus* spp.). Requires large wood deposits at or below the soil surface. Not found in heartwood rot. Obligate saproxylic. Life cycle length three to seven years.	Near Threatened	Nieto et al. 2010 Harvey et al. 2011
Oryctes nasicornis Scarabaeidae, Dynastinae, Oryctini	Feeds widely on decaying deciduous plant material, including wood, roots and leaf humus. Facultative in heartwood rot. Facultative saproxylic. Life cycle two to four years.	Not assessed. Unlikely to be at risk considering its massive range and flexibility	Colón 2003
Propomacrus bimucronatus and *P. cypriacus* Scarabaeidae, Euchirinae	Poorly understood, but appears to be a specialist in oak heartwood rot. Occasionally feeds in heartwood rot of other species (carob, cherry). Obligate saproxylic in heartwood rot. Life cycle length unknown, likely to be two to four years.	*P. bimucronatus*: Near Threatened *P. cypriacus*: Critically Endangered	Nardi et al. 2010 Nieto et al. 2010
Valgus hemipterus Scarabaeidae, Cetoniinae, Valgini	Rotten wood from deciduous trees, especially on fallen rotten logs and stumps. Obligate saproxylic facultative in heartwood rot. Yearly life cycle.	Not assessed. Unlikely to be at risk	Coope 2010
Trichius fasciatus Scarabaeidae, Cetoniinae, Trichiini	Obligate saproxylic heartwood rot feeder, most commonly on tree-birch (*Betula* spp.), but some host flexibility. Yearly lifecycle.	Least Concern	Mannerkoski et al. 2010c
Gnorimus nobilis Scarabaeidae, Cetoniinae, Trichiini	Obligate saproxylic heartwood rot feeder, flexible on deciduous trees, especially oak and beech (*Fagus* spp.). One to three year lifecycle.	Least Concern	Mannerkoski et al. 2010a Bates et al. 2014

Gnorimus variabilis Scarabaeidae, Cetoniinae, Trichiini	Obligate saproxylic heartwood rot feeder, flexible on deciduous trees, especially oak and chestnut (*Castanea* spp.). Rarely in coniferous trees. One to three year lifecycle.	Near Threatened	Mannerkoski et al. 2010b
Osmoderma barnabita Scarabaeidae, Cetoniinae, *incertae sedis*	Obligate saproxylic heartwood rot feeder, flexible on very large deciduous trees. Life cycle usually two years.	Near Threatened	Alexander et al. 2010
Osmoderma eremita Scarabaeidae, Cetoniinae, *incertae sedis*	Obligate saproxylic heartwood rot feeder, flexible on very large deciduous trees. Life cycle usually two years.	Near Threatened	Nieto et al. 2010
Cetonia aurata Scarabaeidae, Cetoniinae, Cetonini	Flexible feeders on decaying plant matter, including compost, mammal faeces, humus, and wood. Facultative saproxylic. Life cycle usually two years.	Not assessed. Unlikely to be at risk	Ahrens et al. 2013
Protaetia mirifica Scarabaeidae, Cetoniinae, Cetonini	Obligate saproxylic heartwood rot feeder, usually in large oaks, in particular *Quercus pubescens*. Life cycle generally two years.	Vulnerable	Nieto et al. 2010
Protaetia speciosissima Scarabaeidae, Cetoniinae, Cetonini	Obligate saproxylic heartwood rot feeder, but flexible on large deciduous trees. Life cycle generally two years.	Near Threatened	Mason et al. 2010

Table 7.2: the number of species distribution records, null model results, median AUC values, and their recommended Maxent fitting parameters for the 13 study species.

Species	Number of records	Null models	AUC vales	Recommended Maxent internal fitting parameters
Dorcus parallelipipedus	1940	Pass	0.885	Linear, Quadratic, Product
Lucanus cervus	6380	Pass	0.895	Linear, Quadratic, Product, Threshold
Oryctes nasicornis	1018	Pass	0.783	Linear, Quadratic, Product
Propomacrus bimucronatus and *P. cypriacus*	39	Pass	0.932	Linear
Valgus hemipterus	497	Pass	0.860	Linear, Quadratic, Product
Trichius fasciatus	5976	Fail	0.830	Linear, Quadratic, Product, Threshold
Gnorimus nobilis	1358	Pass	0.841	Linear, Quadratic, Product
Gnorimus variabilis	345	Pass	0.843	Linear, Quadratic
Osmoderma barnabita	353	Fail	0.826	Linear, Quadratic, Product
Osmoderma eremita	1680	Pass	0.888	Linear, Quadratic, Product
Cetonia aurata	4955	Pass	0.835	Linear, Product, Threshold
Protaetia mirifica	39	Pass	0.862	Threshold
Protaetia speciosissima	165	Pass	0.862	Linear, Quadratic

Table 7.3. The percentage of variation within the complete bioclimatic dataset explained by each principle component. Together the first three principle components explain 82.75% of the total variation

Component	Percentage explained
1	39.90
2	33.14
3	9.708
4	7.796
5	4.793
6	2.838
7	0.763
8	0.409
9	0.240
10	0.135
11	0.099
12	0.073
13	0.041
14	0.021
15	0.019
16	0.015
17	0.005
18	0.001
19	2.85E-29

Table 7.4: the 19 Bioclimatic variables used in the study, the climatic factor they explain, and the proportion of the variation in the first three principle components that they explain. The most extreme/important 10 variables are highlighted in red for each PC.

Variable	Explains	PCA1	PCA2	PCA3
BIO1	Ann. mean temp	-0.229	0.283	-0.146
BIO2	Mean diurnal temp range	0.019	0.271	0.136
BIO3	Isothermality (Bio2 div Bio7)	-0.285	0.152	-0.086
BIO4	Temp. seasonality	0.320	0.016	0.141
BIO5	Max temp. warmest	-0.049	0.356	-0.019
BIO6	Min temp. coldest	-0.305	0.179	-0.184
BIO7	Temp ann. range	0.307	0.074	0.193
BIO8	Mean temp. wet 1/4	0.167	0.087	-0.291
BIO9	Mean temp. dry 1/4	-0.265	0.236	0.045
BIO10	Mean temp. warm 1/4	-0.083	0.338	-0.077
BIO11	Mean temp. cold 1/4	-0.298	0.204	-0.158
BIO12	An precip.	-0.270	-0.234	0.127
BIO13	Precip. wettest	-0.258	-0.164	0.349
BIO14	Precip. driest	-0.179	-0.297	-0.224
BIO15	Precip. seasonality	-0.016	0.194	0.574
BIO16	Precip. wettest 1/4	-0.264	-0.173	0.327
BIO17	Precip. driest 1/4	-0.198	-0.286	-0.211
BIO18	Precip. warmest 1/4	-0.012	-0.339	-0.108
BIO19	Precip. coldest 1/4	-0.315	-0.070	0.250

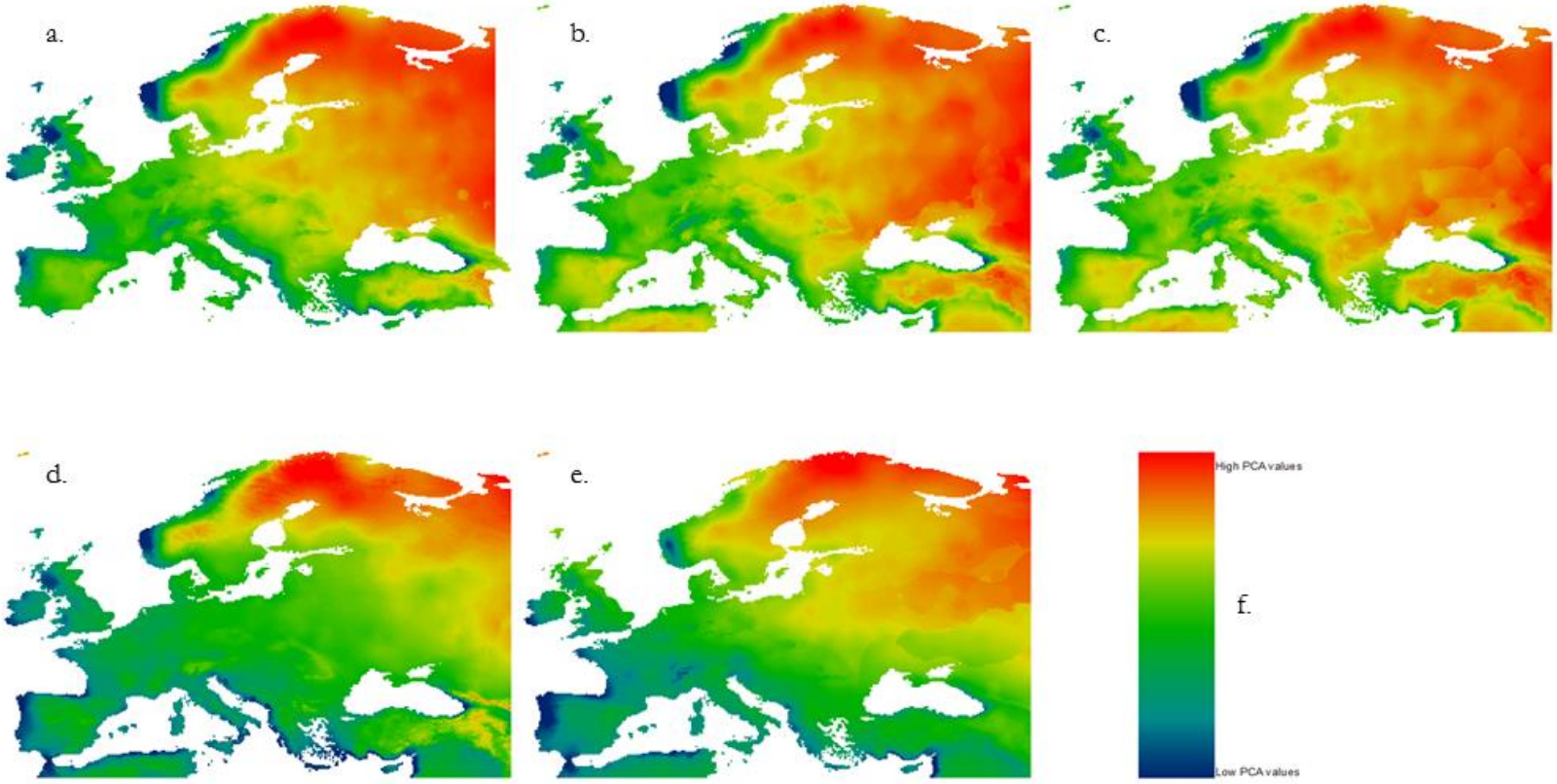


Figure 7.1: Scores for Principle Component 1 across the study area at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – arbitrary scale for the highest and lowest score for each individual projection

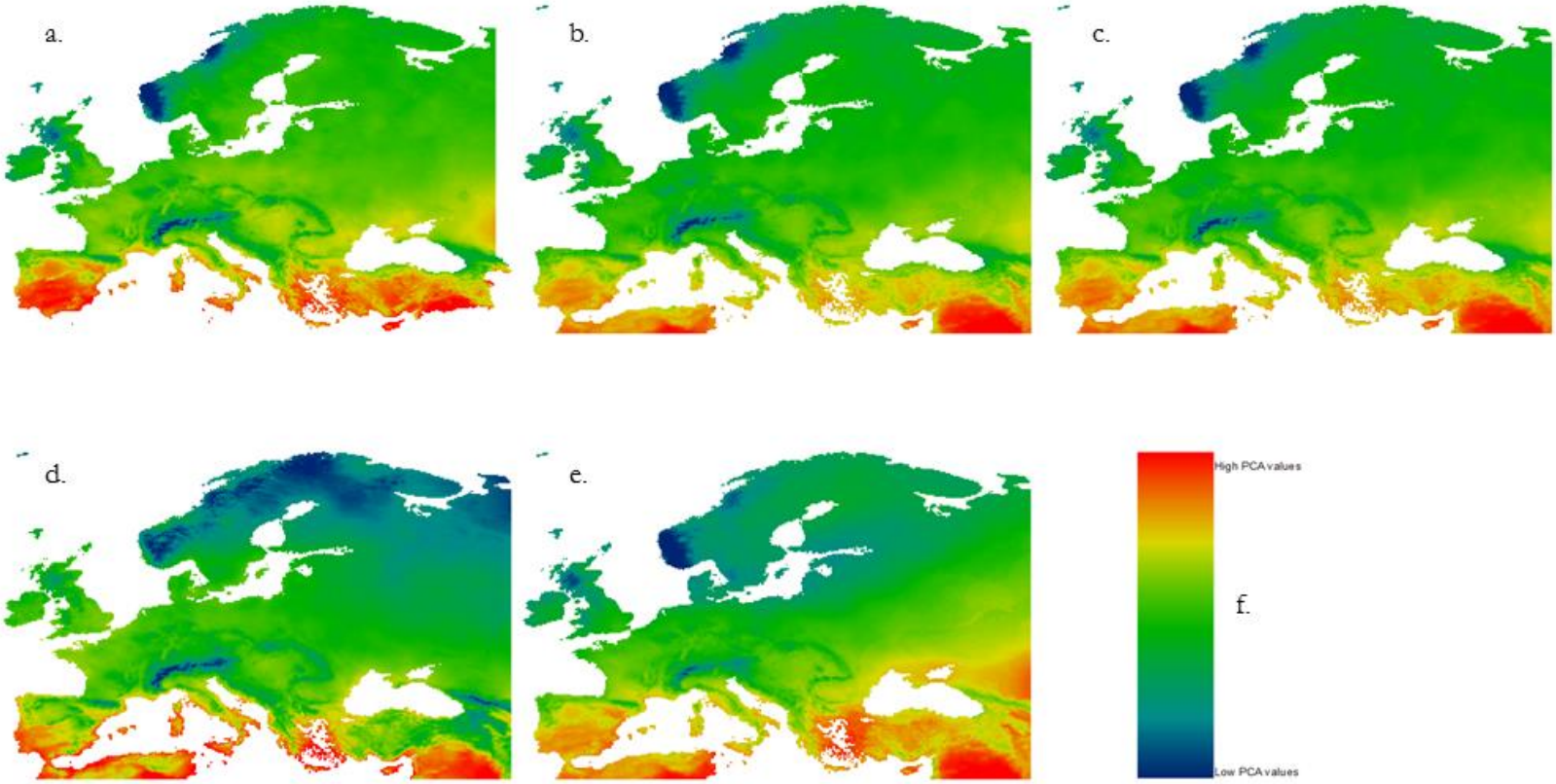


Figure 7.2: Scores for Principle Component 2 across the study area at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – arbitrary scale for the highest and lowest score for each individual projection

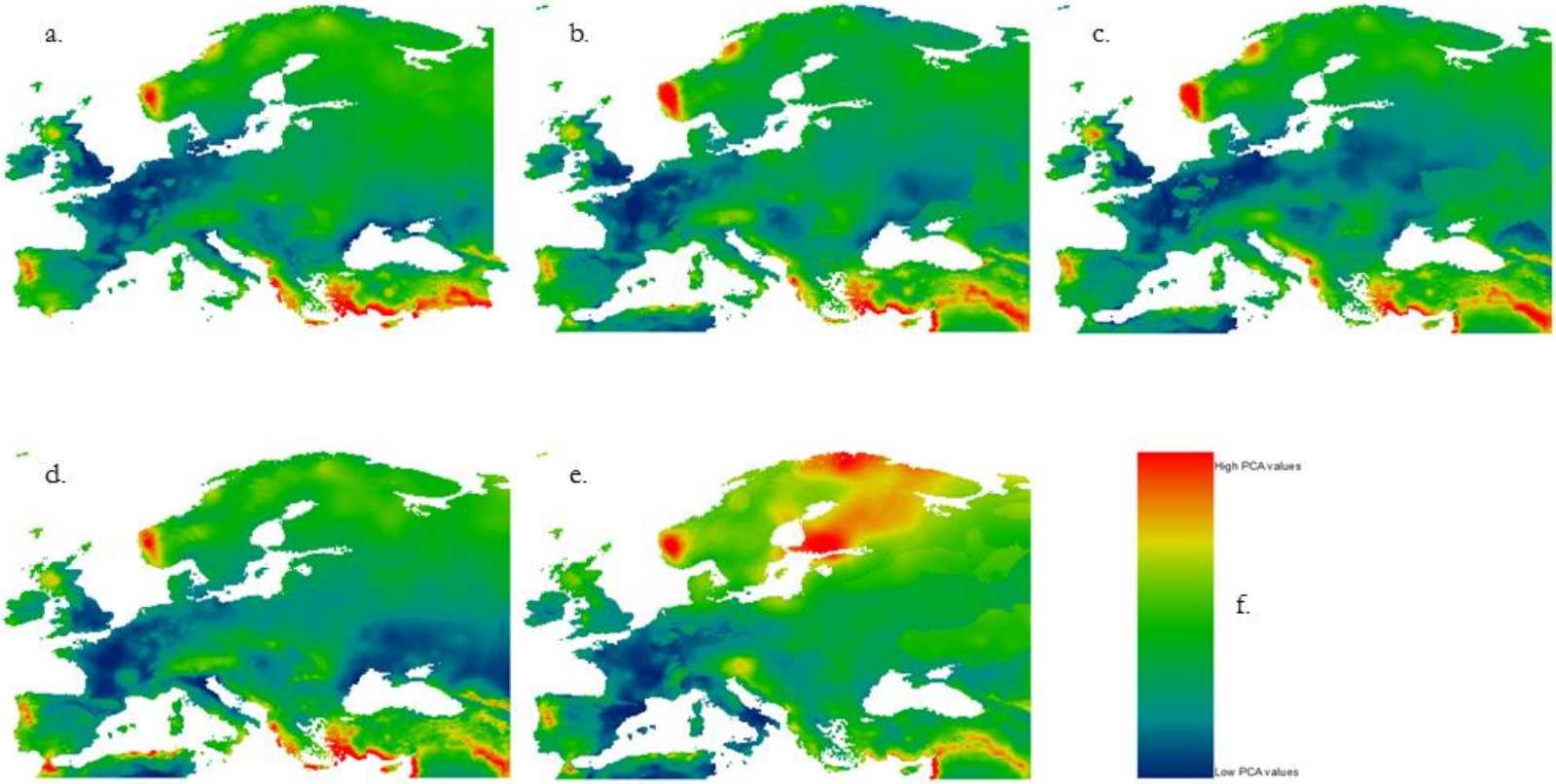
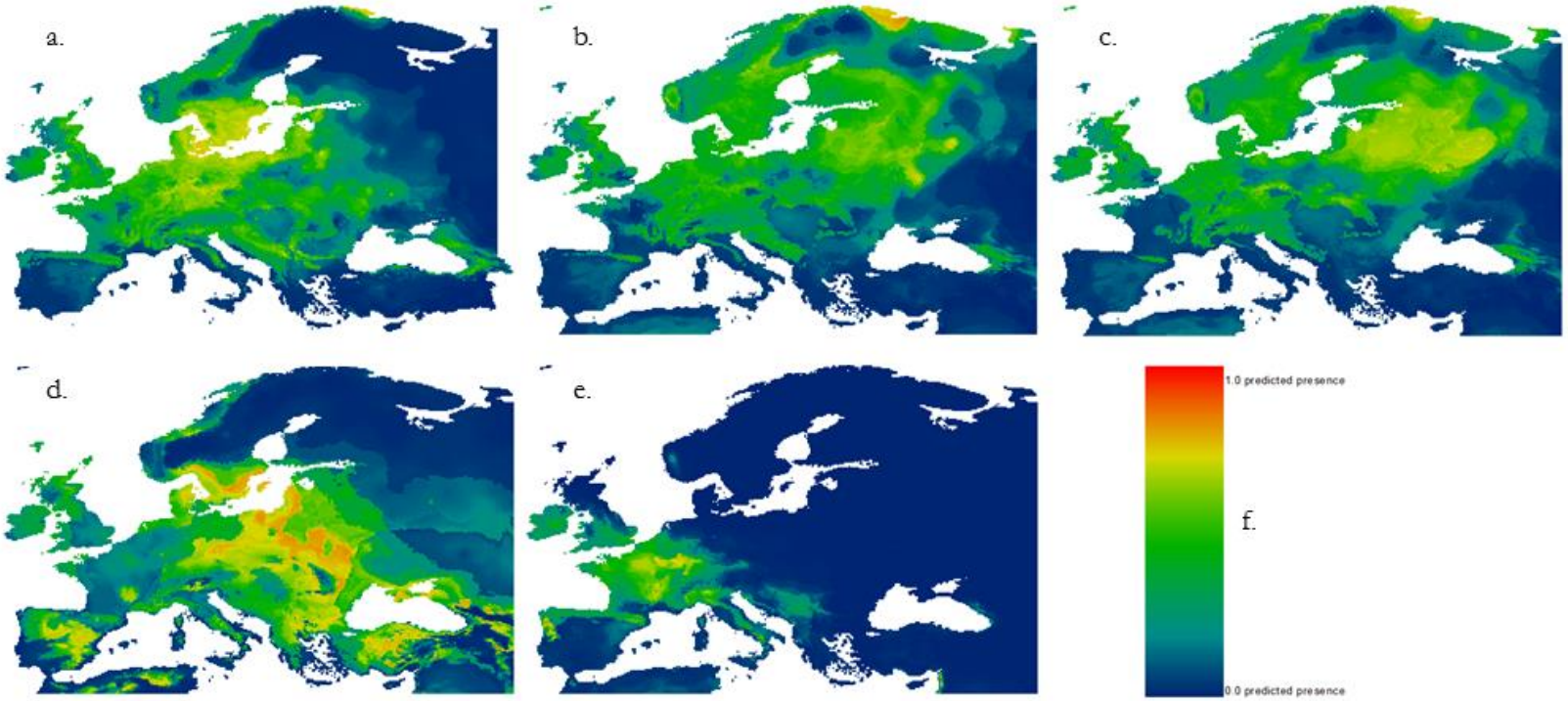


Figure 7.3: Scores for Principle Component 3 across the study area at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – arbitrary scale for the highest and lowest score for each individual projection

Figure 7.4: Predicted Presence Scores for *Cetonia aurata* at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – scale between 0 and 1 for the per-cell probability of presence, g – the importance of each PCA variable to the SDM, h – how the species responds to each PCA



g.

Variable	Percent contribution	Permutation importance
PCA1	45.5	44.0
PCA2	27.6	32.8
PCA3	27.0	23.1

h.

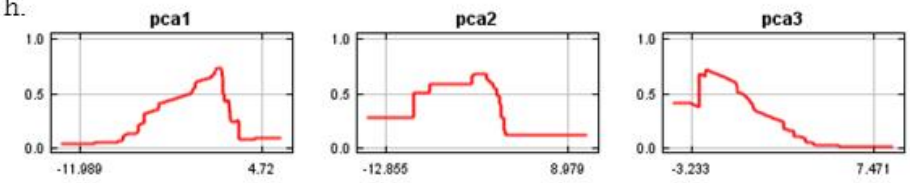
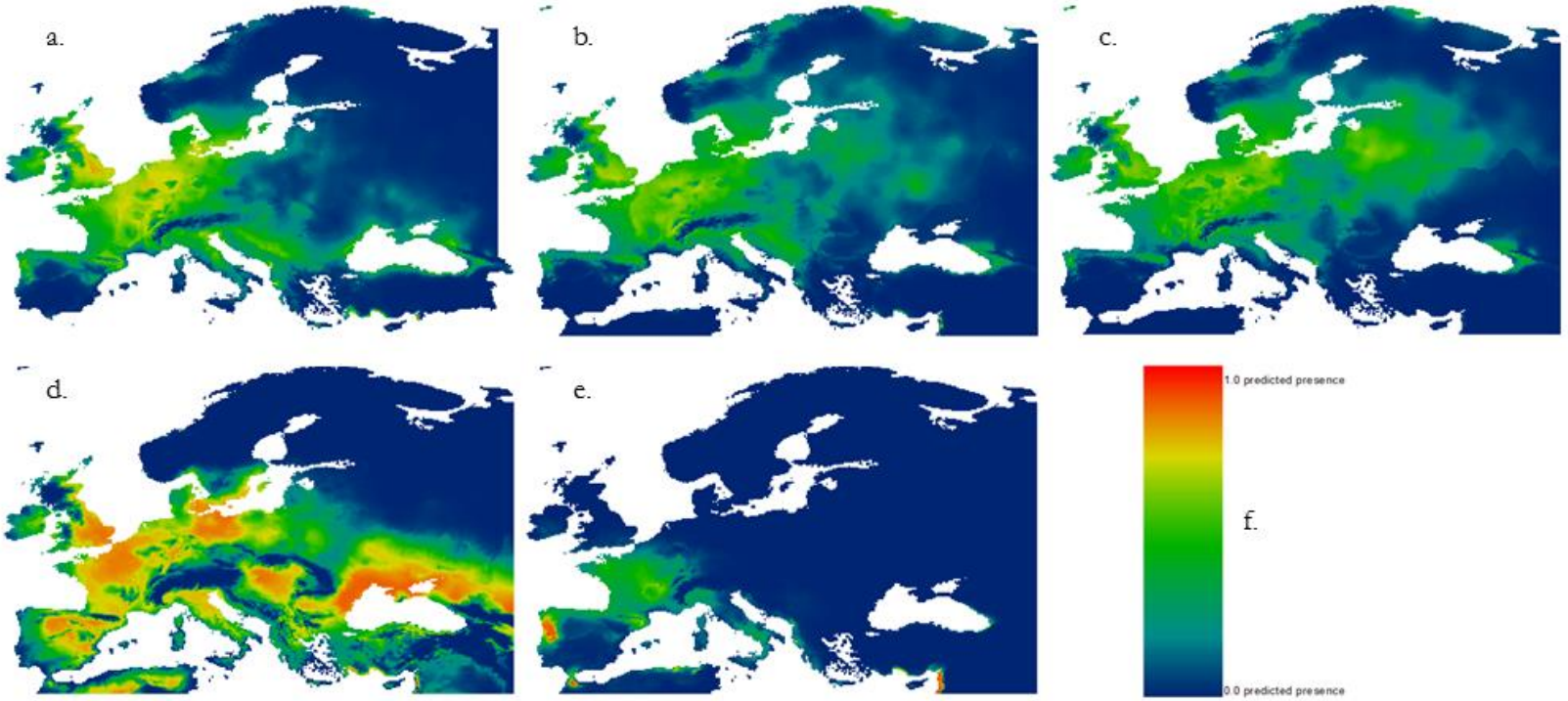


Figure 7.5: Predicted Presence Scores for *Dorcus parallelipipedus* at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – scale between 0 and 1 for the per-cell probability of presence, g – the importance of each PCA variable to the SDM, h – how the species responds to each PCA



g.

Variable	Percent contribution	Permutation importance
PCA3	69.5	44.3
PCA1	23.3	32.4
PCA2	7.2	23.4

h.

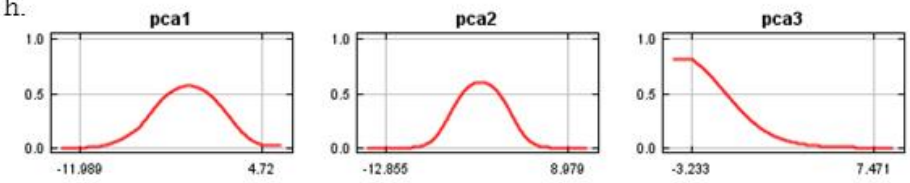
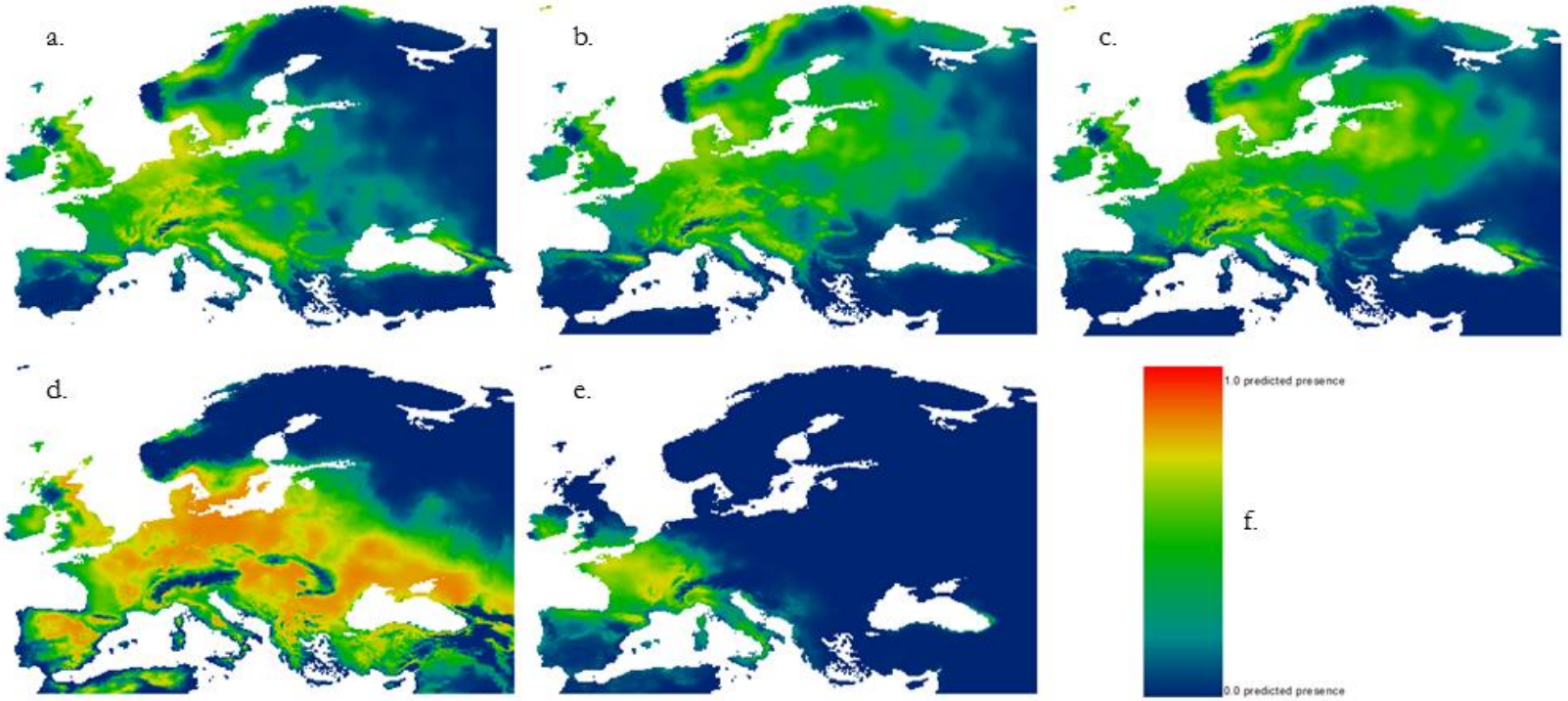


Figure 7.6: Predicted Presence Scores for *Gnorimus nobilis* at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – scale between 0 and 1 for the per-cell probability of presence, g – the importance of each PCA variable to the SDM, h – how the species responds to each PCA



g.

Variable	Percent contribution	Permutation importance
PCA1	44.8	49.8
PCA3	34.0	17.3
PCA2	21.3	32.9

h.

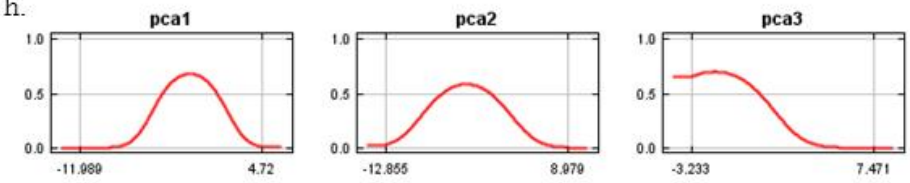
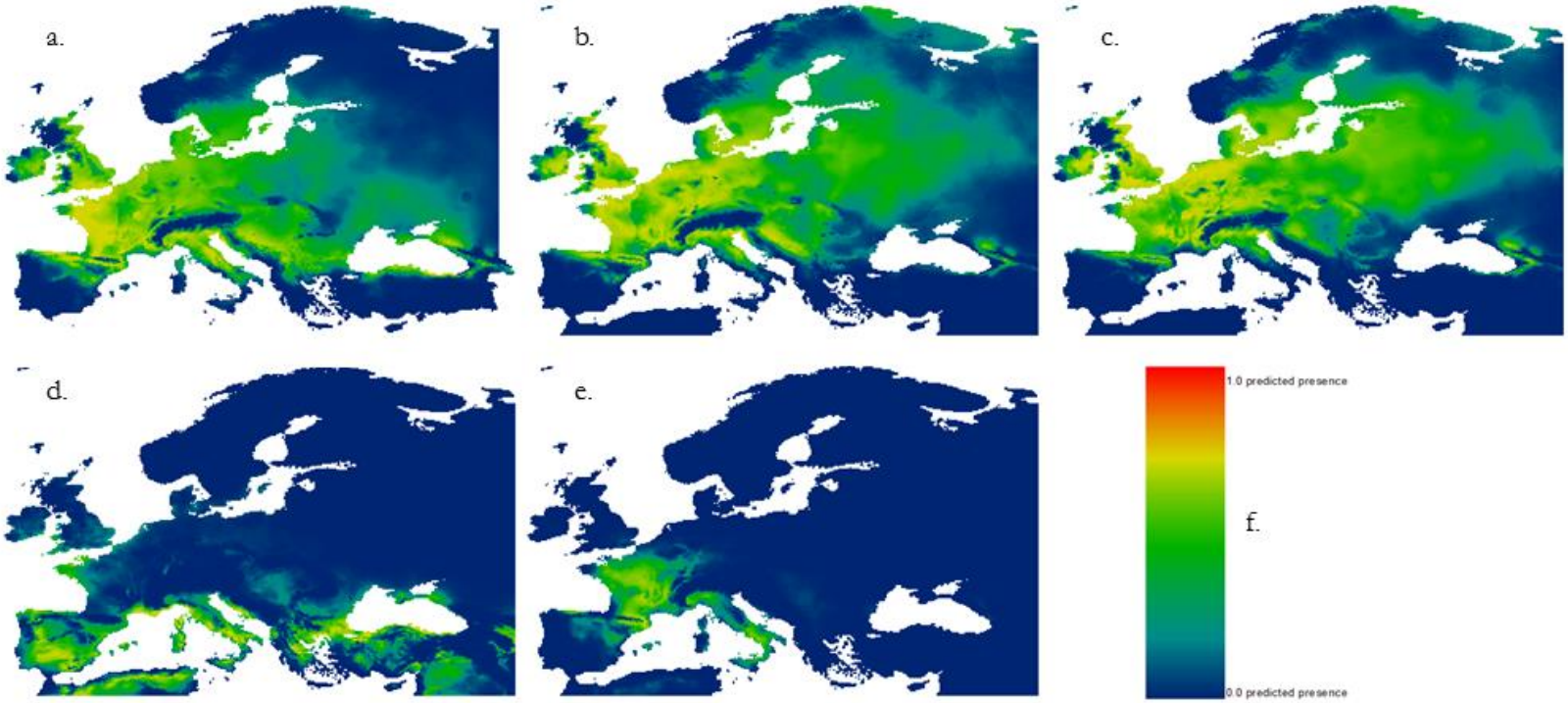


Figure 7.7: Predicted Presence Scores for *Gnorimus variabilis* at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – scale between 0 and 1 for the per-cell probability of presence, g – the importance of each PCA variable to the SDM, h – how the species responds to each PCA



g.

Variable	Percent contribution	Permutation importance
PCA3	58.8	34.7
PCA1	23.9	36.7
PCA2	17.3	28.6

h.

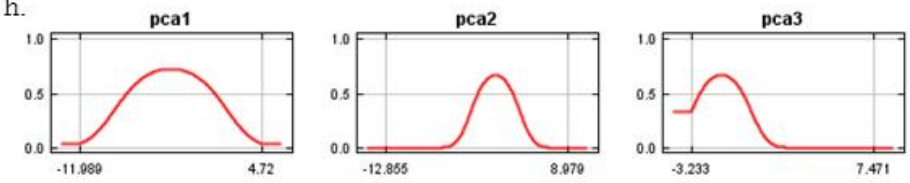
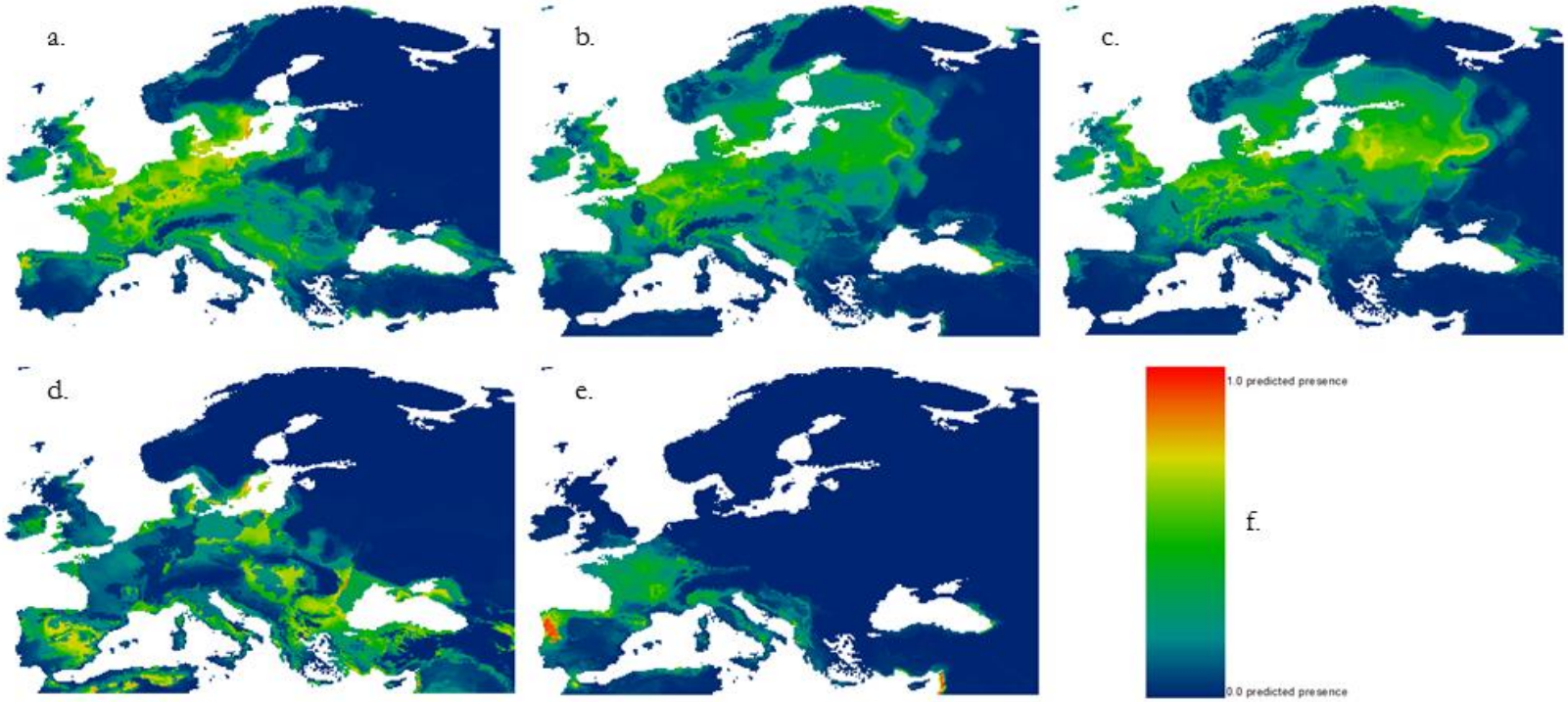


Figure 7.8: Predicted Presence Scores for *Lucanus cervus* at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – scale between 0 and 1 for the per-cell probability of presence, g – the importance of each PCA variable to the SDM, h – how the species responds to each PCA



g.

Variable	Percent contribution	Permutation importance
PCA1	43.4	59.3
PCA3	36.3	18.6
PCA2	20.4	22.1

h.

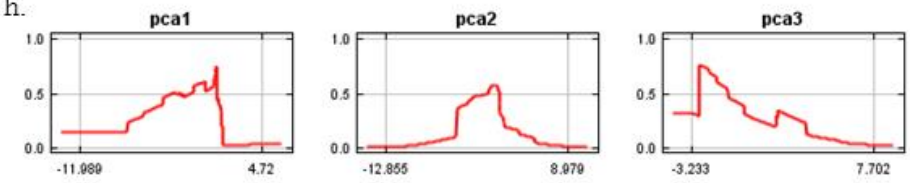
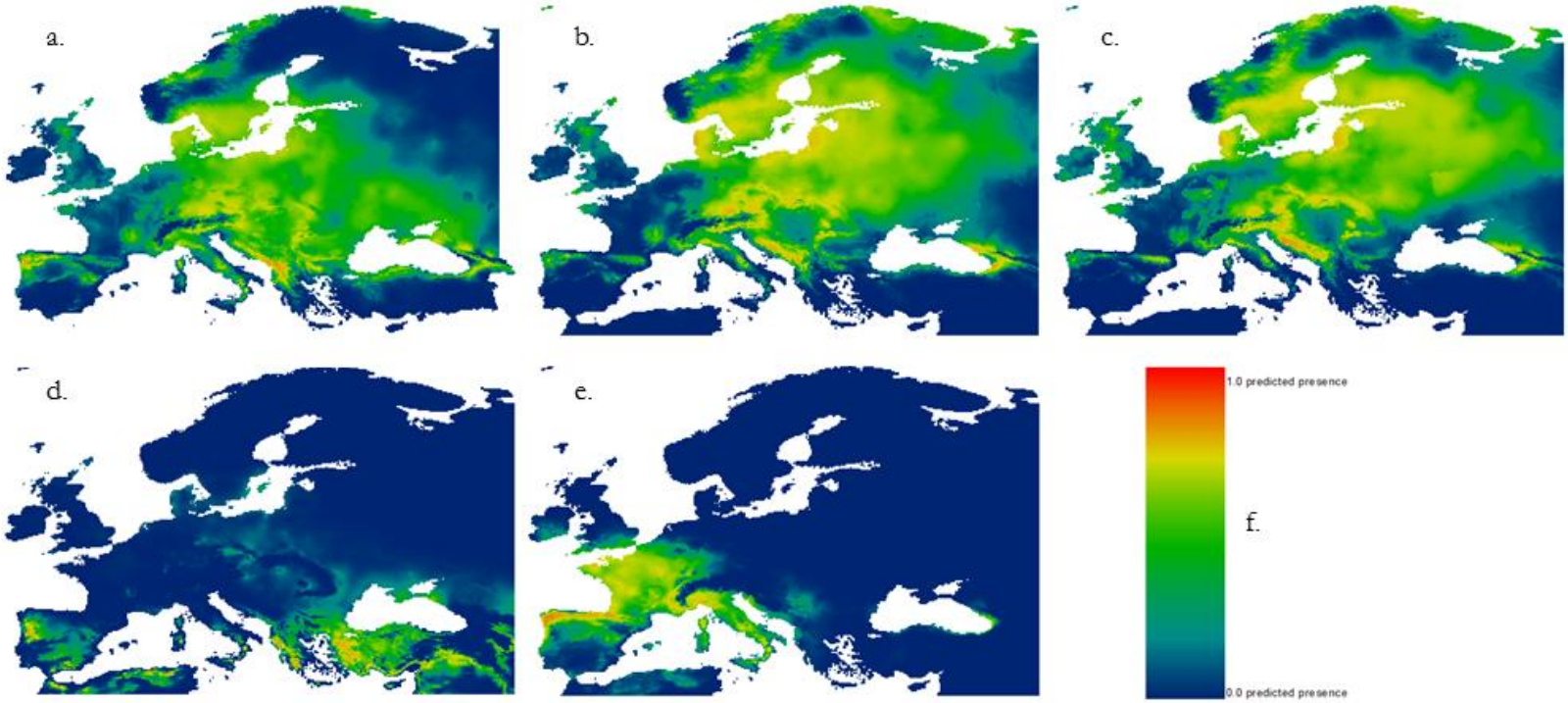


Figure 7.9: Predicted Presence Scores for *Osmoderma barnabita* at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – scale between 0 and 1 for the per-cell probability of presence, g – the importance of each PCA variable to the SDM, h – how the species responds to each PCA



g.

Variable	Percent contribution	Permutation importance
PCA1	48.4	46.5
PCA2	27.3	20.3
PCA3	24.2	33.1

h.

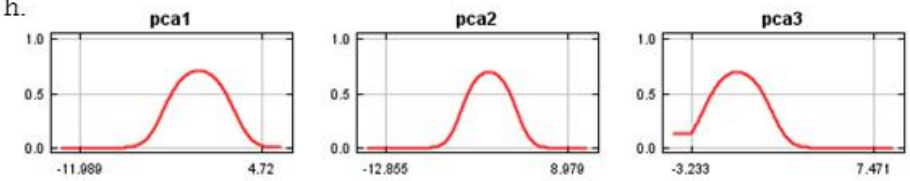
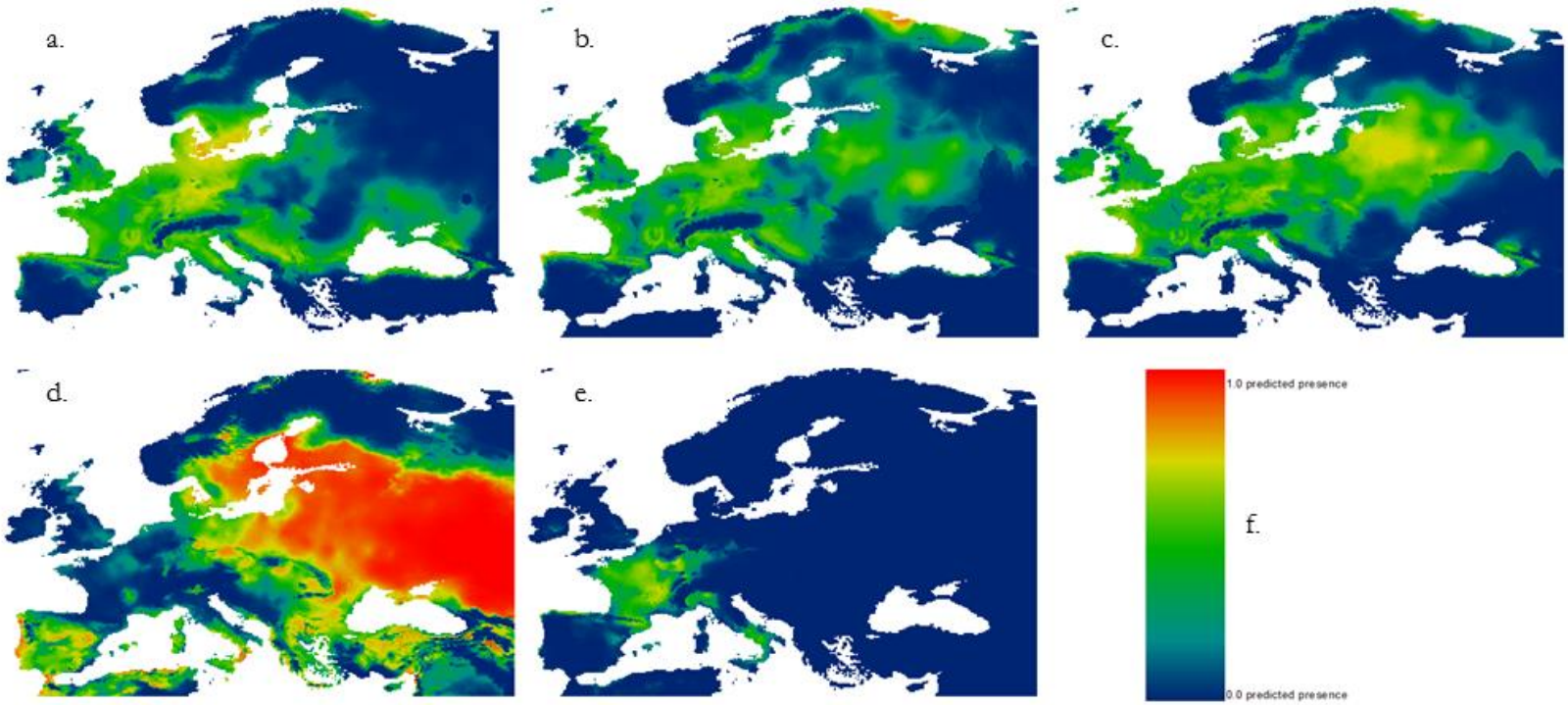


Figure 7.10: Predicted Presence Scores for *Osmoderma eremita* at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – scale between 0 and 1 for the per-cell probability of presence, g – the importance of each PCA variable to the SDM, h – how the species responds to each PCA



g.

Variable	Percent contribution	Permutation importance
PCA3	68.3	62.3
PCA1	19.5	15.8
PCA2	12.2	21.9

h.

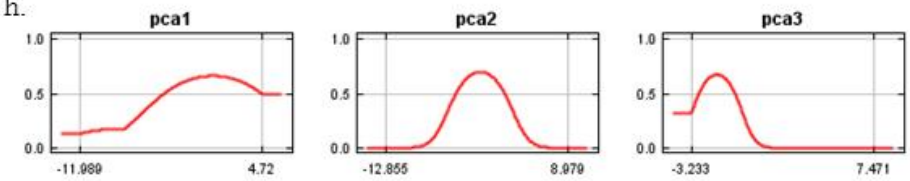
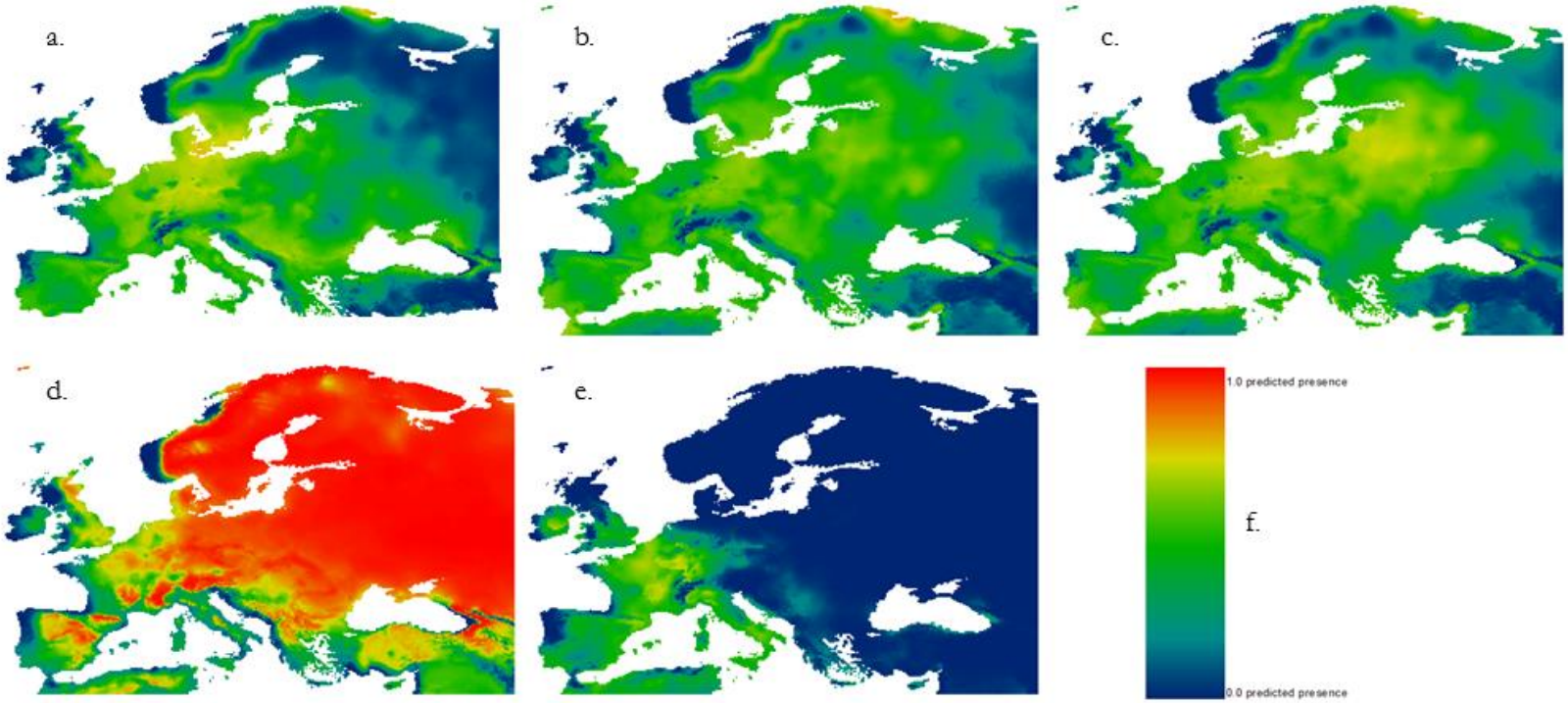


Figure 7.11: Predicted Presence Scores for *Oryctes nasicornis* at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – scale between 0 and 1 for the per-cell probability of presence, g – the importance of each PCA variable to the SDM, h – how the species responds to each PCA



g.

Variable	Percent contribution	Permutation importance
PCA1	47.6	40.2
PCA3	47.0	44.2
PCA2	5.4	15.7

h.

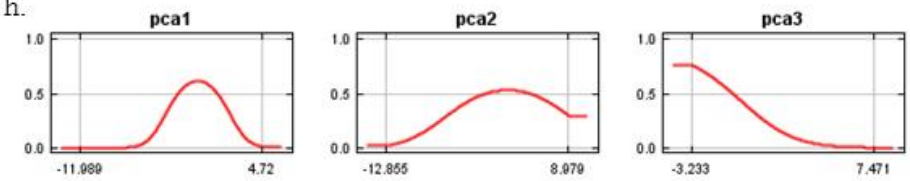
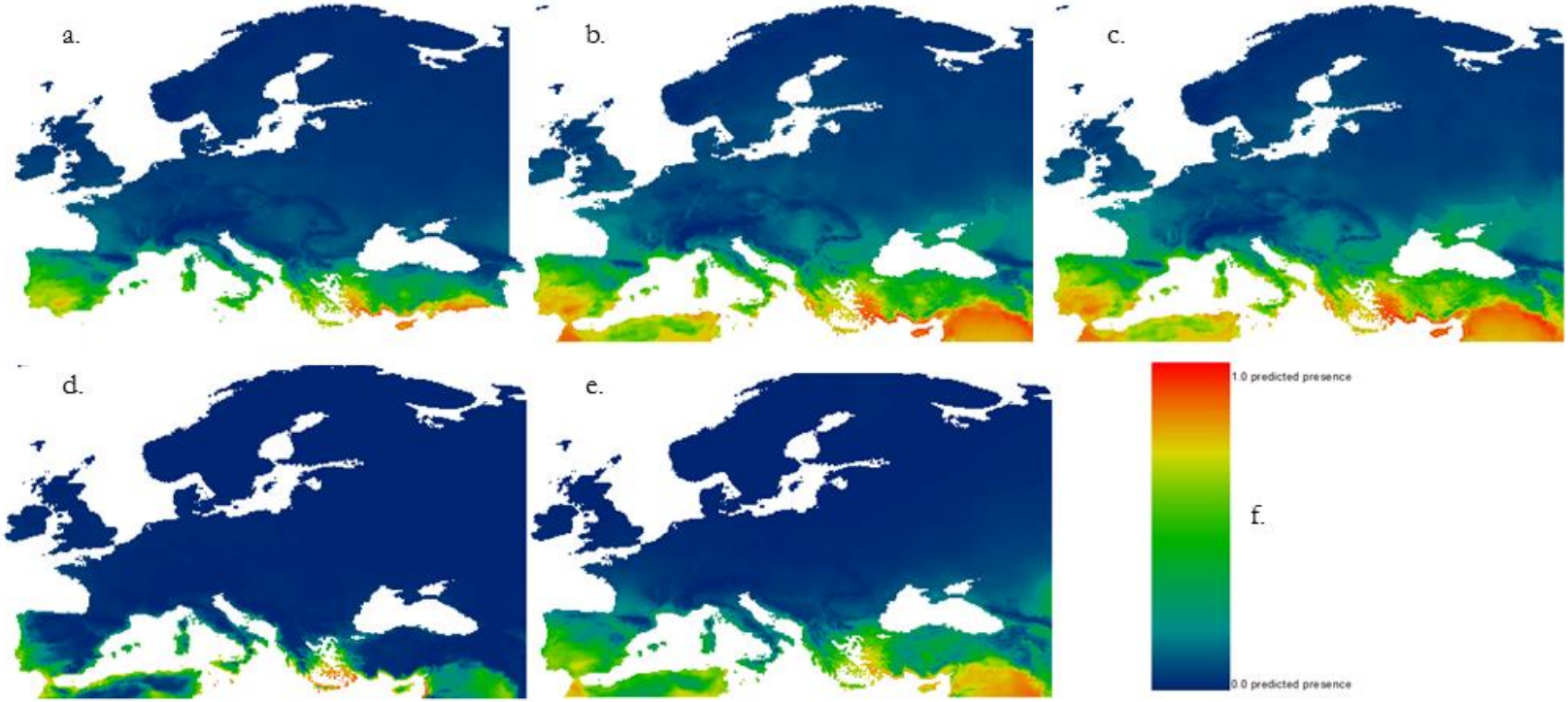


Figure 7.12: Predicted Presence Scores for *Propomacrus bimucronatus* at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – scale between 0 and 1 for the per-cell probability of presence, g – the importance of each PCA variable to the SDM, h – how the species responds to each PCA



g.

Variable	Percent contribution	Permutation importance
PCA2	70.6	91.6
PCA3	16.8	1.8
PCA1	12.6	6.6

h.

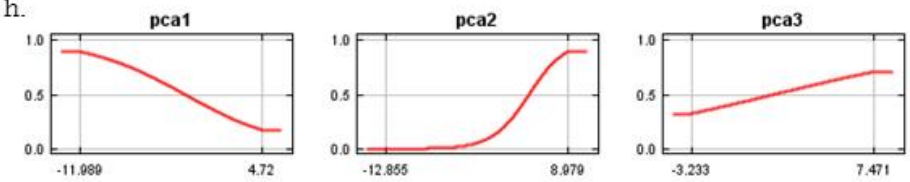
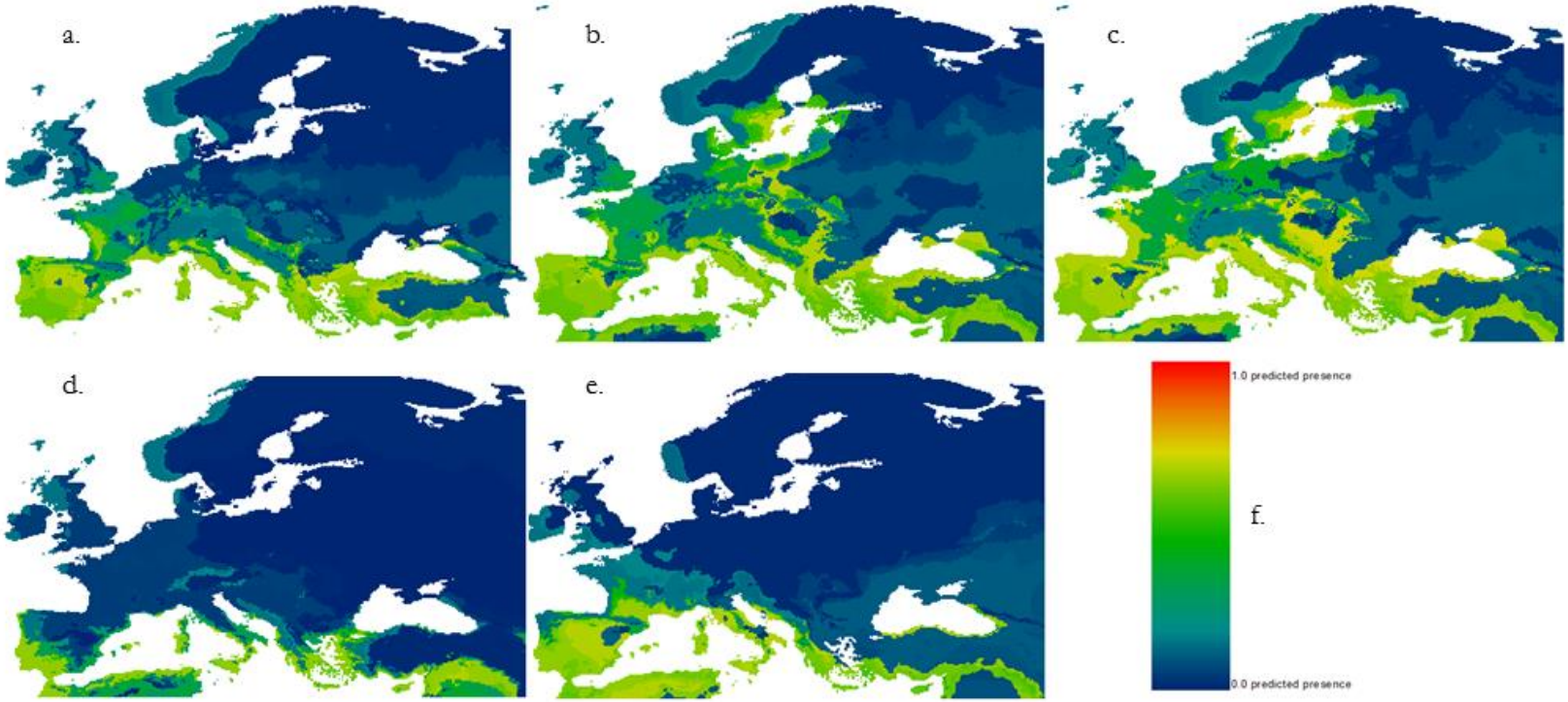


Figure 7.13: Predicted Presence Scores for *Protaetia mirifica* at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – scale between 0 and 1 for the per-cell probability of presence, g – the importance of each PCA variable to the SDM, h – how the species responds to each PCA



g.

Variable	Percent contribution	Permutation importance
PCA1	54.8	71.2
PCA2	39.0	22.4
PCA3	6.2	6.4

h.

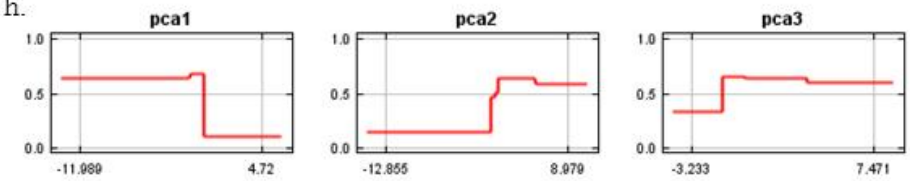
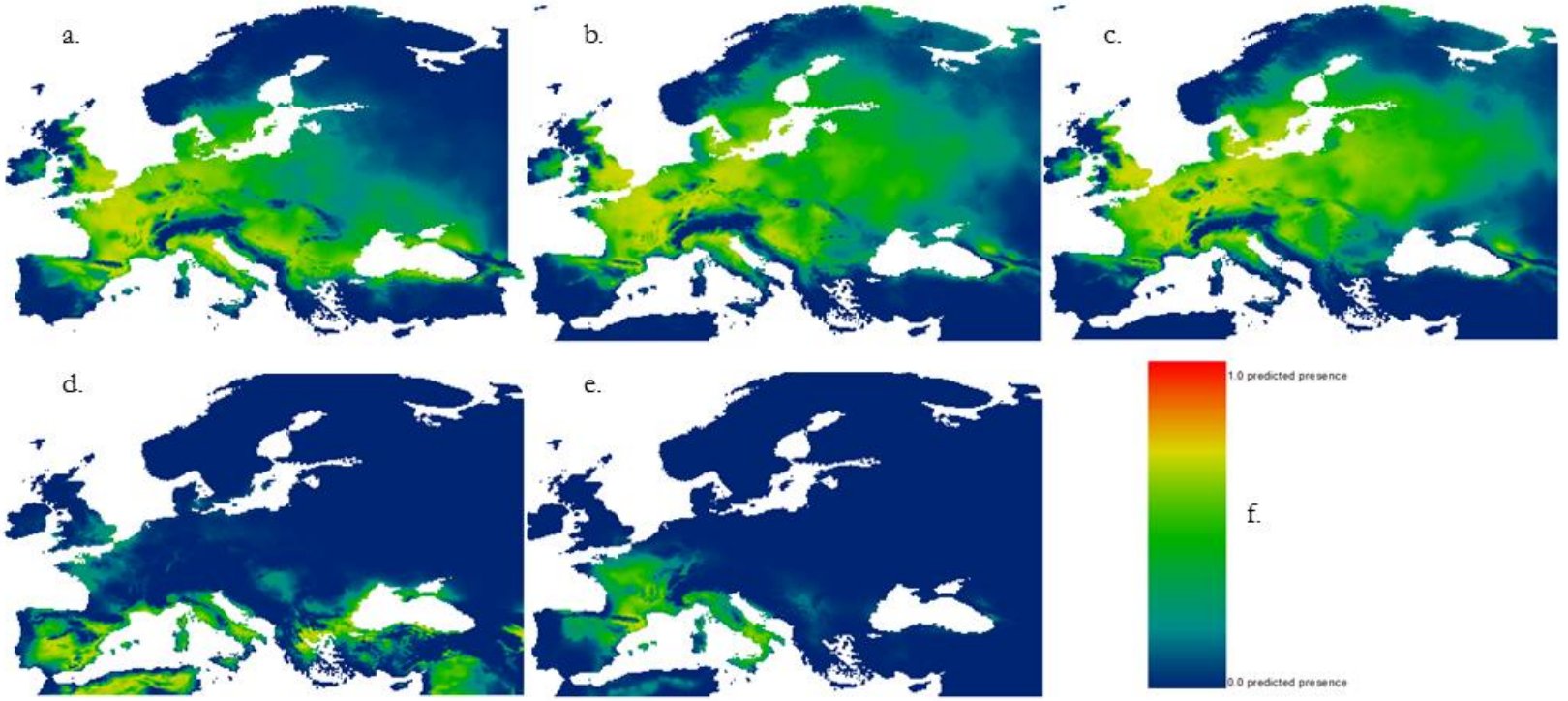


Figure 7.14: Predicted Presence Scores for *Protaetia speciosissima* at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – scale between 0 and 1 for the per-cell probability of presence, g – the importance of each PCA variable to the SDM, h – how the species responds to each PCA



g.

Variable	Percent contribution	Permutation importance
PCA3	71.5	28.2
PCA1	16.2	42.8
PCA2	12.3	29.0

h.

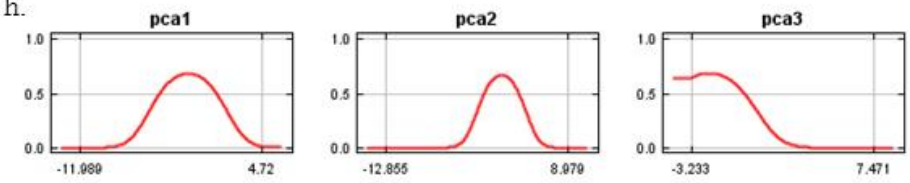
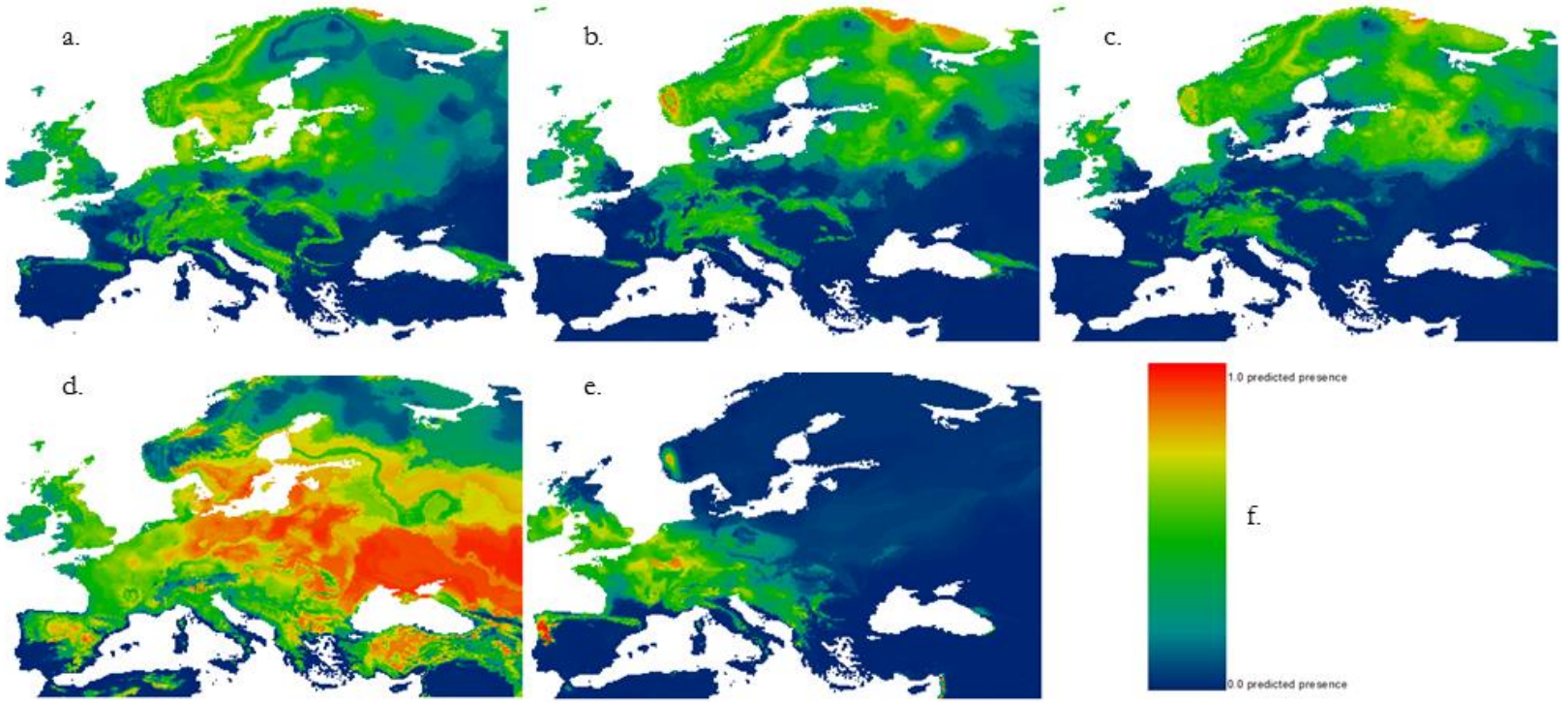


Figure 7.15: Predicted Presence Scores for *Trichius fasciatus* at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – scale between 0 and 1 for the per-cell probability of presence, g – the importance of each PCA variable to the SDM, h – how the species responds to each PCA



g.

Variable	Percent contribution	Permutation importance
PCA2	60.5	55.2
PCA1	33.3	27.7
PCA3	6.2	17.1

h.

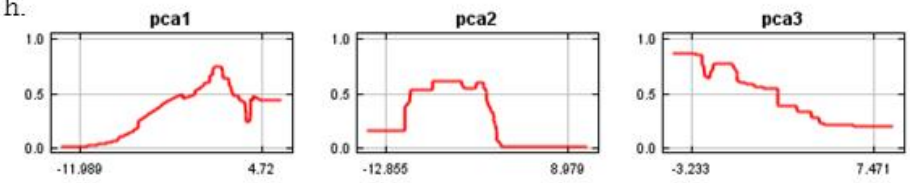
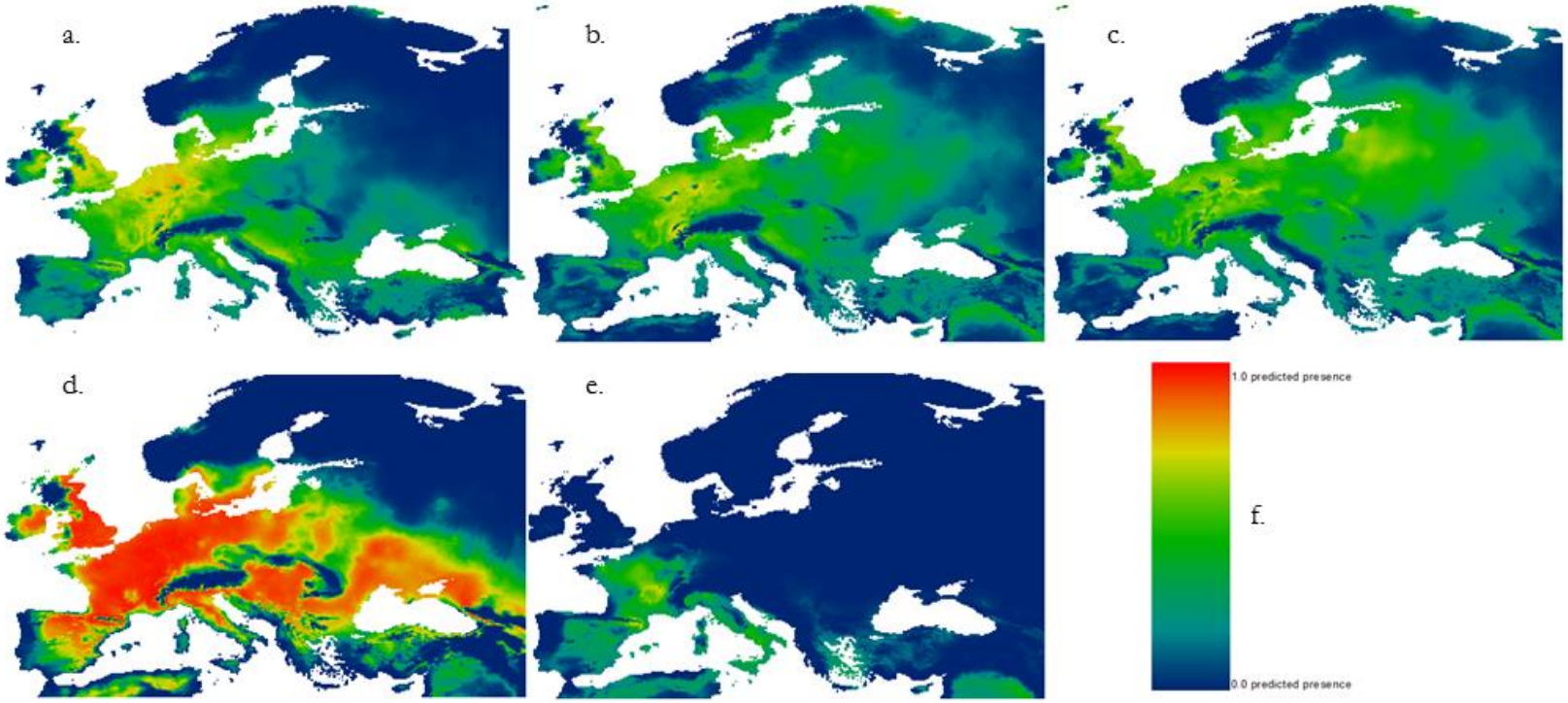


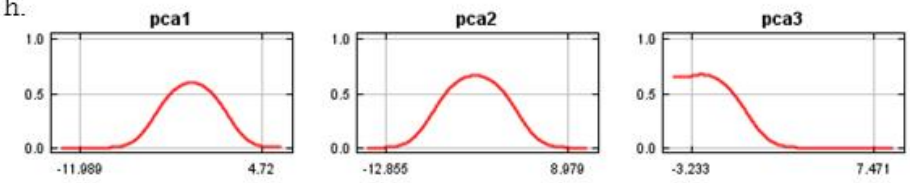
Figure 7.16: Predicted Presence Scores for *Valgus hemipterus* at different time points: a – current, b – 2050 forecast, c – 2070 forecast, d – 6kya hindcast, e – 22kya hindcast, f – scale between 0 and 1 for the per-cell probability of presence, g – the importance of each PCA variable to the SDM, h – how the species responds to each PCA



g.

Variable	Percent contribution	Permutation importance
PCA3	71.0	47.6
PCA1	21.0	29.4
PCA2	7.9	23.0

h.



Chapter 8

General Discussion

8.1 Initial Questions

At the outset of this study, a number of questions to answer were drawn up to further the knowledge of saproxylic scarab beetle population genetics and how this relates to their conservation. The success of this work to answer said questions can now be analysed:

- Which standard patterns of post-glacial recolonisation do *Lucanus cervus*, *Trichius* spp. and *Gnorimus* spp. adhere to, and can any differences be attributed to their biology?

The stag beetle (*Lucanus cervus*) shows little population structure and vicariance across Europe, with COI haplotypes belonging to a star-shaped haplogroup-phylogeny, as is typical of populations that have experienced recent expansions. However, microsatellite genotyping shows that genetic diversity within Spain (a putative glacial refugium for this species) is far higher than diversity within the UK, which is genetically depauperate in comparison.

Both *Gnorimus nobilis* and *G. variabilis* have phylogeographic patterns similar to the ecologically related *Osmoderma* species complex, with a distinct divide between western and eastern European populations, which is more pronounced in *G. nobilis*. Though the sampling density was too low to be able to develop strong hypotheses about the post-glacial expansion of *G. variabilis*, *G. nobilis* appears to have expanded from a northern refugium around eastern France, Switzerland, Austria and southern Germany, leading to low diversity in mitochondrial DNA throughout Europe, and a star-shaped haplogroup-phylogeny similar to that in *L. cervus*. Locally restricted divergent mtDNA haplotypes in northern Spain, southern Italy and possibly the Balkans and species distribution modelling (SDM) suggest evidence for glacial refugia which did not contribute to post-glacial expansions of *G. nobilis*.

In contrast, the *Trichius fasciatus* COI haplogroup shows little population structuring, with populations separated by over 6,000km sharing the same closely related haplotypes (similar to, but more extreme than, the pattern in *L. cervus*). SDM suggests that this species may have persisted in northern refugia throughout western and central Europe during the last glacial maximum, though as with *Gnorimus* the distribution of *Trichius* is constrained by the availability of suitable host trees with rot-holes. The difference in distribution between *Trichius* and *Gnorimus* during the LGM and a different post-glacial recolonisation pattern

may in part be due to differences in thermal tolerance between the genera, and differences in host-species: *Trichius* utilises birch as a primary host, which is a genus more tolerant of extreme environmental conditions than the oaks, chestnut and beech that *Gnorimus* usually utilise.

- What is the status of the three proposed subspecies of *Gnorimus nobilis*, and are they reliably identifiable by genetic and morphological means?

Only *Gnorimus nobilis nobilis* and *G. nobilis bolshakovi* were supported as distinct clades within *G. nobilis*, from mtDNA and morphology. The suggestion in the literature that *G. nobilis bolshakovi* could be identified based on the absence of a mesotibial callus/spoon-shaped protrusion in males was disputed by the results of the present study as most males throughout eastern Europe possessed such mesotibia. *G. nobilis bolshakovi* bodies were considerably narrower than *G. nobilis nobilis*, but did not significantly differ in any other morphological variable tested. The status of the other subspecies, *G. nobilis macedonicus* has not been clearly resolved. Of the three individuals used in the study from Macedonia, one had a COI haplotype belonging to the Eastern (*G. nobilis bolshakovi*) clade and the other two possessed haplotypes belonging to the most common Western haplogroup. As two specimens were damaged, only one could be used in the geometric-morphometric analysis, which grouped more closely with *G. nobilis bolshakovi* in 'morphospace' than with *G. nobilis nobilis*. Thus the status of this subspecies remains ambiguous.

- How closely related are the three *Trichius* species, and are their morphological differences robust enough to allow for identification in non-standard conditions?

Relationships between the three species are complicated. Though three distinctive COI haplogroups were confirmed (on GenBank, and recovered in the study) at a level of divergence that could indicate distinct species, these haplogroups do not correlate clearly with the defined morphotypes. The morphotypes themselves (as defined from the literature) are distinctive, but one well sampled population possessed males identified as both *T. fasciatus* and *T. gallicus* where all individuals possessed COI haplotypes belonging to the assumed *T. fasciatus* clade, and another population where all individuals were morphologically identified as *T. sexualis* but possessed a mixture of *T. fasciatus* and *T. sexualis* haplotypes. Morphological, mtDNA and nuclear DNA (*Wingless*) variation suggest that *T. fasciatus* and *T. sexualis* are, or were at some point in the recent past, well-defined

species, but the status of *T. gallicus* is still unclear. It may be a hybrid form between the other two species (its morphology is intermediate between *T. fasciatus* and *T. sexualis*, *Wingless* shows no population differentiation between it and *T. fasciatus*), or may once have been a 'good' species that has been genetically swamped by the other species via introgressive hybridisation.

- Do markers under selection show similar patterns to neutral markers in analysing *Gnorimus* phylogeography? Does the "southern richness, northern purity" paradigm still hold for genes under selection?

Both Heat Shock Protein 70 and *Orco* showed strong evidence of selection in every sub-*Gnorimus* haplogroup, where HSP70 indicated positive selection and *Orco* indicated purifying selection was taking place on these genes within *Gnorimus*. Western *G. variabilis* showed uniformly strong levels of population differentiation from other populations (other than Eastern *G. variabilis*), and there was some differentiation between different populations of *G. nobilis*. As between-group divergences in *G. nobilis* are relatively low in Western, Southern and Central Europe, it is unlikely that there has been enough evolutionary time for population divergence to have occurred via drift. No evidence was found for the "southern richness, northern purity" paradigm in either HSP70 or *Orco* throughout Europe.

- Can species distribution modelling be used in tandem with multi-locus phylogeography to provide the best possible prediction of glacial refugia, or are there too many dataset biases? Is there a general "saproxylic post-glacial recolonisation pattern", or are different species predicted to have utilised different refugia?

Both SDM and population genetics predict the existence of a non-southern European (i.e. central Europe) glacial refugium for *G. nobilis* during the last glacial maximum, with subsequent post-glacial expansion from France, Switzerland, Austria and Germany. In addition, SDM also predicted refugia in Spain, Italy and the Balkans, also hinting at a small refugium in the Caucasus, all of which were supported by the presence of localised differentiated haplotypes restricted to these regions. In addition, a more restricted northern refugium is predicted for *G. variabilis*, corresponding with the evidence presented by Trizzino et al. (2013) indicating that *G. variabilis* is found at lower elevations than *G. nobilis*. These patterns contrast to some extent with the data from *Trichius* spp. which do not show evidence of being confined to southern refugia. In general however, multiple species of

saproxylic scarab showed similar SDM estimated refugial patterns to *G. nobilis*, but species confined to Mediterranean regions show radically different patterns consistent with adaptation to a warmer, drier environment. *Lucanus cervus* is predicted to have had a large refugium in Iberia, which matches with data showing high diversity in microsatellite alleles in this region, compared to the genetically depauperate northern populations.

- How strongly do previous results from other researchers match with SDM predictions about glacial refugia?

Evidence from *Cetonia aurata* (from COI) indicates an expansion from a wide northern refugium from France through to Germany, and another population predicted in the Balkans. Regions with divergent haplotypes, but which didn't contribute haplotypes to the primary expansion existed in Italy, Iberia and Lebanon, were predicted by the SDM approach. An almost identical pattern was detected in *Lucanus cervus*, with isolated refugia in the Caucasus and Lebanon.

- Can areas predicted to become unsuitable for rare saproxylic scarabs be identified and can new climatically suitable areas be identified with climate change predictions?

Many saproxylics are predicted to expand into the northern European plain (Belarus, Lithuania, Latvia, Estonia and North West Russia) and further north throughout Britain, Norway and Sweden. However, increasingly extreme climates in the south are predicted to result in large areas of the Mediterranean region, particularly in the lowlands, becoming unsuitable for many species. However, Mediterranean specialists are predicted not to find the more extreme climate unsuitable, but are also expected to find additional suitable areas of habitat further north and east from their current distributions. No major differences between generalist saproxylics and rot-hole specialists were found, but the expansions of all species assume that enough suitable habitat is found in new climatically suitable regions for these species (specialists in particular).

- How robustly do the three main methods utilised (SDMs, morphological modelling, and phylogeography) support each other, and where are the weakest areas of support?

SDM approaches only support a small refugium in the southern Caucasus for *Gnorimus nobilis*. Additionally, areas where *G. nobilis bolshakovi* is currently found in Ukraine and

Russia are predicted via SDM as being unsuitable for the species as a whole. This may indicate that this putative subspecies is suited to a more continental climate than *G. nobilis nobilis*, and therefore may respond differently to climate change. This suggests that SDM is sensitive to biological differences between differentiated groups within species, which need to be taken into account when results are interpreted. However, SDM failed to report a strong climatic signal in the distribution of *Trichius fasciatus*, which may be partially due to the complicated taxonomic relationships between the European species. In addition, morphological modelling and phylogeographic / phylogenetic analysis in *Trichius* showed the genus to have a complex pattern of unreported morphological and genetic incongruence, thus more work with more neutral nuclear markers and larger sample sizes should be employed to better analyse the taxonomy of this genus.

Future work should investigate a number of areas:

- Testing other species of saproxylic scarab to see if the predicted patterns of postglacial recolonisation hold true, and to investigate if there are multiple evolutionarily significant units present throughout their range
- Utilising neutral nuclear markers (microsatellites, single nucleotide polymorphisms) to improve the resolution of post-glacial recolonisation in *Gnorimus* and *Trichius*, and expanding the study range of *Lucanus cervus*
- Investigating the effects that rot-hole dwelling saproxylics have on their environment via environmental DNA sequencing, focussing on Cetoniids to analyse the hypothesis that they are secondary ecosystem engineers
- Adding trophic layers (such as host trees and fungi) into the SDM, as well as early-stage habitat modifying beetles to better understand the predicted refugial regions for saproxylics, and their combined responses to climate change

Though populations of saproxylic beetles are generally low in genetic diversity away from southern refugia, effective, active management of their populations may be needed to allow the species to colonise new habitat as regions of suitable climate begin to open up under predicted climate change scenarios. As mature woodland is generally fragmented in distribution across Europe, active translocation of populations may be needed to encourage species to colonise new habitat. Such translocations should come from the nearest population

adapted to a similar climate, confirmed by both species distribution modelling and analysis of non-neutral genetic markers. In addition, as southern regions are predicted to become increasingly unsuitable for some saproxylics, populations in such areas should be genetically monitored to assess if populations are adapting to the changing conditions, or if localised extinctions are likely.

8.2 Conclusions

As a multidisciplinary approach focussing on combining population genetics with morphological and species distribution modelling analyses, its success at advancing the knowledge surrounding conservation management of saproxylic beetles can be judged not only by the long-term use of its findings in beetle conservation, but also by the extent to which each discipline enhances and informs each other. In this regard, combining these three disciplines to analyse *Gnorimus nobilis* phylogeography in particular has been a success: strong divergences between subspecies and glacial refugia have been identified, with similar refugia predicted by both population genetics and species distribution modelling. Future research on species included in the thesis, and those evolutionarily and ecologically related to them, should find useful information throughout the thesis, either discussing the trials and tribulations with working on scarab genetics, or utilising the predictions discussed in the modelling chapter. Reviews of the utility both of novel genetic markers and DNA extraction techniques will also allow for future researchers to better utilise non-invasive DNA solutions as well as museum samples. In particular, the use of non-invasive by-products should be encouraged where possible, as long as some caveats are maintained (use of short targeted regions in particular). With increasing understanding of the role saproxylics play in their environment, and the high risk of extinction some species face, the work presented here should encourage researchers to identify genetic conservation units of saproxylics using multi-gene approaches, to use cross-discipline methods to enhance conclusions, and to work on multiple species to better benefit future conservation studies and management.

References

Abaza, S. M. (2015) Heat shock proteins: Part III. Arthropods. *Parasitologists United Journal* 8 95-100

Abbadie, L., Lepage, M. & Le Roux, X. (1992). Soil fauna at the forest-savanna boundary: role of termite mounds in nutrient cycling. *Nature and dynamics of forest-savanna boundaries*, 473-484.

Aberlenc, H. P. (2006). La grande Cétoine bleue, Graal du patrimoine naturel de Païolive. *La Viste*, 19, 42-48.

Adams, D. C., Rohlf, F. J., & Slice, D. E. (2004). Geometric morphometrics: ten years of progress following the 'revolution'. *Italian Journal of Zoology*, *71*(1), 5-16.

Adams, D. C., Rohlf, F. J., & Slice, D. E. (2013). A field comes of age: geometric morphometrics in the 21st century. *Hystrix*, *24*(1), 7-14.

Ahrens, D., Fabrizi, S., Šipek, P., & Lago, P. K. (2013). Integrative analysis of DNA phylogeography and morphology of the European rose chafer (*Cetonia aurata*) to infer species taxonomy and patterns of postglacial colonisation in Europe. *Molecular phylogenetics and evolution*,69(1), 83-94.

Aitken, S. N., Yeaman, S., Holliday, J. A., Wang, T., & Curtis-McLane, S. (2008). Adaptation, migration or extirpation: climate change outcomes for tree populations. *Evolutionary Applications*,*1*(1), 95-111.

Alcaide, M., Rodríguez, A., & Negro, J. J. (2011). Sampling strategies for accurate computational inferences of gametic phase across highly polymorphic major histocompatibility complex loci. *BMC research notes*,*4*(1), 151.

Alekseev, V. I., & Nikitsky, N. B. (2008). Rare and new for the fauna of the Baltic states beetles (Coleoptera) from the Kaliningrad region. *Acta Zoologica Lituanica*, 18(4), 261-266.

Alexander, K. N. A. (2002). The invertebrates of living and decaying timber in Britain and Ireland - a provisional annotated checklist. English Nature Research Report 467, Peterborough (142pp)

Alexander, K. N. A. (2003 a.) The British saproxylic invertebrate fauna. Proceedings of the second pan-European conference on Saproxylic Beetles. People's Trust for Endangered Species.

Alexander, K. N. A. (2003 b.). Changing distributions of Cantharidae and Buprestidae within Great Britain (Coleoptera). In *Proceedings of the 13th International Colloquium of the European Invertebrate Survey. European Invertebrate Survey, Leiden* (pp. 87-91).

Alexander, K. N. A. (2008 a.) Tree biology and saproxylic Coleoptera: issues of definitions and conservation language. *Revue d'Ecologie* 63 1-5

Alexander, K. N. A. (2008 b.) The spatial importance of traditional orchards for invertebrate conservation, with a case study of the BAP priority species the noble chafer Gnorimus nobilis. 12-18 in: Orchards and Groves: Their History, Ecology, Culture and Archaeology (ed.: Rotherham, I. D.)

Alexander, K. N. A., Buche, B., Dodelin, B. & Schlaghamersky, J. 2010. *Osmoderma barnabita*. The IUCN Red List of Threatened Species 2010: e.T157901A5169119.

Alexander K.N.A., Mannerkoski, I., Pettersson, R., Mason, F. & Schmidl, J. 2010. *Dorcus parallelipipedus*. The IUCN Red List of Threatened Species 2010: e.T157667A5120866.

Algar, A. C., Mahler, D. L., Glor, R. E., & Losos, J. B. (2013). Niche incumbency, dispersal limitation and climate shape geographical distributions in a species-rich island adaptive radiation. *Global Ecology and Biogeography*, 22(4), 391-402.

Alibert, P., Moureau, B., Dommergues, J. L., & David, B. (2001). Differentiation at a microgeographical scale within two species of ground beetle, *Carabus auronitens* and *C. nemoralis* (Coleoptera, Carabidae): a geometrical morphometric approach. *Zoologica Scripta*, *30*(4), 299-311.

Allard, V. (1985) Goliathini 2. Les Col6opteres du Monde. Vol. 6. Sciences Nat, Compiegne.

Allard, V. (1986) Goliathini 3. Les Coleopteres du Monde. Vol. 7. Sciences Nat, Compiegne.

Allard, V. (1991) Goliathini 4. Les Col6opteres du Monde. Vol. 11. Sciences Nat, Compiegne.

Allendorf, F. W. (1986). Genetic drift and the loss of alleles versus heterozygosity. *Zoo biology*, *5*(2), 181-190.

Allendorf, F. W., & Luikart, G. (2007). Units of conservation. *Conservation and the genetics of populations*, 380-420.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology*, *215*(3), 403-410.

Amaral, D. T., Mitani, Y., Oliveira, G., Ohmiya, Y., & Viviani, V. R. (2016). Revisiting Coleoptera a+ T-rich region: structural conservation, phylogenetic and phylogeographic approaches in mitochondrial control region of bioluminescent Elateridae species (Coleoptera). *Mitochondrial DNA Part A*, 1-10.

Anderson, S. J., Conrad, K. F., Gillman, M. P., Woiwod, I. P. & Freeland, J. R. (2008) Phenotypic changes and reduced genetic diversity have accompanied the rapid decline of the garden tiger moth (*Arctia caja*) in the U.K. *Ecological Entomology* 33 638-645

Andersen, K. K., Svensson, A., Johnsen, S. J., Rasmussen, S. O., Bigler, M., Röthlisberger, R., et al. & Vinther, B. M. (2006). The Greenland ice core chronology 2005, 15–42ka. Part 1: Constructing the time scale. *Quaternary Science Reviews*,*25*(23), 3246-3257.

Ariana, S., Kastriot, M., & Neira, M. (2011). Some endangered Coleoptera species of Albania. *Natura Montenegrina*, 10(2), 93-100.

Arinton, M. (2011) Data concerning the diversity of scarabeoids (Insecta, Coleoptera, Scarabaeoidea) from Slănic Moldova, Bacău County. *Actual problems of protection and sustainable use of the animal world diversity*, 81.

Arnone, M., Cascio, P. L., & Grita, F. (2009). Gli scarabeoidei dell'isola di nisyros (dodecaneso, grecia) (Coleoptera, Scarabaeoidea). *Fragmenta Entomologica*, 41(1), 113-122.

Audisio, P., Brustel, H., Carpaneto, G. M., Coletti, G., Mancini, E., Trizzino, M., Antonini, G. & De Biase, A. (2009). Data on molecular taxonomy and genetic diversification of the European Hermit beetles, a species complex of endangered insects (Coleoptera: Scarabaeidae, Cetoniinae, *Osmoderma). Journal of Zoological Systematics and Evolutionary Research*, 47(1), 88-95.

Avgin, S. S., Dertli, İ., & Barševskis, A. (2014). A review of Turkish saproxylic beetles from the European Red List. In Annales de la Société entomologique de France (NS) (Vol. 50, No. 1, pp. 13-50). Taylor & Francis.

Avise, J. C. (1995). Mitochondrial DNA polymorphism and a connection between genetics and demography of relevance to conservation. *Conservation Biology*, *9*(3), 686-690.

Babik, W., Branicki, W., Crnobrnja-Isailović, J., Cogălniceanu, D., Sas, I., Olgun, K., Poyarkov, N. A., Garcia-París, M. & Arntzen, J. W. (2005). Phylogeography of two European newt species—discordance between mtDNA and morphology. *Molecular Ecology*, *14*(8), 2475-2491.

Bai, M., Beutel, R. G., Liu, W. G., Li, S., Zhang, M. N., Lu, Y. Y., Song, K., Ren, D. & Yang, X. K. (2014). Description of a new species of Glaresidae (Coleoptera: Scarabaeoidea) from the Jehol Biota of China with a geometric morphometric evaluation. *Arthropod Systematics & Phylogeny*,*72*(3), 223-236.

Baldwin, R. A. (2009). Use of maximum entropy modeling in wildlife research. *Entropy*, *11*(4), 854-866.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*(4), 729-744.

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(1), 37-48.

Banks, M. A., Blouin, M. S., Baldwin, B. A., Rashbrook, V. K., Fitzgerald, H. A., Blankenship, S. M., & Hedgecock, D. (1999). Isolation and inheritance of novel microsatellites in chinook salmon (*Oncorhynchus tschawytscha*).*Journal of Heredity*, *90*(2), 281-288.

Baraud, J. (1992) Coléoptères Scarabaeoidea d'Europe. Faune de France, France et Régions Limitrophes, 78, I–IX+1–856

Baringou, S., Rouault, J. D., Koken, M., Hardivillier, Y., Hurtado, L., & Leignel, V. (2016). Diversity of cytosolic HSP70 Heat Shock Protein from decapods and their phylogenetic placement within Arthropoda. *Gene*, *591*(1), 97-107.

Barluenga, M., Stölting, K. N., Salzburger, W., Muschick, M., & Meyer, A. (2006). Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature*,*439*(7077), 719-723.

Barnard, P. C. (2011) The Royal Entomological Society book of British insects. Published by Wiley-Blackwell (Chichester, West Sussex) and the Royal Entomological Society

Barševskis, A. (2001). New and rare species of beetles (Insecta: Coleoptera) in the Baltic states and Belarus. *Baltic Journal of Coleopterology*, 1(1-2), 3-18.

Barševskis, A., Valainis, U., Bičevskis, M., Savenkovs, N., Cibuļskis, R., Kalniņš, M., & Strode, N. (2004). Faunistic records of the beetles (Hexapoda: Coleoptera) in Latvia. 1. *Acta biologica universitatis daugavpiliensis*, 4(2), 93-106.

Bates, A., Blake, M., Harvey, D., Bower, L., Alexander, L., Jenkins, L., Sadler, J., McKeown, N., Shaw, P., Gange, A. & Green, A. (2014) Introduction to the Noble Chafer *Gnorimus nobilis* (Linnaeus, 1758) Leverhulme research project. *Worcestershire Record* 35 6-8

Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, 19, 10-15.

Beale, C. M., Lennon, J. J., & Gimona, A. (2008). Opening the climate envelope reveals no macroscale associations with climate in European birds. *Proceedings of the National Academy of Sciences*, 105(39), 14908-14912.

Beere, H. M., Wolf, B. B., Cain, K., Mosser, D. D., Mahboubi, A., Kuwana, T., Tailor, P., Morinoto, R. I., Cohen, G. M. & Green, D. R. (2000). Heat-shock protein 70 inhibits apoptosis by preventing recruitment of procaspase-9 to the Apaf-1 apoptosome. *Nature cell biology*, *2*(8), 469-475.

Bekchiev, R., & Zlatkov, B. (2010). On the distribution of *Propomacrus bimucronatus* (Pallas, 1781)(Col.: Euchiridae) in Bulgaria. *Entomologist's Record and Journal of Variation*, *122*(6), 274.

Bellmann, A. (2002). Die Trogidae, Geotrupidae, Scarabaeidae und Lucanidae (Coleoptera) des Weser-Ems-Gebietes. *DROSERA-Naturkundliche Mitteilungen aus Norddeutschland*, 2002(1/2), 109-128.

Benítez, H. A., Sanzana, M. J., Jerez, V., Parra, L. E., Hernández, C. E., & Canales-Aguirre, C. B. (2013). Sexual shape and size dimorphism in carabid beetles of the genus *Ceroglossus*: is geometric body size similar between sexes due to sex ratio?. *Zoological science*, *30*(4), 289-295.

Berrigan, D. (1991). The allometry of egg size and number in insects. *Oikos*, 313-321.

Besnard, G., Bertrand, J. A., Delahaie, B., Bourgeois, Y. X., Lhuillier, E., & Thébaud, C. (2016). Valuing museum specimens: high-throughput DNA sequencing on historical collections of New Guinea crowned pigeons (*Goura*). *Biological Journal of the Linnean Society*,*117*(1), 71-82.

Bi, K., Linderoth, T., Vanderpool, D., Good, J. M., Nielsen, R., & Moritz, C. (2013). Unlocking the vault: next-generation museum population genomics. *Molecular ecology*, *22*(24), 6018-6032.

Birks, H. J. B. (1989) Holocene isochrones maps and patterns of tree-spreading in the British Isles. *Journal of Biogeography* 16 503-540

Birks, H. J. B., & Willis, K. J. (2008). Alpines, trees, and refugia in Europe. *Plant Ecology & Diversity*, 1(2), 147-160.

Birky, C. W., Fuerst, P., & Maruyama, T. (1989). Organelle gene diversity under migration, mutation, and drift: equilibrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. *Genetics*, *121*(3), 613-627.

Birky, C. W., Maruyama, T., & Fuerst, P. (1983). An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics*, *103*(3), 513-527.

Bishop, W.W. & Coope, G.R. (1977). Stratigraphical and Faunal Evidence for Lateglacial and Early Flandrian Environments in South-West Scotland. In J.M.Gray & J.J.Lowe (eds.) Studies in the Scottish Lateglacial Environment, 61-88. Pergamon, Oxford.

Blackledge, T. A. (2012) Spider silk: a brief review and prospectus on research linking biomechanics and ecology in draglines and orb webs. *The Journal of Arachnology* **40** 1-12

Blackledge, T. A., Scharff, N., Coddington, J. A., Szüts, T., Wenzel, J. W., Hayashi, C. Y. & Agnarsson, I. (2009) Reconstructing web evolution and spider diversification in the molecular era. *Proceedings of the National Academy of Sciences* **106** 5229-5234

Boddy, L. (2001). Fungal community ecology and wood decomposition processes in angiosperms: from standing tree to complete decay of coarse woody debris. *Ecological Bulletins*, 43-56.

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., Yu, D. W. & de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology & Evolution*, *29*(6), 358-367.

Bouget, C., Larrieu, L., & Brin, A. (2014). Key features for saproxylic beetle diversity derived from rapid habitat assessment in temperate forests. *Ecological Indicators*, *36*, 656-664.

Bowen, B. W. (1999). Preserving genes, species, or ecosystems? Healing the fractured foundations of conservation policy. *Molecular ecology*, *8*(s1).

Brelih, S., Kajzer, A., & Pirnat, A. (2010). Gradivo za favno hroščev (Coleoptera) Slovenije 4. Prispevek: Polyphaga: Scarabaeoidea (= Lamellicornia). *Scopolia*, 70, 13-86.

Brito, P. H., & Edwards, S. V. (2009). Multilocus phylogeography and phylogenetics using sequence-based markers. *Genetica*, *135*(3), 439-455.

Brower, A. V. Z. (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences*, *91*(14), 6491-6495.

Brower, A. V. Z., & DeSalle, R. (1998). Patterns of mitochondrial versus nuclear DNA sequence divergence among nymphalid butterflies: the utility of wingless as a source of characters for phylogenetic inference. *Insect molecular biology*, *7*(1), 73-82.

Browne, J. & Scholtz, C. H. (1999) A phylogeny of the families of Scarabaeoidea (Coleoptera*). Systematic Entomology* 24 51-84

de Bruyn, M., Hoelzel, A. R., Carvalho, G. R., & Hofreiter, M. (2011). Faunal histories from Holocene ancient DNA. *Trends in Ecology & Evolution*, *26*(8), 405-413.

Buckland, P.I. & Buckland, P.C. (2006). BugsCEP Coleopteran Ecology Package. IGBP PAGES/World Data Center for Paleoclimatology Data Contribution Series # 2006-116. NOAA/NCDC Paleoclimatology Program, Boulder CO, USA. URL: http://www.ncdc.noaa.gov/paleo/insect.html

Buckland, P.I. (2007). "The Development and Implementation of Software for Palaeoenvironmental and Palaeoclimatological Research: The Bugs Coleopteran Ecology Package (BugsCEP)". PhD thesis, Environmental Archaeology Lab., Department of Archaeology & Sámi Studies. University of Umeå, Sweden. Archaeology and Environment 23, 236 pp + CD. Available online: http://www.diva-portal.org/umu/abstract.xsql?dbid=1105

Bunalski, M. (2000). Contribution to the knowledge Scarabaeoidea (Coleoptera) of Bulgaria. Part I. Results of the Expeditions 1996 and 1998.*Wiadomosti entomologiczne*, 19(2), 85-92.

Bunalski, M., Przewoźny, M., Ruta, R., Borowiak-Sobkowiak, B., Trzciński, P., & Sienkiewicz, P. (2012). Materiały do poznania rozmieszczenia chrząszczy (Coleoptera) Zachodniej Polski. *Część*, 2, 90-99.

Buse, J., Griebeler, E. M., & Niehuis, M. (2013). Rising temperatures explain past immigration of the thermophilic oak-inhabiting beetle *Coraebus florentinus* (Coleoptera: Buprestidae) in south-west Germany. *Biodiversity and conservation*, *22*(5), 1115-1131.

Buse, J., Ranius, T., & Assmann, T. (2008 a.). An endangered longhorn beetle associated with old oaks and its possible role as an ecosystem engineer. *Conservation Biology*, *22*(2), 329-337.

Buse, J., Schröder, B., & Assmann, T. (2007). Modelling habitat and spatial distribution of an endangered longhorn beetle–a case study for saproxylic insect conservation. *Biological Conservation*, 137(3), 372-381

Buse, J., Zábranský, P., & Assmann, T. (2008 b.). The xylobiontic beetle fauna of old oaks colonised by the endangered longhorn beetle *Cerambyx cerdo* Linnaeus, 1758 (Coleoptera: Cerambycidae). *Mitt. Dtsch. Ges. allg. angew. Ent*, *16*, 109-112.

Bussler, H., Müller, J., & Dorka, V. (2005). European natural heritage: the saproxylic beetles in the proposed Parcul National Defileul Jiului. *Analele Icas*, 18, 55-71.

Butcher, B. A., Smith, M. A., Sharkey, M. J., & Quicke, D. L. (2012). A turbo-taxonomic study of Thai *Aleiodes* (*Aleiodes*) and *Aleiodes* (*Arcaleiodes*)(Hymenoptera: Braconidae: Rogadinae) based largely on COI barcoded specimens, with rapid descriptions of 179 new species. *Zootaxa*, *3457*, 1-232.

Butler, J. M., Shen, Y., & McCord, B. R. (2003). The development of reduced size STR amplicons as tools for analysis of degraded DNA. *Journal of forensic sciences*, *48*(5), 1054-1064.

Calvignac, S., Konecny, L., Malard, F., & Douady, C. J. (2011). Preventing the pollution of mitochondrial datasets with nuclear mitochondrial paralogs (numts). *Mitochondrion*, *11*(2), 246-254.

Campanaro, A., Zapponi, L., Hardersen, S., Méndez, M., Al Fulaij, N., Audisio, P., Bardiani, M., Carpaneto, G. M., Corezzola, S., Della Rocca, F., Harvey, D., Hawes, C., Kadej, M., Karg, J., Rink, M., Smolis, A., Sprecher, E., Thomaes, A., Toni, I., Vrezec, A., Zauli, A., Zilioli, M. & Chiari, S. (2016). A European monitoring protocol for the stag beetle, a saproxylic flagship species. *Insect Conservation and Diversity*. doi: 10.1111/icad.12194

Campbell, N. A., & Atchley, W. R. (1981). The geometry of canonical variate analysis. *Systematic Biology*, *30*(3), 268-280.

Campbell, G., & Robinson, M. (2007). Environment and land use in the valley bottom. *The Raunds Area Project. A Neolithic and Bronze Age Landscape in Northamptonshire*, *1*, 18-34.

Cardoso, P., Borges, P. A., Triantis, K. A., Ferrández, M. A., & Martín, J. L. (2011a). Adapting the IUCN Red List criteria for invertebrates. *Biological Conservation*, *144*(10), 2432-2440.

Cardoso, P., Erwin, T. L., Borges, P. A., & New, T. R. (2011b). The seven impediments in invertebrate conservation and how to overcome them. *Biological Conservation*, *144*(11), 2647-2655.

Cardoso, P., Silva, I., de Oliveira, N. G., & Serrano, A. R. (2004). Higher taxa surrogates of spider (Araneae) diversity and their efficiency in conservation. *Biological Conservation*, *117*(4), 453-459.

Carpaneto, G. M., Baviera, C., Biscaccianti, A. B., Brandmayr, P., Mazzei, A., Mason, F., Battistoni, A., Teofili, C., Rondinini, C., Fattorini, S. & Audisio, P. (2015). A Red List of Italian Saproxylic Beetles: taxonomic overview, ecological features and conservation issues (Coleoptera). *Fragmenta entomologica*, *47*, 53-126.

Centeno-Cuadros, A., Delibes, M., & Godoy, J. A. (2009). Phylogeography of Southern Water Vole (*Arvicola sapidus*): evidence for refugia within the Iberian glacial refugium*? Molecular Ecology*, 18(17), 3652-3667.

Cha, S. Y., Yoon, H. J., Lee, E. M., Yoon, M. H., Hwang, J. S., Jin, B. R., Han, Y. S. & Kim, I. (2007). The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae). *Gene*, *392*(1), 206-220.

Chamberlain, S., Ram, K., Barve, V., & Mcglinn, D. (2016). rgbif: Interface to the Global Biodiversity Information Facility API. R package version 0.9, 3.

Chapman, D. S. (2010). Weak climatic associations among British plant distributions. *Global Ecology and Biogeography*, 19(6), 831-841

Chapman, J. R., Nakagawa, S., Coltman, D. W., Slate, J., & Sheldon, B. C. (2009). A quantitative review of heterozygosity–fitness correlations in animal populations. *Molecular Ecology*, *18*(13), 2746-2765.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, *10*(3), 195-205.

Charlesworth, B., Nordborg, M., & Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genetical research*, *70*(02), 155-174.

Chen, M., Yao, R., Su, J. F., Wang, H., Li, Y. P., Liu, Y. Q., Jiang, X.-F., Qin, L., Wang, Z.-D. & Lu, C. (2012). Length polymorphism and structural organization of the A+ T-rich region of mitochondrial DNA in *Antheraea pernyi* (Lepidoptera: Saturniidae).*Biochemical Systematics and Ecology*, *43*, 169-177.

Chiari, S., Zauli, A., Audisio, P., & Carpaneto, G. M. (2014). Interactions between larvae of the threatened saproxylic beetle *Osmoderma eremita* and other flower chafers in Mediterranean woodlands: implications for conservation. *Insect Conservation and Diversity*, *7*(5), 462-469.

Chapman, R. F. (1998). *The insects: structure and function*. Cambridge university press.

Chown, S. L., & Gaston, K. J. (2010). Body size variation in insects: a macroecological perspective. *Biological Reviews*, *85*(1), 139-169.

Colón, J. I. L. (2003). Datos sobre la alimentación y distribución ibérica de *Oryctes nasicornis grypus* Illiger, 1803 (Coleoptera, Scarabaeidae, Dynastinae). *Boletín de la SEA*, (33), 183-188.

Concha, C., Edman, R. M., Belikoff, E. J., Schiemann, A. H., Carey, B., & Scott, M. J. (2012). Organization and expression of the Australian sheep blowfly (Lucilia cuprina) hsp23, hsp24, hsp70 and hsp83 genes. *Insect molecular biology*, *21*(2), 169-180.

Coope, G. R. (1977). Fossil Coleopteran Assemblages as Sensitive Indicators of Climatic Changes During the Devensian (Last) Cold Stage. *Philosophical Transactions of the Royal Society of London Series B*, 280, 313-337.

Coope, G. R. (1998). Insects. In, R. C. Preece & D. R. Bridgeland (eds.) Late Quaternary environmental change in North-west Europe: Excavations at Holywell Coombe, South-east England, 213-233. Chapman & Hall, London

Coope, G. R. (2010). Coleopteran faunas as indicators of interglacial climates in central and southern England. *Quaternary Science Reviews*,29(13), 1507-1514.

Cornuet, J. M, Piry, S., Luikart, G., Estoup, A. & Solignac, M (1999) New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153 1989-2000

Courchamp, F., Angulo, E., Rivalan, P., Hall, R. J., Signoret, L., Bull, L., & Meinard, Y. (2006). Rarity value and species extinction: the anthropogenic Allee effect. *PLoS Biol*, *4*(12), e415.

Cox, K., Thomaes, A., Antonini, G., Zilioli, M., De Gelas, K., Harvey, D., Solano, E., Audisio, P., McKeown, N., Shaw, P. & Minetti, R. (2013) Testing the performance of a fragment of the COI gene to identify western Palaearctic stag beetle species (Coleoptera, Lucanidae). *ZooKeys*, (365), p.105.

Crandall, K. A., Bininda-Emonds, O. R. P, Mace, G. M. & Wayne, R. K. (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution* 15 290-295

Crespo, L. C., Silva, I., Borges, P. A., & Cardoso, P. (2014). Assessing the conservation status of the strict endemic Desertas wolf spider, *Hogna ingens* (Araneae, Lycosidae). *Journal for Nature Conservation*, *22*(6), 516-524.

Criscione, C. D., Poulin, R., & Blouin, M. S. (2005). Molecular ecology of parasites: elucidating ecological and microevolutionary processes. *Molecular ecology*, 14(8), 2247-2257.

Crnokrak, P., & Roff, D. A. (1999). Inbreeding depression in the wild. *Heredity*, *83*(3), 260-270.

Crockatt, M. E., Campbell, A., Allum, L., Ainsworth, A. M., & Boddy, L. (2010). The rare oak polypore *Piptoporus quercinus*: Population structure, spore germination and growth. *Fungal Ecology*, 3(2), 94-106.

Crow, J. F., & Morton, N. E. (1955). Measurement of gene frequency drift in small populations. *Evolution*, 202-214.

Crowson, R. A. (1981). The biology of the Coleoptera. Academic Press.

de la Giroday, H. M. C., Carroll, A. L., & Aukema, B. H. (2012). Breach of the northern Rocky Mountain geoclimatic barrier: initiation of range expansion by the mountain pine beetle. *Journal of Biogeography*, *39*(6), 1112-1123.

De Queiroz, K. (2011) Branches in the lines of descent: Charles Darwin and the evolution of the species concept. *Biological Journal of the Linnean Society* 103 19-35

de Vries, H. H., denBoer, P. J. & vanDijk, T. S. (1996) Ground beetle species in heathland fragments in relation to survival, dispersal, and habitat preference. *Oecologia* 107 332-342

Dalmaris, E., Ramalho, C. E., Poot, P., Veneklaas, E. J., & Byrne, M. (2015). A climate change context for the decline of a foundation tree species in south-western Australia: insights from phylogeography and species distribution modelling. *Annals of botany*, *116*(6), 941-952.

Damgaard, J., & Sperling, F. A. (2001). Phylogeny of the water strider genus *Gerris* Fabricius (Heteroptera: Gerridae) based on COI mtDNA, EF-1α nuclear DNA and morphology. *Systematic Entomology*, *26*(2), 241-254.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, *9*(8), 772-772.

De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. *Molecular Ecology Resources*, *14*(2), 306-323.

Deagle, B. E., Eveson, J. P., & Jarman, S. N. (2006). Quantification of damage in DNA recovered from highly degraded samples–a case study on DNA in faeces. *Frontiers in Zoology*,*3*(1), 1.

Dhananjeyan, K. J., Paramasivan, R., Tewari, S. C., Rajendran, R., Thenmozhi, V., Victor Jerald Leo, S., Venkatesh, A. & Tyagi, B. K. (2010). Molecular identification of mosquito vectors using genomic DNA isolated from eggshells, larval and pupal exuvium. *Trop Biomed*, *27*(1), 47-53.

Donald, H. M., Wood, C. W., Benowitz, K. M., Johnson, R. A., Brodie III, E. D. & Formica, V. A. (2012) Nondestructive sampling of insect DNA from defensive secretion. *Molecular Ecology Resources* **12** 856-860

Drag, L., Hauck, D., Bérces, S., Michalcewicz, J., Šerić Jelaska, L., Aurenhammer, S., & Cizek, L. (2015). Genetic differentiation of populations of the threatened saproxylic beetle Rosalia longicorn, *Rosalia alpina* (Coleoptera: Cerambycidae) in Central and South-east Europe. *Biological Journal of the Linnean Society*, 116(4), 911-925.

Drees, C., Hüfner, S., Matern, A., Nève, G. & Assmann, T. (2011) Repeated sampling detects gene flow in a flightless ground beetle in a fragmented landscape. *Hereditas* 148 36-45

Drees, C., Matern, A. Rasplus, J.-V., Terlutter, H., Assmann, T. & Weber, F. (2008) Microsatellites and allozymes as the genetic memory of habitat fragmentation and defragmentation in populations of the ground beetle *Carabus auronitens* (Col., Carabidae). *Journal of Biogeography* 35 1937-1949

Dubois, G. F., Vernon, P., & Brustel, H. (2009). A flight mill for large beetles such as *Osmoderma eremita* (Coleoptera: Cetoniidae). In *Buse J, Alexander KNA, Ranius T, Assmann T, Saproxylic Beetles: Their Role and Diversity in European Woodland and Tree Habitats-Proceedings of the 5th Symposium and Workshop. Pensoft Publishers, Sofia* (pp. 219-224).

Dubois, G. F., Vignon, V., Delettre, Y. R., Rantier, Y., Vernon, P. & Burel, F. (2009) Factors affecting the occurrence of the endangered saproxylic beetle *Osmoderma eremitia* (Scopoli, 1763) (Coleoptera: Cetoniidae) in an agricultural landscape. *Landscape and Urban Planning* 91 152-159

Duff, A. (1993). Beetles of Somerset: their status and distribution. Somerset Archaeological & Natural History Society, Taunton.

Dyall, S. D., Brown, M. T., & Johnson, P. J. (2004). Ancient invasions: from endosymbionts to organelles. *Science*, *304*(5668), 253-257.

Eldred, T., Meloro, C., Scholtz, C., Murphy, D., Fincken, K., & Hayward, M. (2016) Does size matter for horny beetles? A geometric morphometric analysis of interspecific and intersexual size and shape variation in *Colophon haughtoni* Barnard, 1929, and *C. kawaii* Mizukami, 1997 (Coleoptera: Lucanidae). *Organisms Diversity & Evolution*, 1-13.

Eliasson, C. U. (2000) Is Gnorimus variabilis (Coleoptera: Scarabaeidae) native at its northern limit? *Entomologisk Tidskrift* 121 173-179

Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 677.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and distributions*, 17(1), 43-57.

Ellegren, H. (1999). Inbreeding and relatedness in Scandinavian grey wolves *Canis lupus*. *Hereditas*, *130*(3), 239-244.

Enyedi, R. (2006). Contributions to the Scarabaeoidea fauna of Hungary (Coleoptera: Scarabaeoidea). *Folia Historico Naturalia Musei Matraensis*, 30, 215-225.

Ersts, P.J. (2016) Geographic Distance Matrix Generator (version 1.2.3). American Museum of Natural History, Center for Biodiversity and Conservation. Accessed on 2016-4-5.

Estoup, A., Jarne, P., & Cornuet, J. M. (2002). Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis. *Molecular ecology*, *11*(9), 1591-1604.

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

Excoffier, L., Foll, M., & Petit, R. J. (2009). Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, *40*, 481-501.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479-491.

Feinstein, J. (2004) DNA sequence from butterfly frass and exuviae. *Conservation Genetics* 5 103-104

Feliner, G. N. (2011). Southern European glacial refugia: a tale of tales. *Taxon*, *60*(2), 365-372.

Ficetola, G. F., Padoa-Schioppa, E., Wang, J., & Garner, T. W. (2010). Polygyny, census and effective population size in the threatened frog, *Rana latastei*. *Animal conservation*, *13*(s1), 82-89.

Fielding, A. H., & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental conservation*, 24(01), 38-49.

Finkelstein, A., Rubin, L. L., & Tzeng, M. C. (1976). Black widow spider venom: effect of purified toxin on lipid bilayer membranes. *Science*,*193*(4257), 1009-1011.

Fisher, R. A. (1930). *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press.

Fitzpatrick, J. L. & Evans, J. P. (2009) Reduced heterozygosity impairs sperm quality in endangered mammals. *Biology Letters* 5 320-323

Flanders, J. O. N., Jones, G., Benda, P., Dietz, C., Zhang, S., Li, G., Sharifi, M. & Rossiter, S. J. (2009). Phylogeography of the greater horseshoe bat, *Rhinolophus ferrumequinum*: contrasting results from mitochondrial and microsatellite data. *Molecular Ecology*, *18*(2), 306-318.

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

Forester, B. R., DeChaine, E. G., & Bunn, A. G. (2013). Integrating ensemble species distribution modelling and statistical phylogeography to inform projections of climate change impacts on species distributions. *Diversity and Distributions*, *19*(12), 1480-1495.

Franc, V. (2002). Beetles (Coleoptera) of the Veľká Fatra Mts. with special reference to bioidicatively significant species. *Matthias Belivs Univ. Proc*, 165-177.

Frankham, R. (1995). Effective population size/adult population size ratios in wildlife: a review. *Genetical research*, *66*, 95-107.

Frankham, R. (2005) Genetics and extinction. *Biological Conservation* 126 131-140

Frankham, R. (2010). Challenges and opportunities of genetic approaches to biological conservation. *Biological conservation*, *143*(9), 1919-1927.

Frankham, R., Ballou, J. D., Eldridge, M. D. B., Lacy. R. C., Ralls, K., Dudash, M. R. & Fenster, C. B. (2011) Predicting the probability of outbreeding depression. *Conservation Biology* 25 465-475

Frankham, R., Bradshaw, C. J., & Brook, B. W. (2014). Genetics in conservation management: revised recommendations for the 50/500 rules, Red List criteria and population viability analyses. *Biological Conservation*, *170*, 56-63.

Frankel, O. H. (1974). Genetic conservation: our evolutionary responsibility. *Genetics*, *78*(1), 53-65.

Franklin, J. (2010). Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions*, 16(3), 321-330.

Freeland, J. R., Kirk, H. & Petersen, S. D. (2011) Molecular Ecology (2nd ed.) Published by Wiley-Blackwell (Chichester, West Sussex)

Fu, Y. X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, *147*(2), 915-925.

Fu, Y. X., & Li, W. H. (1993). Statistical tests of neutrality of mutations. *Genetics*, 133(3), 693-709.

Fumagalli, L., Taberlet, P., Favre, L., & Hausser, J. (1996). Origin and evolution of homologous repeated sequences in the mitochondrial DNA control region of shrews. *Molecular Biology and Evolution*, *13*(1), 31-46.

Funk, D. J., & Omland, K. E. (2003). Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, 397-423.

Garcia de Leaniz, C., Fleming, I. A., Einum, S., Verspoor, E., Jordan, W. C., Consuegra, S., Aubin-North, N., Lajus, D., Letcher, B. H., Youngson, A. F., Webb, J. H., ¬Vøllestad, L. A., Villanueva, B., Ferguson, A. & Quinn, T. P. a (2007). A critical review of adaptive genetic variation in Atlantic salmon: implications for conservation. *Biological Reviews*, *82*(2), 173-211.

Gardiner, M. M., Allee, L. L., Brown, P. M., Losey, J. E., Roy, H. E., & Smyth, R. R. (2012). Lessons from lady beetles: accuracy of monitoring data from US and UK citizen-science programs. *Frontiers in Ecology and the Environment*, *10*(9), 471-476.

Garnier, S., Magniez-Jannin, F., Rasplus, J. Y., & Alibert, P. (2005). When morphometry meets genetics: inferring the phylogeography of *Carabus solieri* using Fourier analyses of pronotum and male genitalia. *Journal of evolautionary biology*,*18*(2), 269-280.

Garrick, R. C., Sands, C. J., Rowell, D. M., Tait, N. N., Greenslade, P., & Sunnucks, P. (2004). Phylogeography recapitulates topography: very fine-scale local endemism of a saproxylic 'giant' springtail at Tallaganda in the Great Dividing Range of south–east Australia. *Molecular Ecology*,13(11), 3329-3344.

Garrick, R. C., Sands, C. J. & Sunnucks, P. (2006) The use and application of phylogeography for invertebrate conservation research and planning. In: Insect biodiversity and dead wood: proceedings of a symposium for the 22d International Congress of Entomology (pp. 15-22). Grove, S.J., Hanula, J.L., eds.

Garrick, R. C., Sunnucks, P., & Dyer, R. J. (2010). Nuclear gene phylogeography using PHASE: dealing with unresolved genotypes, lost alleles, and systematic bias in parameter estimation. *BMC Evolutionary Biology*, *10*(1), 1.

Garza, J. C., & Williamson, E. G. (2001). Detection of reduction in population size using data from microsatellite loci. *Molecular ecology*,*10*(2), 305-318.

Ge, X., Zong, S., He, S., Liu, Y., & Kong, X. (2014). Areas of China predicted to have a suitable climate for *Anoplophora chinensis* under a climate-warming scenario. *Entomologia Experimentalis et Applicata*, *153*(3), 256-265.

Georgiev, G. T., & Doychev, D. (2010). New xylophagous beetles (Insecta: Coleoptera) on poplars in Bulgaria. *Acta zool. bulg*, 62, 175-180.

Gibbons, P., Lindenmayer, D. B., Fischer, J., Manning, A. D., Weinberg, A., Seddon, J., Ryan, P. & Barrett, G. (2008). The future of scattered trees in agricultural landscapes. *Conservation Biology*,*22*(5), 1309-1319.

Gilbert, M. T. P., Moore, W., Melchoir, L. & Worobey, M. (2007) DNA extraction from dry museum beetles without conferring external morphological damage. PLoS ONE 2 e272

Giribet, G., & Edgecombe, G. D. (2006). Conflict between datasets and phylogeny of centipedes: an analysis based on seven genes and morphology. *Proceedings of the Royal Society of London B: Biological Sciences*, *273*(1586), 531-538.

Gladieux, P., Feurtey, A., Hood, M.E., Snirc, A., Clavel, J., Dutech, C., Roy, M. and Giraud, T. (2015). The population biology of fungal invasions. *Molecular Ecology*, 24(9), pp.1969-1986.

Goldman, R. L., Daily, G. C. & Kareiva, P. (2010) Trade-offs in making ecosystem services and human well-being conservation priorities. *In: Trade-offs in Conservation, Wiley-Blackwell, Chichester*

Gompert, Z., Nice, C. C., Fordyce, J. A, Forister, M. L. & Shapiros, A. M. (2005) Identifying units for conservation using molecular systematics: the cautionary tale of the Karner Blue butterfly. *Molecular Ecology* 15 1759-1768

Gough, L. A., Sverdrup-Thygeson, A., Milberg, P., Pilskog, H. E., Jansson, N., Jonsell, M., & Birkemoe, T. (2015). Specialists in ancient trees are more affected by climate than generalists. *Ecology and evolution*, *5*(23), 5632-5641.

Gould, S. J. (1966). Allometry and size in ontogeny and phylogeny. *Biological Reviews*,*41*(4), 587-638.

Gokturk, T., & Mihli, A. (2015). New contributions to Scarabaeidae (Insecta: Coleoptera) fauna of the Artvin province in Turkey. *Annals of Agrarian Science*, 13(1), 71-78.

Gray, J. (1944). Studies in the Mechanics of the Tetrapod Skeleton. *Journal of Experimental Biology*, *20*(2), 88-116.

Grove, S. J. (2002). Saproxylic insect ecology and the sustainable management of forests. *Annual review of ecology and systematics*, 1-23.

Guéorguiev, B., Lobo, J., & Chehlarov, E. (2011). The Scarabaeoid Beetles (Insecta: Coleoptera: Scarabaeoidea) in the Bulgarian Section of the Western Rhodopes. *Biodiversity of Bulgaria*, 4, 237-265.

Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology*, *52*(5), 696-704

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

Gusakov, A.A., (2002) Plastinchatousye zhuki podsemeistva Trichiinae (Coleoptera, Scarabaeidae) v faune Rossii. Vidy roda *Gymnodus* Kirby. Poligrafia, *Kaluga*, 1-51

Haavik, L. J., Billings, S. A., Guldin, J. M., & Stephen, F. M. (2015). Emergent insects, pathogens and drought shape changing patterns in oak decline in North America and Europe. *Forest Ecology and Management*, *354*, 190-205.

Habel, J. C., Drees, C., Schmitt, T., & Assmann, T. (2010). Review: refugial areas and postglacial colonizations in the western Palearctic. In: Relict Species (pp. 189-197). Habel, J. C. & Assman, T. eds. Springer Berlin Heidelberg.

Habel, J. C., Roedder, D., Schmitt, T., & Neve, G. (2011). Global warming will affect the genetic diversity and uniqueness of *Lycaena helle* populations. *Global Change Biology*, *17*(1), 194-205.

Habel, J. C., Schmitt, T., & Müller, P. (2005). The fourth paradigm pattern of post-glacial range expansion of European terrestrial species: the phylogeography of the Marbled White butterfly (Satyrinae, Lepidoptera). *Journal of Biogeography*,32(8), 1489-1497.

Hadjiconstantis, M., Dimitriou, M., Makris, C. & Sfenthourakis, S. (2015) New data on '*Propomacrus cypriacus* Alexis & Makris, 2002' (Coleoptera: Euchiridae) of ecological, taxonomic and conservation significance. 13th International Congress on the Zoogeography and Ecology of Greece and Adjacent Regions

Hajibabaei, M., Singer, G. A., Hebert, P. D., & Hickey, D. A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *TRENDS in Genetics*,*23*(4), 167-172.

Hajibabaei, M., Smith, M., Janzen, D. H., Rodriguez, J. J., Whitfield, J. B., & Hebert, P. D. (2006). A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, *6*(4), 959-964.

Hall, S. J. G. & Bunce, R. G. H. (2011) Mature trees as keystone structures in Holarctic ecosystems – a quantitative species comparison in a northern English park. *Plant Ecology & Diversity* 4 243-250

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser*. 41:95-98.

Halliburton, R. (2004).*Introduction to population genetics*. Upper Saddle River: Pearson/Prentice Hall.

Hamilton, M. B. (2009) *Population Genetics*. Published by Wiley-Blackwell (Chichester, West Sussex)

Hampe, A., & Petit, R. J. (2005). Conserving biodiversity under climate change: the rear edge matters. *Ecology letters*, *8*(5), 461-467.

Hanski, I. (1982). Dynamics of regional distribution: the core and satellite species hypothesis. *Oikos*, 210-221.

Harmon, M. E., Franklin, J. F., Swanson, F. J., Sollins, P., Gregory, S. V., Lattin, J. D., Anderson, N. H., Cline, S. P., Aumen, N. G., Sedell, J. R., Lienkaemper, G. W., Cromack Jr., K. W. & Cummins, K. W. (1986). Ecology of coarse woody debris in temperate ecosystems. *Advances in ecological research*, *15*, 133-302.

Harrison, R. G. (1980). Dispersal polymorphisms in insects. *Annual Review of Ecology and Systematics*, *11*, 95-118.

Harpending, H. C., Sherry, S. T., Rogers, A. R., & Stoneking, M. (1993). The genetic structure of ancient human populations. *Current Anthropology*,*34*(4), 483-496.

Harper, G. L., Maclean, N., & Goulson, D. (2006). Analysis of museum specimens suggests extreme genetic drift in the adonis blue butterfly (*Polyommatus bellargus*). *Biological Journal of the Linnean Society*, *88*(3), 447-452.

Harvey, D. J., Gange, A. C., Hawes, C. J. & Rink, M. (2011) Bionomics and distribution of the stag beetle, *Lucanus cervus* (L.) across Europe. *Insect Conservation and Diversity* 4 23-38

Hauser, L., Adcock, G. J., Smith, P. J., Ramírez, J. H. B., & Carvalho, G. R. (2002). Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*).*Proceedings of the National Academy of Sciences*, *99*(18), 11742-11747.

Hawkins, R. D., Hon, G. C., & Ren, B. (2010). Next-generation genomics: an integrative approach. *Nature Reviews Genetics*, *11*(7), 476-486.

Hebert, P. D., Cywinska, A., & Ball, S. L. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biological Sciences*, *270*(1512), 313-321.

Hedin, J., Ranius, T., Nilsson, S. G. & Smith, H. G (2008) Restricted dispersal in a flying beetle assessed by telemetry. *Biodiversity and Conservation* 17 675-684

Hedrick, P. W. (1999). Perspective: highly variable loci and their interpretation in evolution and conservation. *Evolution*, 313-318.

Hedrick, P. W. & Fredrickson, R. (2010) Genetic rescue guidelines with examples from Mexican wolves and Florida panthers. *Conservation Genetics* 11 615-626

Hedrick, P. W., & Kalinowski, S. T. (2000). Inbreeding depression in conservation biology. *Annual review of ecology and systematics*, 139-162.

Hendrich, L., Morinière, J., Haszprunar, G., Hebert, P. D., Hausmann, A., Köhler, F., & Balke, M. (2015). A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. *Molecular Ecology Resources*, *15*(4), 795-818.

Hendrixson, B. E., Guice, A. V., & Bond, J. E. (2015). Integrative species delimitation and conservation of tarantulas (Araneae, Mygalomorphae, Theraphosidae) from a North American biodiversity hotspot. *Insect Conservation and Diversity*, *8*(2), 120-131.

Hertzog, L. R., Besnard, A., & Jay-Robert, P. (2014). Field validation shows bias-corrected pseudo-absence selection is the best method for predictive species-distribution modelling. *Diversity and distributions*, 20(12), 1403-1413.

Hewitt, G. M. (1988). Hybrid zones-natural laboratories for evolutionary studies. *Trends in Ecology & Evolution*, 3(7), 158-167.

Hewitt, G. M. (1996). Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological journal of the Linnean Society*, *58*(3), 247-276.

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

Hewitt, G. M. (2000). The genetic legacy of the Quaternary ice ages. *Nature*,405(6789), 907-913.

Hewitt, G. M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 359(1442), 183-195.

Hewitt, G. M. (2011). Mediterranean peninsulas: the evolution of hotspots. In: Biodiversity hotspots (pp. 123-147). Zachos, F. E. & Habel, J. C eds. Springer Berlin Heidelberg.

Hickling, R., Roy, D. B., Hill, J. K., Fox, R., & Thomas, C. D. (2006). The distributions of a wide range of taxonomic groups are expanding polewards. *Global change biology*, *12*(3), 450-455.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G., & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International journal of climatology*,25(15), 1965-1978.

Hijmans, R.J. (2012) Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, 93, 679–688.

Hill, W. G. (1981) Estimation of effective population size from data on linkage disequilibriums. *Genetical Research* 38 209-216

Hill, K. B., Marshall, D. C., Moulds, M. S., & Simon, C. (2015). Molecular phylogenetics, diversification, and systematics of Tibicen Latreille 1825 and allied cicadas of the tribe Cryptotympanini, with three new genera and emphasis on species from the USA and Canada (Hemiptera: Auchenorrhyncha: Cicadidae).*Zootaxa*, *3985*(2), 219.

Hoberg, E. P., & Brooks, D. R. (2015). Evolution in action: climate change, biodiversity dynamics and emerging infectious disease. *Phil. Trans. R. Soc. B*, *370*(1665), 20130553.

Hoberg, E. P., Galbreath, K. E., Cook, J. A., Kutz, S. J., & Polley, L. (2012). 1 Northern Host-Parasite Assemblages: History and Biogeography on the Borderlands of Episodic Climate and Environmental Transition. *Advances in parasitology*, 79, 1.

Hoelzel, A. R., Halley, J., O'Brien, S. J., Campagna, C., Arnborm, T., Le Boeuf, B., Ralls, K. & Dover, G. A. (1993). Elephant seal genetic variation and the use of simulation models to investigate historical population bottlenecks. *Journal of Heredity*, *84*(6), 443-449.

Hohenlohe, P. A., Bassman, S., Etter, P. D., Stiffler, N., Johnson, E. A., Cresko, W. A. (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* 6 e1000862

Hofreiter, M., & Stewart, J. (2009). Ecological change, range fluctuations and population dynamics during the Pleistocene. *Current biology*,*19*(14), R584-R594.

Holderegger, R., Kamm, U., & Gugerli, F. (2006). Adaptive vs. neutral genetic diversity: implications for landscape genetics. *Landscape Ecology*, *21*(6), 797-807.

Honeycutt, R. L., Hillis, D. M., & Bickham, J. W. (2010). Biodiversity Discovery and its importance to conservation. *Molecular approaches in natural resource conservation and management. Cambridge University Press, New York*

Hopkins, J. J. & Kirby, K. J. (2007) Review: ecological change in British broadleaves woodland since 1947. *Ibis* 149 (supplement 2) 29-40

Horák, J. (2014) Fragmented habitats of traditional fruit orchards are important for dead wood-dependent beetles associated with open canopy deciduous woodlands. *Naturwissenschaften* DOI 10.1007/s00114-014-1179-x (online only)

Horák, J., Chobot, K., & Horáková, J. (2012). Hanging on by the tips of the tarsi: a review of the plight of the critically endangered saproxylic beetle in European forests. *Journal for Nature Conservation*, 20(2), 101-108.

Horák, J., Hui, C., Roura-Pascual, N., & Romportl, D. (2013). Changing roles of propagule, climate, and land use during extralimital colonization of a rose chafer beetle. *Naturwissenschaften*, *100*(4), 327-336.

Hortal, J., Lobo, J. M., & Jiménez-Valverde, A. (2012). Basic questions in biogeography and the (lack of) simplicity of species distributions: putting species distribution models in the right place. *Natureza & Conservação*, 10(2), 108-118.

Hou, Z., Li, J., & Li, S. (2014). Diversification of low dispersal crustaceans through mountain uplift: a case study of *Gammarus* (Amphipoda: Gammaridae) with descriptions of four novel species. *Zoological Journal of the Linnean Society*, *170*(4), 591-633.

Hudson, R. R., Boos, D. D., & Kaplan, N. L. (1992). A statistical test for detecting geographic subdivision. *Molecular biology and evolution*, *9*(1), 138-151.

Hudson, R. R., & Turelli, M. (2003). Stochasticity overrules the "three-times rule": genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, *57*(1), 182-190.

Huang, J.-P. & Lin, C.-P. (2010) Diversification in subtropical mountains: phylogeography, Pleistocene demographic expansion, and evolution of polyphonic mandibles in Taiwanese stag beetle, *Lucanus formosanus*. *Molecular Phylogenetics and Evolution* 57 1149-1161

Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, *17*(8), 754-755.

Huff, D. D., Miller, L. M, Chizinski, C. J. & Vondracek, B. (2011) Mixed-source reintroductions lead to outbreeding depression in second-generation descendents of a native North American fish. *Molecular Ecology* 20 4246-4258

Hughes, A. L., & Nei, M. (1989). Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proceedings of the National Academy of Sciences*, *86*(3), 958-962.

Humphrey, J. W., Sippola, A. L., Lempérière, G., Dodelin, B., Alexander, K. N. A., & Butler, J. E. (2005). Deadwood as an indicator of biodiversity in European forests: from theory to operational guidance. *EFI Proc*, *51*, 193-206.

Hundsdoerfer, A. K., Mende, M. B., Kitching, I. J., & Cordellier, M. (2011). Taxonomy, phylogeography and climate relations of the Western Palaearctic spurge hawkmoth (Lepidoptera, Sphingidae, Macroglossinae). *Zoologica Scripta*, *40*(4), 403-417.

Hundsdoerfer, A. K., & Wink, M. (2006). Incongruence of morphology and genetic markers in *Hyles tithymali* (Lepidoptera: Sphingidae) from the Canary Islands. *Journal of Zoological Systematics and Evolutionary Research*,*44*(4), 316-322.

Hurst, G. 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 of London B: Biological Sciences*, *272*(1572), 1525-1534.

Husemann, M., Schmitt, T., Zachos, F. E., Ulrich, W., & Habel, J. C. (2014). Palaearctic biogeography revisited: evidence for the existence of a North African refugium for Western Palaearctic biota. *Journal of Biogeography*, 41(1), 81-94.

Hwang, J. S., Kim, Y. J., Bang, H. S., Yun, E. Y., Kim, Y. T., Kim, S. R., Park, K. H., Kim, N. S., Kim, I., Hwang, S. J. & Jeon, J. P. (2008). Isolation of antibacterial response genes from the dung beetle *Copris tripartitus* (Coleoptera: Scarabaeidae) immunized with Escherichia coli. *European Journal of Entomology*, *105*(2), 355

Hwang, U. W., & Kim, W. (1999). General properties and phylogenetic utilities of nuclear ribosomal DNA and mitochondrial DNA commonly used in molecular systematics. *The Korean journal of parasitology*, *37*(4), 215-228.

IBM Corp. (2013). IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.

Issac, N. J., Turvey, S. T., Collen, B., Waterman, C. & Baillie, J. E. (2007) Mammals on the EDGE: conservation priorities based on threat and phylogeny. *PLoS ONE* 2 e296

IUCN Red List, Version 3.1 (2010). Accessed on 14.01.2013

IUCN (2012) IUCN Red List of Threatened Species. Version 2012.2

IUCN (2016a) IUCN Red List of Threatened Species. Version 2016.2

IUCN (2016b) Guidelines for Using the IUCN Red List Categories and Criteria (Version 12.0).

Iwata, K. (1966) Ovarian eggs in the Scarabaeoidea (Coleoptera). *Seibutsu-Kenkyū* 10, 1-3

Jackson, R. R. & Brassington, R. J. (1987) The biology of *Pholcus phalangioides* (Araneae, Pholcidae): predatory versatility, araneophagy and aggressive mimicry. *Journal of Zoology* **211** 227-238

Jenkins, D. G., Carey, M., Czerniewska, J., Fletcher, J., Hether, T., Jones, A., Knight, S., Knox, J., Long, T., Mannino, M., McGuire, M., Riffle, A., Segelsky, S., Shappell, L., Sterner, A., Strickler, T., & Tursi, R. (2010). A meta-analysis of isolation by distance: relic or reference standard for landscape genetics?. *Ecography*, *33*(2), 315-320.

Jerde, C. L., Mahon, A. R., Chadderton, W. L., & Lodge, D. M. (2011). "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, *4*(2), 150-157.

Jessop, L. (1986) Dung beetles and chafers (Coleoptera: Scarabaeoidea). Handbooks for the Identification of British Insects vol. 5, part 11.

Jiggins, F. M., Bentley, J. K., Majerus, M. E., & Hurst, G. D. (2001). How many species are infected with Wolbachia? Cryptic sex ratio distorters revealed to be common by intensive sampling. *Proceedings of the Royal Society of London B: Biological Sciences*, *268*(1472), 1123-1126

Jiménez-Valverde, A., Lobo, J. M., & Hortal, J. (2008). Not as good as they seem: the importance of concepts in species distribution modelling. *Diversity and distributions*, 14(6), 885-890.

Johnson, H. (2008) The traditional British orchard; a precious and fragile resource. Article written online for Building Conservation, accessed on 20.05.2013 http://www.buildingconservation.com/articles/traditional-orchards/traditional-orchards.htm

Johnstone, R. A., & Hurst, G. D. (1996). Maternally inherited male-killing microorganisms may confound interpretation of mitochondrial DNA variability. *Biological Journal of the Linnean Society*, *58*(4), 453-470.

Jones, M. R., & Good, J. M. (2016). Targeted capture in evolutionary and ecological genomics. *Molecular ecology*, *25*(1), 185-202.

Jones, W. D., Nguyen, T. A. T., Kloss, B., Lee, K. J., & Vosshall, L. B. (2005). Functional conservation of an insect odorant receptor gene across 250 million years of evolution. *Current Biology*, *15*(4), R119-R121.

Jones, P. L., Pask, G. M., Rinker, D. C., & Zwiebel, L. J. (2011). Functional agonism of insect odorant receptor ion channels. *Proceedings of the National Academy of Sciences*, *108*(21), 8821-8825.

Jorde, P. E., & Ryman, N. (1995). Temporal allele frequency change and estimation of effective size in populations with overlapping generations. *Genetics*, *139*(2), 1077-1090.

Junqueira, A. C. M., Lessinger, A. C., & Azeredo-Espin, A. M. L. (2002). Methods for the recovery of mitochondrial DNA sequences from museum specimens of myiasis-causing flies. *Medical and veterinary entomology*, *16*(1), 39-45.

Juřičková, L., Horáčková, J., & Ložek, V. (2014). Direct evidence of central European forest refugia during the last glacial period based on mollusc fossils. *Quaternary Research*, *82*(1), 222-228.

Justice, J. L., Weese, D. A., & Santos, S. R. (2016). Phylogenetic utility, and variability in structure and content, of complete mitochondrial genomes among genetic lineages of the Hawaiian anchialine shrimp Halocaridina rubra Holthuis 1963 (Atyidae: Decapoda). *Mitochondrial DNA Part A*,*27*(4), 2710-2718.

Kadej, M., Zając, K., Smolis, A., Malkiewicz, A., Tarnawski, D., Kania, J., ... & Józefczuk, J. (2013). Nowe dane o rozsiedleniu wybranych gatunków poświętnikowatych (Insecta: Coleoptera: Scarabaeidae) w Polsce południowo-zachodniej. *Przyroda Sudetów*, 16, 95-114.

Kalinina, O. I. & Shabalin, S. A. (2008) Morphological features of larvae of *Anthracophora rusticola* Burmeister, 1842 (Coleoptera, Scarabaeidea, Cetoniinae), testing the affinity of the subfamilies Cetoniinae and Trichiinae. *Entomological Review* 88 579-581

Karl, S. A., Toonen, R. J., Grant, W. S. & Bowen, B. W. (2012) Common misconceptions in molecular ecology: echoes of the modern synthesis. *Molecular Ecology* 21 4171-4189

Kasic-Lelo, M., Lelo, S., & Zekic, D. Z. (2010). Comparative analyses of the representation of the cetoniidae leach family on mount ozren near sarajevo, in accordance with the stated temperature and barometric pressure. *Poljoprivreda i Sumarstvo*, 56(1-4), 39.

Kawano, K. (1995). Horn and wing allometry and male dimorphism in giant rhinoceros beetles (Coleoptera: Scarabaeidae) of tropical Asia and America. *Annals of the Entomological Society of America*, *88*(1), 92-99.

Kawano, K. (2006). Sexual dimorphism and the making of oversized male characters in beetles (Coleoptera). *Annals of the Entomological Society of America*, *99*(2), 327-341.

Keller, I., Excoffier, L. & Largiadèr, C. R. (2005) Estimation of effective population size and detection of a recent population decline coinciding with habitat fragmentation in a ground beetle. *Journal of Evolutionary Biology* 18 90-100

Keller, I., Nentwig, W. & Largiadèr, C. R. (2004) Recent habitat fragmentation due to roads can lead to significant genetic differentiation in an abundant flightless ground beetle. *Molecular Ecology* 13 2983-2994

Kim, M. J., Im, H. H., Lee, K. Y., Han, Y. S., & Kim, I. (2014). Complete mitochondrial genome of the whiter-spotted flower chafer, *Protaetia brevitarsis* (Coleoptera: Scarabaeidae).*Mitochondrial DNA*, 25(3), 177-178.

Kim, M. J., Jeong, S. Y., Jeong, J. C., Kim, S. S., & Kim, I. (2016). Complete mitochondrial genome of the endangered flower chafer *Osmoderma opicum* (Coleoptera: Scarabaeidae). *Mitochondrial DNA Part B*, 1-2.

Kimura, M. (1983). Rare variant alleles in the light of the neutral theory. *Molecular Biology and Evolution*, *1*(1), 84-93.

King, G. F., & Hardy, M. C. (2013). Spider-venom peptides: structure, pharmacology, and potential for control of insect pests. *Annual review of entomology*, *58*, 475-496.

Kirk, H., & Freeland, J. R. (2011). Applications and implications of neutral versus non-neutral markers in molecular ecology. *International journal of molecular sciences*, *12*(6), 3966-3988.

Klingenberg, C. P. (2011). MorphoJ: an integrated software package for geometric morphometrics. *Molecular ecology resources*, *11*(2), 353-357.

Klingenberg, C. P., Barluenga, M., & Meyer, A. (2002). Shape analysis of symmetric structures: quantifying variation among individuals and asymmetry. *Evolution*, *56*(10), 1909-1920.

Klingenberg, C. P., Duttke, S., Whelan, S., & Kim, M. (2012). Developmental plasticity, morphological variation and evolvability: a multilevel analysis of morphometric integration in the shape of compound leaves. *Journal of evolutionary biology*,*25*(1), 115-129.

Knowles, L. L., Carstens, B. C., & Keat, M. L. (2007). Coupling genetic and ecological-niche models to examine how past population distributions contribute to divergence. *Current Biology*, 17(11), 940-946.

Knowles, L. L., & Maddison, W. P. (2002). Statistical phylogeography. *Molecular Ecology*,*11*(12), 2623-2635.

Koch, K. (1989). Die Käfer Mitteleuropas, Ökologie 2. Goecke & Evers, Krefeld.

Kodandaramaiah, U., Simonsen, T. J., Bromilow, S., Wahlberg, N., & Sperling, F. (2013). Deceptive single-locus taxonomy and phylogeography: Wolbachia-associated divergence in mitochondrial DNA is not reflected in morphology and nuclear markers in a butterfly species. *Ecology and evolution*, *3*(16), 5167-5176.

Köhler F. (2000) Totholzkafer in Naturwaldzel- len des nordlichen Rheinlandes. Vergleich- ende Studien zur Totholzkdferfauna Deutsch- lands und deutschen Naturwaldforschung Recklinghausen: Landesamt Agrarordnung NRW. 351 pp.

Köhler, F. (2014). *Die klimabedingte Veränderung der Totholzkäferfauna (Coleoptera) des nördlichen Rheinlandes: Analysen zur Gesamtfauna und am Beispiel von Wiederholungsuntersuchungen in ausgewählten Naturwaldzellen*. Landesbetrieb Wald und Holz NRW.

Kohn, M. H., & Wayne, R. K. (1997). Facts from feces revisited. *Trends in ecology & evolution*,*12*(6), 223-227.

Kohnen, A., Wissemann, V., & Brandl, R. (2009). No genetic differentiation in the rose-infesting fruit flies *Rhagoletis alternata* and *Carpomya schineri* (Diptera: Tephritidae) across central Europe. *European Journal of Entomology*, *106*(2), 315.

Kolokotronis, S. O., MacPhee, R. D., & Greenwood, A. D. (2007). Detection of mitochondrial insertions in the nucleus (NuMts) of Pleistocene and modern muskoxen. *BMC evolutionary biology*, *7*(1), 1.

Kool, E. T. (2001). Hydrogen bonding, base stacking, and steric effects in DNA replication. *Annual review of biophysics and biomolecular structure*, *30*(1), 1-22.

Kopp, A., & True, J. R. (2002). Phylogeny of the Oriental *Drosophila melanogaster* species group: a multilocus reconstruction. *Systematic biology*, *51*(5), 786-805.

Koren, T., Rojko, I., & Laus, B. (2011, July). Additions to the faunal list of scarabaeoid beetles (insecta, scarabaeoidea) of the river zrmanja and its surroundings, Croatia. In Annales: Series Historia Naturalis (Vol. 21, No. 2, p. 125). Scientific and Research Center of the Republic of Slovenia.

Koressaar T. & Remm M. (2007) Enhancements and modifications of primer design program Primer3. *Bioinformatics* **23** 1289-91

Král, D., & Malý, V. (1993). New records of Scarabaeoidea (Coleoptera) from Bulgaria. *Acta Societatis Zoologicae Bohemoslovacae*, 57, 17-29.

Krehenwinkel, H., Kennedy, S., Pekár, S., & Gillespie, R. (2016). A cost efficient and simple protocol to enrich prey DNA from extractions of predatory arthropods for large scale gut content analysis by Illumina sequencing. *Methods in Ecology and Evolution*.

Krehenwinkel, H., & Pekar, S. (2015). An analysis of factors affecting genotyping success from museum specimens reveals an increase of genetic and morphological variation during a historical range expansion of a European spider. *PloS one*, *10*(8), e0136337.

Krell, F. T. (2010). *Trichius sexualis* Bedel, 1906–nomen protectum (Coleoptera, Scarabaeidae, Trichiinae). *Deutsche Entomologische Zeitschrift*, 57(2), 219-222.

Krell, F. T. (2012). On nomenclature and synonymy of *Trichius rosaceus, T. gallicus*, and *T. zonatus* (Coleoptera: Scarabaeidae: Cetoniinae: Trichiini). *Zootaxa*, 3278, 61-68.

Krell, F.T., Ballerio, A., Smith, A. B. T. & Audisio, P. (2006) *Gnorimus* Le Peletier de Saint- Fargeau & Serville, 1828 and *Osmoderma* Le Peletier de Saint-Fargeau & Serville, 1828 (Insecta, Coleoptera): proposed conservation of the generic names. *Bulletin of Zoological Nomenclature* 63 Case 3349

Krell, F. T., Rey, A., Micó Balaguer, E., & Dutto, M. (2012). On nomenclature and identity of *Scarabaeus aeruginosus* Linnaeus, *S. aeruginosus* Drury and *S. speciosissimus* Scopoli (Coleoptera: Scarabaeoidea: Cetoniinae and Rutelinae). *Revue Suisse de Zoologie* 119: 99–110.

Krell, F. T., & Wheeler, Q. D. (2014). Specimen collection: Plan for the future. *Science*, *344*(6186), 815-816.

Kremen, C., & Chaplin-Kramer, R. (2007). Insects as providers of ecosystem services: crop pollination and pest control. In *Insect conservation biology: proceedings of the royal entomological society's 23rd symposium* (pp. 349-382).

Kunz, W. (2013). *Do Species Exist: Principles of Taxonomic Classification*. John Wiley & Sons.

Lachaume, G. (1983) Goliathini 1. Goliathus, Argyrophegges, Fornasinius, Hegemus. Les Coléopterès du Monde. Published by Sciences Nat, Venette, Compiègne.

Lande, R., & Barrowclough, G. F. (1987). Effective population size, genetic variation, and their use in population management. *Viable populations for conservation*, 87-123.

Landvik, M., Niemelä, P., & Roslin, T. (2016). Mother knows the best mould: an essential role for non-wood dietary components in the life cycle of a saproxylic scarab beetle. *Oecologia*, 1-13.

Landvik, M., Wahlberg, N., & Roslin, T. (2013). The identity of the Finnish *Osmoderma* (Coleoptera: Scarabaeidae, Cetoniinae) population established by COI sequencing. *Entomologica Fennica*, *24*(3), 147-155.

Larsson, M. C., Domingos, A. I., Jones, W. D., Chiappe, M. E., Amrein, H., & Vosshall, L. B. (2004). Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron*, *43*(5), 703-714.

Larsson, M. C., Hedin, J., Svensson, G. P., Tolasch, T. & Francke, W. (2003) Characteristic odor of *Osmoderma eremita* identified as a male-released pheromone. *Journal of Chemical Ecology* 29 575-587

Lawrence, E. (2008) Henderson's Dictionary of Biology (14th ed.). Pearson Education Limited, Harlow, Essex, England

Lawrence, J. M., & Samways, M. J. (2003). Litter breakdown by the Seychelles giant millipede and the conservation of soil processes on Cousine Island, Seychelles. *Biological Conservation*, *113*(1), 125-132.

Lawton, J. H. & May, R. M. (1995) *Extinction Rates*. Oxford University Press. Oxford

Le Corre, V., & Kremer, A. (1998). Cumulative effects of founding events during colonisation on genetic diversity and differentiation in an island and stepping-stone model. *Journal of Evolutionary Biology*, *11*(4), 495-512.

Lee, H. L. & Boulding, E. G. (2009). Spatial and temporal population genetic structure of four northeastern Pacific littorinid gastropods: the effect of mode of larval development on variation at one mitochondrial and two nuclear DNA markers. *Molecular Ecology*, *18*(10), 2165-2184.

Lefort, M.-C., Boyer, S., Worner, S. P. & Armstrong, K. (2012) Noninvasive molecular methods to identify live scarab larvae: an example of sympatric pest and nonpest species in New Zealand. Molecular Ecology Resources 12 389-395

Lenoir, J., & Svenning, J. C. (2015). Climate-related range shifts–a global multidimensional synthesis and new research directions. *Ecography*, *38*(1), 15-28.

Levins, R. (1969) Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the Entomological Society of America* 15 237-240

Li, Z., & Srivastava, P. (2004). Heat-Shock Proteins. *Current Protocols in Immunology*, A-1T.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), 1451-1452.

Liefting, M., & Ellers, J. (2008). Habitat-specific differences in thermal plasticity in natural populations of a soil arthropod. *Biological Journal of the Linnean Society*, *94*(2), 265-271.

Lin, C. P., Huang, J. P., Lee, Y. H., & Chen, M. Y. (2011). Phylogenetic position of a threatened stag beetle, Lucanus datunensis (Coleoptera: Lucanidae) in Taiwan and implications for conservation. *Conservation Genetics*, *12*(1), 337-341.

Lindner, M., Maroschek, M., Netherer, S., Kremer, A., Barbati, A., Garcia-Gonzalo, J., Seidl, R., Delzon, S., Corona, P., Kolström, M., Lexer, M. J. & Marchetti, M. (2010). Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. *Forest Ecology and Management*,*259*(4), 698-709.

Lindenmayer, D. B., Piggott, M. P., & Wintle, B. A. (2013). Counting the books while the library burns: why conservation monitoring programs need a plan for action. *Frontiers in Ecology and the Environment*, *11*(10), 549-555.

Lindroth C.H. (1969) The biological importance of pleistocene refugia. In: The Kodiak Island Refugium: Its Geology, Flora, Fauna And History (pp. 7-17). Karlstrom, T. N. V. & Ball, G. E. eds. Ryerson Press: Toronto, Canada

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. *Global ecology and Biogeography*, *17*(2), 145-151.

Loiselle, B. A., Howell, C. A., Graham, C. H., Goerck, J. M., Brooks, T., Smith, K. G., & Williams, P. H. (2003). Avoiding pitfalls of using species distribution models in conservation planning. *Conservation biology*, 17(6), 1591-1600.

López-Colón, j. I., & Blasco-Zumeta, J. (1997). Nuevos registros de Scarabaeoidea de Aragón (Coleoptera). *Bol. SEA*, 18, 19-23.

Luck, G. W., Daily, G. C. & Ehrlich, P. R. (2003) Population diversity and ecosystem services. *Trends in Ecology and Evolution* 18 331-336

Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K., & Allendorf, F. W. (2010). Estimation of census and effective population sizes: the increasing usefulness of DNA-based approaches. *Conservation Genetics*, *11*(2), 355-373.

Lumbsch, H. T., Buchanan, P. K., May, T. W., & Mueller, G. M. (2008). Phylogeography and biogeography of fungi. *Mycological Research*,112(4), 423-424.

Lunt, D. H., Whipple, L. E., & Hyman, B. C. (1998). Mitochondrial DNA variable number tandem repeats (VNTRs): utility and problems in molecular ecology. *Molecular ecology*, *7*(11), 1441.

Ma, C., Liu, C., Yang, P., & Kang, L. (2009). The complete mitochondrial genomes of two band-winged grasshoppers, *Gastrimargus marmoratus* and *Oedaleus asiaticus*. *BMC genomics*, *10*(1), 1.

Macharia, R., Mireji, P., Murungi, E., Murilla, G., Christoffels, A., Aksoy, S., & Masiga, D. (2016). Genome-Wide Comparative Analysis of Chemosensory Gene Families in Five Tsetse Fly Species. *PLoS Negl Trop Dis*, *10*(2), e0004421.

Magri, D., Vendramin, G. G., Comps, B., Dupanloup, I., Geburek, T., Gömöry, D., ... & Tantau, I. (2006). A new scenario for the Quaternary history of European beech populations: palaeobotanical evidence and genetic consequences. *New Phytologist*, 171(1), 199-221.

Magurran, A. E., & Henderson, P. A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*, *422*(6933), 714-716.

Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., Chiquillo, E., Rønsted, N. & Antonelli, A. (2015). Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? *Global Ecology and Biogeography*, *24*(8), 973-984.

Maliouchenko, O., Palmé, A. E., Buonamici, A., Vendramin, G. G., & Lascoux, M. (2007). Comparative phylogeography and population structure of European *Betula* species, with particular focus on *B. pendula* and *B. pubescens*. *Journal of Biogeography*, 34(9), 1601-1610.

Malpel, S., Merlin, C., François, M. C., & Jacquin-Joly, E. (2008). Molecular identification and characterization of two new Lepidoptera chemoreceptors belonging to the *Drosophila melanogaster* OR83b family. *Insect molecular biology*, *17*(5), 587-596.

Mamidala, P., Wijeratne, A. J., Wijeratne, S., Poland, T., Qazi, S. S., Doucet, D., Cusson, M., Beliveau, C. & Mittapalli, O. (2013). Identification of odor-processing genes in the emerald ash borer, *Agrilus planipennis*. *PloS one*, *8*(2), e56555.

Mancini, E., De Biase, A., Mariottini, P., Bellini, A., & Audisio, P. (2008). Structure and evolution of the mitochondrial control region of the pollen beetle *Meligethes thalassophilus* (Coleoptera: Nitidulidae). *Genome*, *51*(3), 196-207.

Mann, D. J. (2006) Scarabaeoidea. 47-58 in: A Coleopterist's Handbook (4th ed.). Cooter, J. & Barclay, M. V. L. eds., The Amateur Entomologists' Society

Mannerkoski, I., Hyvärinen, E., Campanaro, A., Alexander, K., Büche, B., Dodelin, B., Mason, F., Pettersson, R., Mico, E. & Méndez, M. (2010a). *Gnorimus nobilis*. The IUCN Red List of Threatened Species 2010: e.T157566A5096743.

Mannerkoski, I., Hyvärinen, E., Campanaro, A., Alexander, K., Büche, B., Dodelin, B., Mason, F., Pettersson, R., Mico, E. & Méndez, M. (2010b). *Gnorimus variabilis*. The IUCN Red List of Threatened Species 2010: e.T157887A5166091

Mannerkoski, I., Hyvärinen, E., Campanaro, A., Alexander, K., Büche, B., Dodelin, B., Mason, F., Pettersson, R., Mico, E. & Méndez, M. (2010c). *Trichius fasciatus*. The IUCN Red List of Threatened Species 2010: e.T157860A5160897.

Mannerkoski, I., Pettersson, R., Mason, F. & Schmidl, J. 2010(d). *Trichius zonatus*. The IUCN Red List of Threatened Species 2010: e.T157853A5159430.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer research*, 27(2 Part 1), 209-220.

Mardulyn, P., Termonia, A., & Milinkovitch, M. C. (2003). Structure and evolution of the mitochondrial control region of leaf beetles (Coleoptera: Chrysomelidae): a hierarchical analysis of nucleotide sequence variation. *Journal of molecular evolution*, *56*(1), 38-45.

Marshall, S. A., & Evenhuis, N. L. (2015). New species without dead bodies: a case for photo-based descriptions, illustrated by a striking new species of *Marleyimyia* Hesse (Diptera, Bombyliidae) from South Africa. *ZooKeys*, (525), 117.

Marshall, T. C., Slate, J., Kruuk, L. E. B. & Pemberton, J. M. (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology* 7 639-655

Martin, D. P., Murrell, B., Golden, M., Khoosal, A., & Muhire, B. (2015). RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evolution*, *1*(1), vev003.

Mason, F., Schlaghamersky, J., Schmidl, J. & Petrakis, P. 2010(a) *Protaetia speciosissima*. The IUCN Red List of Threatened Species 2010: e.T157859A5160391

Mason, F., Schlaghamersky, J., Schmidl, J. & Petrakis, P. 2010(b) *Trichius sexualis*. The IUCN Red List of Threatened Species 2010: e.T157775A5143990

Massilani, D., Guimaraes, S., Brugal, J. P., Bennett, E. A., Tokarska, M., Arbogast, R. M., et al. & Madelaine, S. (2016). Past climate changes, population dynamics and the origin of *Bison* in Europe. *bioRxiv*, 063032.

MathWorks, (2012). *Bioinformatics Toolbox: User's Guide (R2012a)*.

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

Mazur, M. A., Holecová, M., Lachowska-Cierlik, D., Lis, A., Kubisz, D., & Kajtoch, Ł. (2016). Selective sweep of Wolbachia and parthenogenetic host genomes–the example of the weevil Eusomus ovulum. *Insect Molecular Biology*, *25*(6), 701-711.

McKay, J. K., & Latta, R. G. (2002). Adaptive population divergence: markers, QTL and traits. *Trends in Ecology & Evolution*, *17*(6), 285-291.

McKeown, N. J., & Shaw, P. W. (2008). Single paternity within broods of the brown crab Cancer pagurus: a highly fecund species with long-term sperm storage. *Marine Ecology Progress Series*,*368*, 209-215.

McKinney, M. L. (1999). High rates of extinction and threat in poorly studied taxa. *Conservation biology*, *13*(6), 1273-1281.

Medvedev, S.I., (1964) *Fauna of the USSR. Coleoptera. Vol. 10, Issue 4. Scarabaeidae of the Subfamilies Euchirinae, Dynastinae, Glaphyrinae, and Trichiinae* (Acad. Sci. USSR, Moscow-Leningrad, 1964)

Menéndez, R., Megías, A. G., Hill, J. K., Braschler, B., Willis, S. G., Collingham, Y., Fox, F., Roy, D. B. & Thomas, C. D. (2006). Species richness changes lag behind climate change. *Proceedings of the Royal Society of London B: Biological Sciences*,*273*(1593), 1465-1470.

Meraner, A., Brandstätter, A., Thaler, R., Aray, B., Unterlechner, M., Niederstätter, H., Parson, W., Zelger, E., Dalla Via, J. & Dallinger, R. (2008). Molecular phylogeny and population structure of the codling moth (Cydia pomonella) in Central Europe: I. Ancient clade splitting revealed by mitochondrial haplotype markers. *Molecular Phylogenetics and Evolution*, *48*(3), 825-837.

Micó, E., García-López, A., Sánchez, A., Juárez, M., & Galante, E. (2015). What can physical, biotic and chemical features of a tree hollow tell us about their associated diversity? *Journal of Insect Conservation*, *19*(1), 141-153.

Micó, E., Morón, M. A., Šipek, P. & Galente, E. (2008) Larval morphology enhances phylogenetic reconstruction in Cetoniidae (Coleoptera: Scarabaeoidea) and allows the interpretation of the evolution of larval feeding habits. *Systematic Entomology* 33 128-144

Miller, J. A., Beentjes, K. K., van Helsdingen, P., & IJland, S. (2013). Which specimens from a museum collection will yield DNA barcodes? A time series study of spiders in alcohol. *ZooKeys*, *365*, 245-261.

Miller, J. A., & Hormiga, G. (2004). Clade stability and the addition of data: a case study from erigonine spiders (Araneae: Linyphiidae, Erigoninae).*Cladistics*, *20*(5), 385-442.

Minelli, A., Boxshall, G., & Fusco, G. (Eds.). (2013). *Arthropod Biology and Evolution: Molecules, Development, Morphology*. Springer Science & Business Media.

Miraldo, A., Hewitt, G. M., Paulo, O. S., & Emerson, B. C. (2011). Phylogeography and demographic history of *Lacerta lepida* in the Iberian Peninsula: multiple refugia, range expansions and secondary contact zones. *BMC Evolutionary Biology*, *11*(1), 1.

Mondol, S., Karanth, K. U., Kumar, N. S., Gopalaswamy, A. M., Andheria, A., & Ramakrishnan, U. (2009). Evaluation of non-invasive genetic sampling methods for estimating tiger population size. *Biological Conservation*,*142*(10), 2350-2360.

Monroe, E. M., Lynch, C., Soluk, D. A. & Britten, H. B. (2010) Nonlethal tissue sampling techniques and microsatellite markers used for first report of genetic diversity in two populations of the endangered *Somatochlora hineana* (Odonata: Corduliidae). *Annals of the Entomological Society of America* **103** 1012-1017

Moore, W. S. (1995). Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, 718-726.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., & Worm, B. (2011). How many species are there on Earth and in the ocean?. *PLoS Biol*, *9*(8), e1001127.

Morando, M., Avila, L. J., Baker, J., & Sites, J. W. (2004). Phylogeny and phylogeography of the *Liolaemus darwinii* complex (Squamata: Liolaemidae): evidence for introgression and incomplete lineage sorting. *Evolution*, *58*(4), 842-859.

Morgan, M., Knisley, C. B., & Vogler, A. P. (2000). New taxonomic status of the endangered tiger beetle *Cicindela limbata albissima* (Coleoptera: Cicindelidae): evidence from mtDNA. *Annals of the Entomological Society of America*, *93*(5), 1108-1115.

Moreau, C. S., Wray, B. D., Czekanski-Moir, J. E., & Rubin, B. E. (2013). DNA preservation: a test of commonly used preservatives for insects. *Invertebrate systematics*, *27*(1), 81-86.

Morin, P. A., Wallis, J., Moore, J. J., Chakraborty, R., & Woodruff, D. S. (1993). Non-invasive sampling and DNA amplification for paternity exclusion, community structure, and phylogeography in wild chimpanzees. *Primates*,*34*(3), 347-356.

Moritz, C. (1994). Defining evolutionarily significant units' for conservation. *Trends in ecology and evolution*, *9*(10), 373-374.

Moss, R. H., Edmonds, J. A., Hibbard, K. A., Manning, M. R., Rose, S. K., Van Vuuren, D. P., et al. & Meehl, G. A. (2010). The next generation of scenarios for climate change research and assessment. *Nature*, *463*(7282), 747-756.

Mouillot, D., Bellwood, D. R., Baraloto, C., Chave, J., Galzin, R., Harmelin-Vivien, M., Kulbicki, M., Lavergne, S., Lavorel, S., Mouquet, N., Paine, C. T., Renaud, J. & Thuiller, W. (2013). Rare species support vulnerable functions in high-diversity ecosystems. *PLoS Biol*,*11*(5), e1001569.

Muafor, F. J., Levang, P., Angwafo, T. E., & Gall, P. L. (2012). Making a living with forest insects: beetles as an income source in Southwest Cameroon. *International Forestry Review*, *14*(3), 314-325.

Müller, J., Brustel, H., Brin, A., Bussler, H., Bouget, C., Obermaier, E., et al. (2015). Increasing temperature may compensate for lower amounts of dead wood in driving richness of saproxylic beetles. *Ecography*, *38*(5), 499-509.

Müller, J., Jarzabek-Müller, A., Bussler, H., & Gossner, M. M. (2014). Hollow beech trees identified as keystone structures for saproxylic beetles by analyses of functional and phylogenetic diversity. *Animal Conservation* 17 (2), 154-162.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*,*403*(6772), 853-858.

Nádai, L., & Merkl, O. (1999). Scarabaeoidea (Coleoptera) from the Aggtelek National Park. *The Fauna of the Aggtelek National Park, I. Hungarian Natural History Museum, Budapest*, 215-220.

Nagylaki, T. (1980). The strong-migration limit in geographically structured populations. *Journal of mathematical biology*, *9*(2), 101-114.

Nakahama, N., & Isagi, Y. (2016). Availability of short microsatellite markers from butterfly museums and private specimens. *Entomological Science*. doi: 10.1111/ens.12215

Nardi, G., Tykarski, P., Munteanu, N., Istrate, P., Horák, J., Tezcan, S. & Putchkov, A. 2010. *Propomacrus bimucronatus*. The IUCN Red List of Threatened Species 2010: e.T157553A5094351

Navajas, M., Le Conte, Y., Solignac, M., Cros-Arteil, S., & Cornuet, J. M. (2002). The complete sequence of the mitochondrial genome of the honeybee ectoparasite mite *Varroa destructor* (Acari: Mesostigmata). *Molecular biology and evolution*, *19*(12), 2313-2317.

Navascués, M., Depaulis, F., & Emerson, B. C. (2010). Combining contemporary and ancient DNA in population genetic and phylogeographical studies. *Molecular Ecology Resources*, *10*(5), 760-772.

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia university press.

Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, 1-10.

Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular biology and evolution*, *3*(5), 418-426.

Nei, M., & Tajima, F. (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics*,*97*(1), 145-163.

Neumann, K., Michaux, J. R., Maak, S., Jansman, H. A. H., Kayser, A., Mundt, G., & Gattermann, R. (2005). Genetic spatial structure of European common hamsters (*Cricetus cricetus*)—a result of repeated range expansion and demographic bottlenecks. *Molecular Ecology*,*14*(5), 1473-1483.

New, T. R. (2010). Beetles in Conservation. Wiley-Blackwell, London.

Nieberding, C. M., Durette-Desset, M. C., Vanderpoorten, A., Casanova, J. C., Ribas, A., Deffontaine, V., Feliu, C., Morand, S., Libois, R. & Michaux, J. R. (2008). Geography and host biogeography matter for understanding the phylogeography of a parasite. *Molecular phylogenetics and evolution*, *47*(2), 538-554.

Nieto, A., Dodelin, B., Campanaro, A. & Méndez, M. 2010. *Protaetia mirifica*. The IUCN Red List of Threatened Species 2010: e.T157803A5149718.

Nieto, A. Mannerkoski, I., Pettersson, R., Mason, F., Méndez, M. & Schmidl, J. 2010. *Lucanus cervus*. The IUCN Red List of Threatened Species 2010: e.T157554A5094499.

Nieto, A., Mannerkoski, I., Putchkov, A., Tykarski, P., Mason, F., Dodelin, B. & Tezcan, S. (2010). *Osmoderma eremita*. The IUCN Red List of Threatened Species 2010: e.T15632A4926651

Nieto, A., Petrakis, P., Schmidl, J. & Pettersson, R. 2010. *Propomacrus cypriacus*. The IUCN Red List of Threatened Species 2010: e.T157773A5143371.

Nieto, A. & Alexander, K. N. A. (2010) European Red List of Saproxylic Beetles. Publications Office of the European Union, Luxembourg.

Noriyuki, S., Kameda, Y., & Osawa, N. (2014). Prevalence of male-killer in a sympatric population of two sibling ladybird species, Harmonia yedoensis and Harmonia axyridis (Coleoptera: Coccinellidae). *European Journal of Entomology*, *111*(2), 307.

Nybom, H. (2004). Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular ecology*, *13*(5), 1143-1155.

Ober, K. A., & Connolly, C. T. (2015). Geometric morphometric and phylogenetic analyses of Arizona Sky Island populations of *Scaphinotus petersi* Roeschke (Coleoptera: Carabidae).*Zoological Journal of the Linnean Society*, *175*(1), 107-118.

Ødegaard, F., & Tømmerås, B. Å. (2000). Compost heaps—refuges and stepping-stones for alien arthropod species in northern Europe. *Diversity and distributions*, *6*(1), 45-59.

Oleksa, A., Chybicki, I. J., Gawroński, R., Svensson, G. P. & Burczyk, J. (2013) Isolation by distance in saproxylic beetles may increase with niche specialization. *Journal of Insect Conservation* **17** 219-233

Oleksa, A., Meyza, K., Cizek, L. & Drag, L. (2016) Microsatellite records: Characterization of microsatellite loci for the European species of hermit beetles (*Osmoderma* spp.) (Coleoptera: Cetoniidae). *Conservation Genetics Resources* 7(4):917-944 DOI: 10.1007/s12686-015-0493-8

Osborne, P. J. (1974). An Insect Assemblage of Early Flandrian Age from Lea Marston, Warwickshire and Its Bearing on the Contemporary Climate and Ecology. *Quaternary Research*, **4**, 471-486.

Paetkau, D., Waits, L. P., Stirling, I. & Strobeck, C. (1995) Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* **4** 347-354

Palmé, A. E., Su, Q., Rautenberg, A., Manni, F., & Lascoux, M. (2003). Postglacial recolonization and cpDNA variation of silver birch, *Betula pendula*. *Molecular ecology*, **12**(1), 201-212.

Palmer, A. R., & Strobeck, C. (1986). Fluctuating asymmetry: measurement, analysis, patterns. *Annual review of Ecology and Systematics*, 391-421.

Palstra, F. P., & Ruzzante, D. E. (2008). Genetic estimates of contemporary effective population size: what can they tell us about the importance of genetic stochasticity for wild population persistence? *Molecular Ecology*, *17*(15), 3428-3447.

Palumbi, S. R., Cipriano, F., & Hare, M. P. (2001). Predicting nuclear gene coalescence from mitochondrial data: the three-times rule. *Evolution*, *55*(5), 859-868.

Pamilo, P., Zhu, D., Fortelius, W., Rosengren, R., Seppä, P. & Sundström, L. (2005) Genetic patchwork of network-building wood ant populations. *Annales Zoologici Fennici* **42** 179-187

Papadopoulou, A., Anastasiou, I., & Vogler, A. P. (2010). Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution*, *27*(7), 1659-1672.

Pape, T. (2016) Taxonomy: Species can be named from photos. *Nature* **537** 307

Pauls, S. U., Nowak, C., Bálint, M., & Pfenninger, M. (2013). The impact of global climate change on genetic diversity within populations and species. *Molecular ecology*, *22*(4), 925-946.

Pearson, R. G., & Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global ecology and biogeography*, *12*(5), 361-371.

People's Trust for Endangered Species (2013) – results of the Traditional Orchard Survey from the website, accessed 07.01.2013. http://www.ptes.org/?page=431

Peterken, G. F. (1993). *Woodland conservation and management*. Springer Science & Business Media.

Petersen, F. T., Damgaard, J., & Meier, R. (2007) DNA taxonomy: how many DNA sequences are needed for solving a taxonomic problem? The case of two parapatric species of louse flies (Diptera: Hippoboscidae: *Ornithomya* Latreille, 1802). *Arthropod. Syst. Phylogeny*, *65*, 119-125.

Petit, R.J., Brewer, S., Bordács, S., Burg, K., Cheddadi, R., Coart, E., Cottrell, J., Csaikl, U.M., van Dam, B., Deans, J.D. *et al.* and Kremer, A. (2002) Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. *Forest ecology and management*, **156** 49-74

Petit, R. J., & Excoffier, L. (2009). Gene flow and species delimitation. *Trends in Ecology & evolution*, *24*(7), 386-393.

Philp, E. G. (2006) Vascular plants and the beetles associated with them. In: *A Coleopterist's Handbook (4th ed.)* (pp. 270-310) Cooter, J. & Barclay, M. V. L. eds., The Amateur Entomologists' Society

Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, *31*(2), 161-175.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological modelling*,*190*(3), 231-259.

Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. In *Proceedings of the twenty-first international conference on Machine learning* (p. 83). ACM.

Pierson, J., Luikart, G., & Schwartz, M. (2015). The application of genetic indicators in wild populations: potential and pitfalls for genetic monitoring. *Indicators and Surrogates of Biodiversity and Environmental Change*, 149.

Piggott, M. P. (2005). Effect of sample age and season of collection on the reliability of microsatellite genotyping of faecal DNA. *Wildlife Research*, *31*(5), 485-493.

Piggott, M. P., & Taylor, A. C. (2003). Remote collection of animal DNA and its applications in conservation management and understanding the population biology of rare and cryptic species. *Wildlife Research*, *30*(1), 1-13.

Pimm, S. L., Russell, G. J., Gittleman, J. L., & Brooks, T. M. (1995). The future of biodiversity. *Science-AAAS-Weekly Paper Edition*, *269*(5222), 347-349.

Pivotti, I., Agoglitta, R., Zunino, M., Piattella, E., Dellacasa, M., Corallini, C., & Mifsud, D. (2011). The Scarabaeoidea of the Maltese Islands (Central Mediterranean)(Coleoptera). *Bulletin of the Entomological Society of Malta*, 85-124.

Pizzo, A., Roggero, A., Palestrini, C., Cervella, P., Del Pero, M., & Rolando, A. (2006). Genetic and morphological differentiation patterns between sister species: the case of *Onthophagus taurus* and *Onthophagus illyricus* (Coleoptera, Scarabaeidae). *Biological Journal of the Linnean Society*, *89*(2), 197-211.

Popadin, K. Y., Nikolaev, S. I., Junier, T., Baranova, M., & Antonarakis, S. E. (2013). Purifying selection in mammalian mitochondrial protein-coding genes is highly effective and congruent with evolution of nuclear genes. *Molecular biology and evolution*, *30*(2), 347-355.

Polzin, T., & Daneshmand, S. V. (2003). On Steiner trees and minimum spanning trees in hypergraphs. *Operations Research Letters*, *31*(1), 12-20.

Post, R. J., Flook, P. K., & Millest, A. L. (1993). Methods for the preservation of insects for DNA studies. *Biochemical systematics and ecology*, *21*(1), 85-92.

Pray, L. A., Goodnight, C. J., Stevens, L., Schwartz, J. M., & Yan, G. (1996). The effect of population size on effective population size: an empirical study in the red flour beetle *Tribolium castaneum*. *Genetical Research*, *68*(02), 151-155.

Provan, J., & Maggs, C. A. (2011). Unique genetic variation at a species's rear edge is under threat from global climate change. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20110536.

Provan, J., & Bennett, K. D. (2008). Phylogeographic insights into cryptic glacial refugia. *Trends in Ecology & Evolution*, **23**(10), 564-571.

Pulgarín-R, P. C., & Burg, T. M. (2012). Genetic signals of demographic expansion in downy woodpecker (*Picoides pubescens*) after the last North American glacial maximum. *PloS one*, *7*(7), e40412.

Pulliam, H. R. (1988) Sources, sinks, and population regulation. *American Naturalist* **132** 652-661

Quinto, J., de los Ángeles Marcos-García, M., Díaz-Castelazo, C., Rico-Gray, V., Galante, E., & Micó, E. (2015). Association patterns in saproxylic insect networks in three Iberian Mediterranean woodlands and their resistance to microhabitat loss. *PloS one*, *10*(3), e0122141.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Raes, N., & ter Steege, H. (2007). A null-model for significance testing of presence-only species distribution models. *Ecography*, *30*(5), 727-736.

Ralls, K., Ballou, J. D., & Templeton, A. (1988). Estimates of lethal equivalents and the cost of inbreeding in mammals. *Conservation biology*,*2*(2), 185-193.

Rand, D. M., Dorfsman, M., & Kann, L. M. (1994). Neutral and non-neutral evolution of *Drosophila* mitochondrial DNA. *Genetics*, *138*(3), 741-756.

Rand, D. M., & Harrison, R. G. (1989). Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution*, 432-449.

Ranius, T. (2000) Minimum viable metapopulation size of a beetle, *Osmoderma eremita*, living in tree hollows. *Animal Conservation* **3** 37-43

Ranius, T., Aguado, L.O., Antonsson, K., Audisio, P., Ballerio, A., Carpaneto, G.M., Chobot, K., Gjurašin, B., Hanssen, O., Huijbregts, H. and Lakatos, F., et al. (2005). *Osmoderma eremita* (Coleoptera, Scarabaeidae, Cetoniinae) in Europe. *Europe*, *28*, pp.1-44.

Ranius, T. (2007) Extinction risks in metapopulations of a beetle inhabiting hollow trees predicted from time series. *Ecography* **30** 716-726

Ranius, T., & Hedin, J. (2001). The dispersal rate of a beetle, *Osmoderma eremita*, living in tree hollows. *Oecologia*, *126*(3), 363-370.

Ranius, T., Niklasson, M., & Berg, N. (2009a). Development of tree hollows in pedunculate oak (*Quercus robur*). *Forest Ecology and Management*, *257*(1), 303-310.

Ranius, T., Svensson, G. P., Berg, N., Niklasson, M. & Larsson, M. C. (2009b) The successional change of hollow oaks affects their suitability for an inhabiting beetle, *Osmoderma eremitia*. *Annales Zoologici Fennici* **46** 205-216

Ratcliffe, B. C. & Jameson, M. L. (2004) The revised classification for Scarabaeoidea: what the hell is going on? *Scarabs* **15** 3-10

Ratcliffe, B. C., & Warner, W. B. (2011). A revision of the genus Gymnetina Casey, 1915 (Coleoptera: Scarabaeidae: Cetoniinae: Gymnetini). *Insecta Mundi* 173 1-28

Raupach, M. J., Amann, R., Wheeler, Q. D., & Roos, C. (2016). The application of "-omics" technologies for the classification and identification of animals. *Organisms Diversity & Evolution*, *16*(1), 1-12.

Raupach, M. J., Astrin, J. J., Hannig, K., Peters, M. K., Stoeckle, M. Y., & Wägele, J. W. (2010). Molecular species identification of Central European ground beetles (Coleoptera: Carabidae) using nuclear rDNA expansion segments and DNA barcodes. *Frontiers in zoology*, *7*(1), 1.

Reboleira, A. S. P. S. & Enghoff, H. (2016) Taxonomics – next-generation taxonomists. *Organisms Diversity & Evolution* doi:10.1007

Reed, D. H., & Frankham, R. (2001). How closely correlated are molecular and quantitative measures of genetic variation? a meta-analysis. *Evolution*, *55*(6), 1095-1103.

Reed, D. H. & Frankham, R. (2003) Correlation between fitness and genetic diversity. *Conservation Biology* **17** 230-237

Reed, D. H., Lowe, E. H., Briscoe, D. A. & Frankham, R. (2003) Fitness and adaptation in a novel environment: effect of inbreeding, prior environment and lineage. *Evolution* **8** 1822-1828

Renault, D., Vernon, P., & Vannier, G. (2005). Critical thermal maximum and body water loss in first instar larvae of three Cetoniidae species (Coleoptera).*Journal of Thermal Biology*, *30*(8), 611-617.

Renault, D., Vernon, P., & Vannier, G. (2004). Comparing the freezing susceptibility of third-instar larvae of Gnorimus variabilis (Cetoniidae: Trichiinae) from three distant geographical regions. *Canadian journal of zoology*, *82*(6), 873-879.

Říčan, O., Zardoya, R., & Doadrio, I. (2008). Phylogenetic relationships of Middle American cichlids (Cichlidae, Heroini) based on combined evidence from nuclear genes, mtDNA, and morphology. *Molecular Phylogenetics and Evolution*, *49*(3), 941-957.

Richter, A., Weinhold, D., Robertson, G., Young, M., Edwards, T., Hnatiuk, S., & Osborne, W. (2013). More than an empty case: a non invasive technique for monitoring the Australian critically endangered golden sun moth, *Synemon plana* (Lepidoptera: Castniidae). *Journal of insect conservation*, *17*(3), 529-536.

Rink, M. & Sinsch, U. (2007) Radio-telemetric monitoring of dispersing stag beetles: implications for conservation. *Journal of Zoology* **272** 235-243

Róbert, E. (2004). szarvaskő lemezescsápú (Coleoptera: scarabaeoidae) faunája. *Folia Historico Naturalia Musei Matraensis*, *28*, 135-140.

Rodrigues, A. S., Pilgrim, J. D., Lamoreux, J. F., Hoffmann, M., & Brooks, T. M. (2006). The value of the IUCN Red List for conservation. *Trends in ecology & evolution*, *21*(2), 71-76.

Roe, H. M., Coope, G. R., Devoy, R. J., Harrison, C. J., Penkman, K. E., Preece, R. C., & Schreve, D. C. (2009). Differentiation of MIS 9 and MIS 11 in the continental record: vegetational, faunal, aminostratigraphic and sea-level evidence from coastal sites in Essex, UK. *Quaternary Science Reviews*, *28*(23), 2342-2373.

Rogers, A. R., & Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Molecular biology and evolution*, *9*(3), 552-569.

Roggero, A., Giachino, P. M., & Palestrini, C. (2013). A new cryptic ground beetle species from the Alps characterised via geometric morphometrics. *Contributions to Zoology*, *82*(4).

Rohlf, F. J., & Slice, D. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Biology*, *39*(1), 40-59.

Rohlf, F. J. (2015). The tps series of software. *Hystrix, the Italian Journal of Mammalogy*, *26*(1), 9-12.

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572-1574.

Rosa, J. J. D. L. (2005). Aportaciones la conocimiento de la corología de algunos escarabeidos florícolas íbero-baleares (Coleoptera: Scarabaeidae).*Boletín de la SEA*, (37), 313-314.

Rosenberg, N. A. & Nordborg, M. (2002) Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nature Reviews Genetics* **3** 380-390

Rosvold, J., Andersen, R., Linnell, J. D., & Hufthammer, A. K. (2013). Cervids in a dynamic northern landscape: Holocene changes in the relative abundance of moose and red deer at the limits of their distributions. *The Holocene*, 0959683613483625.

Rougerie, R., Lopez-Vaamonde, C., Barnouin, T., Delnatte, J., Moulin, N., Noblecourt, T., Parmain, G., Soldati, F. & Bouget, C. (2015). PASSIFOR: A reference library of DNA barcodes for French saproxylic beetles (Insecta, Coleoptera). *Biodiversity data journal*, (3).

Rowe, H. C., Renaut, S. & Guiggisberg, A. (2011) RAD in the realm of next-generation sequencing technologies. *Molecular Ecology* **20** 3499-3502

Roy, M. S., Geffen, E., Smith, D., Ostrander, E. A., & Wayne, R. K. (1994). Patterns of differentiation and hybridization in North American wolflike canids, revealed by analysis of microsatellite loci. *Molecular Biology and Evolution*, *11*(4), 553-570.

Rozner, I., & Rozner, G. (2009). Data to the Lamellicornia fauna of the Republic of Macedonia (Coleoptera: Lamellicornia). *Natura Somogyiensis*,*15*, 57-68.

Rukke, B. A. (2000) Effects of habitat fragmentation: increased isolation and reduced habitat size reduces the incidence of dead wood fungi beetles in a fragmented forest landscape. *Ecography* **23** 492-502

Sabatinelli, G. (1977). Note su alcuni Lucanidae e Scarabaeoidea floricoli di Turchia, con descrizione di due nuove specie (Coleoptera). *Fragmenta entomologica*, *12*(2), 71-96.

Sabatinelli, G. (2008) *Trichius orientalis. The Scarabs of the Levant*, http://www.glaphyridae.com/Cetoniinae/Tr_orientalis.html

Sætre, G. P., Borge, T., Lindell, J., Moum, T., Primmer, C. R., Sheldon, B. C., Haavie, J., Johnsen A. & Ellegren, H. (2001). Speciation, introgressive hybridization and nonlinear rate of molecular evolution in flycatchers. *Molecular Ecology*, *10*(3), 737-749.

Saito, S., Tamura, K., & Aotsuka, T. (2005). Replication origin of mitochondrial DNA in insects. *Genetics*, *171*(4), 1695-1705.

Sallé, A., Nageleisen, L. M., & Lieutier, F. (2014). Bark and wood boring insects involved in oak declines in Europe: Current knowledge and future prospects in a context of climate change. *Forest Ecology and Management*, *328*, 79-93.

Samways, M. J. (2010) Conserving Invertebrates: How many can be saved, and how? *In: Trade-offs in Conservation*, Wiley-Blackwell, Chichester

Sánchez-Galván, I. R., Quinto, J., Micó, E., Galante, E., & Marcos-García, M. A. (2014). Facilitation among saproxylic insects inhabiting tree hollows in a Mediterranean forest: the case of cetonids (Coleoptera: Cetoniidae) and syrphids (Diptera: Syrphidae). *Environmental entomology*,*43*(2), 336-343.

Santibáñez-López, C. E., Francke, O. F., & Prendini, L. (2014). Phylogeny of the North American scorpion genus *Diplocentrus* Peters, 1861 (Scorpiones: Diplocentridae) based on morphology, nuclear and mitochondrial DNA. *Arthropod Syst. Phylo*, *72*, 257-279.

Scheidegger, C., Bilovitz, P. O., Werth, S., Widmer, I., & Mayrhofer, H. (2012). Hitchhiking with forests: population genetics of the epiphytic lichen *Lobaria pulmonaria* in primeval and managed forests in southeastern Europe. *Ecology and evolution*, *2*(9), 2223-2240.

Schenke, J. M. L. (2010) The autecology of the Noble Chafer (*Gnorimus nobilis* L.) within Worcestershire, England. Master of Science thesis, Department of Environmental and Geographical Sciences, The Manchester Metropolitan University.

Schmitt, T. (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Frontiers in zoology*, *4*(1), 1.

Schmitt, T., & Varga, Z. (2012). Extra-Mediterranean refugia: The rule and not the exception?. *Frontiers in Zoology*, *9*(1), 1.

Schmuki, C., Vorburger, C., Runciman, D., Maceachern, S. & Sunnucks, P. (2006) When log-dwellers meet loggers: impacts of forest fragmentation on two endemic log-dwelling beetles in southeastern Australia. *Molecular Ecology* **15** 1481-1492

Schönrogge, K., Barr, B., Wardlaw, J. C., Napper, E., Gardner, M. G., Breen, J., Elmes, G. W. & Thomas, J. A. (2002). When rare species become endangered: cryptic speciation in myrmecophilous hoverflies. *Biological Journal of the Linnean Society*, *75*(3), 291-300.

Schuldt, A., & Assmann, T. (2010). Invertebrate diversity and national responsibility for species conservation across Europe–a multi-taxon approach. *Biological Conservation*, *143*(11), 2747-2756.

Schwartz, M. K., Luikart, G., & Waples, R. S. (2007). Genetic monitoring as a promising tool for conservation and management. *Trends in ecology & evolution*, *22*(1), 25-33.

Scriven, J. J., Woodall, L. C. & Goulson, D. (2013) Nondestructive DNA sampling from bumblebee faeces. *Molecular Ecology Resources* **13** 225-229

Seidl, R., Schelhaas, M. J., & Lexer, M. J. (2011). Unraveling the drivers of intensifying forest disturbance regimes in Europe. *Global Change Biology*, *17*(9), 2842-2852.

Salander, R. K. (1983) Evolutionary consequences of inbreeding. *In: Genetics and Conservation: A Reference for Managing Wild Animals and Plant Population*. Benjamin-Cummings, London

Sebek, P., Altman, J., Platek, M., & Cizek, L. (2013). Is active management the key to the conservation of saproxylic biodiversity? Pollarding promotes the formation of tree hollows. *PloS one*,*8*(3), e60456.

Seeber, J., Rief, A., Seeber, G. U., Meyer, E., & Traugott, M. (2010). Molecular identification of detritivorous soil invertebrates from their faecal pellets. *Soil Biology and Biochemistry*, *42*(8), 1263-1267.

Selkoe, K. A. & Toonen, R. J. (2006) Microsatellites for ecologists: a practical guide to using and evaluating microsatellite markers. *Ecology Letters* **9** 615-629

Şenyüz, Y., & Şahin, Y. (2009). Faunistic studies on Cetoniinae, Dynastinae, Melolonthinae, Rutelinae (Coleoptera: Scarabaeidae) Geotrupinae (Geotrupidae) of Kütahya province, Turkey. *Munis Entomology & Zoology*,*4*(2), 536-541.

Shabalin, S. A., & Bezborodov, V. G. (2009). Larvae of scarab beetles of the subfamily Trichiinae (Coleoptera, Scarabaeidae) from the Russian Far East. *Entomological review*, *89*(9), 1131-1136.

Shaw, K. L. (2002). Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: what mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences*, *99*(25), 16122-16127.

Sheedy, E. M., Van de Wouw, A. P., Howlett, B. J., & May, T. W. (2015). Population genetic structure of the ectomycorrhizal fungus *Laccaria* sp. A resembles that of its host tree *Nothofagus cunninghamii*. *Fungal Ecology*, *13*, 23-32.

Sheffield, N. C., Song, H., Cameron, S. L., & Whiting, M. F. (2008). A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. *Molecular biology and evolution*, *25*(11), 2499-2509.

Shen, J. M., Hu, L. M., Bin, S. Y., & Lin, J. T. (2011). Cloning and expression profiling of an olfactory receptor gene in *Bactrocera cucurbitae* (Coquillett) (Diptera: Tephritidae). *Acta Entomologica Sinica*, *54*(3), 265-271.

Shennan, I., Bradley, S., Milne, G., Brooks, A., Bassett, S. & Hamilton, S. (2006) Relative sea-level changes, glacial isostatic modelling and ice-sheet reconstructions from the British Isles since the Last Glacial Maximum. *Journal of Quaternary Science* **21** 585-599

Siitonen, J. (2001). Forest management, coarse woody debris and saproxylic organisms: Fennoscandian boreal forests as an example. *Ecological bulletins*, 11-41.

Simon, C., Buckley, T. R., Frati, F., Stewart, J. B., & Beckenbach, A. T. (2006). Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction

primers for animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, 545-579.

Šipek, P., Gill, B. D. & Grebennikov, V. V. (2009) Afromontane *Coelocorynus* (Coleoptera: Scarabaeidae: Cetoniinae): Larval descriptions, biological notes and phylogenetic placement. *European Journal of Entomology* **106** 95-106

Šípek, P., Fabrizi, S., Eberle, J., & Ahrens, D. (2016). A molecular phylogeny of rose chafers (Coleoptera: Scarabaeidae: Cetoniinae) reveals a complex and concerted morphological evolution related to their flight mode. *Molecular phylogenetics and evolution*, *101*, 163-175.

Seppä, H., Bjune, A. E., Telford, R. J., Birks, H. J. B., & Veski, S. (2009). Last nine-thousand years of temperature variability in Northern Europe.*Climate of the Past*, *5*(3), 523-535.

Seppä, H., Schurgers, G., Miller, P. A., Bjune, A. E., Giesecke, T., Kühl, N., Renssen, H. & Salonen, J. S. (2015). Trees tracking a warmer climate: The Holocene range shift of hazel (*Corylus avellana*) in northern Europe. *The Holocene*, *25*(1), 53-63.

Sint, D., Thurner, I., Kaufmann, R., & Traugott, M. (2015). Sparing spiders: faeces as a non-invasive source of DNA. *Frontiers in Zoology* ***12*** 3

Sites, J. W., & Marshall, J. C. (2003). Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology & Evolution*, *18*(9), 462-470.

Slatkin, M. (1977). Gene flow and genetic drift in a species subject to frequent local extinctions. *Theoretical population biology*, *12*(3), 253-262.

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*,*236*(4803), 787-792.

Slatkin, M. (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution*, 264-279.

Slatyer, R. A., Hirst, M., & Sexton, J. P. (2013). Niche breadth predicts geographical range size: a general ecological pattern. *Ecology letters*, *16*(8), 1104-1114.

Smith, M. N. (2003) Saproxylic beetles in Britain, an overview of the status and distribution of four Biodiversity Action Plan species. *Proceedings of the second pan-European conference on Saproxylic Beetles*. People's Trust for Endangered Species.

Smith, A. B. T., Hawks, D. C. & Heraty, J. M. (2006) An overview of the classification and evolution of the major scarab beetle clades (Coleoptera: Scarabaeoidea) based on preliminary molecular analyses. *Papers in Entomology* Paper 121

Snäll, N., Huoponen, K., Savontaus, M. L., & Ruohomäki, K. (2002). Tandem repeats and length variation in the mitochondrial DNA control region of Epirrita autumnata (Lepidoptera: Geometridae).*Genome*, *45*(5), 855-861.

Sokal, R. R. & Crovello, T. J. (1970) The biological species concept: a critical evaluation. *The American Naturalist* 104 127-153

Solano, E., Mancini, E., Ciucci, P., Mason, F., Audisio, P., & Antonini, G. (2013). The EU protected taxon Morimus funereus Mulsant, 1862 (Coleoptera: Cerambycidae) and its western Palaearctic allies: systematics and conservation outcomes. *Conservation Genetics*, *14*(3), 683-694.

Solano, E., Thomaes, A., Cox, K., Carpaneto, G.M., Cortellessa, S., Baviera, C., Bartolozzi, L., Zilioli, M., Casiraghi, M., Audisio, P. and Antonini, G. (2016) When morphological identification meets genetic data: the case of *Lucanus cervus* and *L. tetraodon* (Coleoptera, Lucanidae). *Journal of Zoological Systematics and Evolutionary Research*. DOI: 10.1111/jzs.12124

Somero, G. N. (2010). The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine 'winners' and 'losers'. *Journal of Experimental Biology*, *213*(6), 912-920.

Sommer, R. S., Fritz, U. W. E., Seppä, H., Ekström, J., Persson, A., & Liljegren, R. (2011). When the pond turtle followed the reindeer: effect of the last extreme global warming event on the timing of faunal change in Northern Europe. *Global Change Biology*, *17*(6), 2049-2053.

Song H, Buhay JE, Whiting MF, Crandall KA. 2008. Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences of the United States of America 105*: 13486–13491.

Sota, T., & Vogler, A. P. (2001). Incongruence of mitochondrial and nuclear gene trees in the carabid beetles *Ohomopterus*. *Systematic biology*,*50*(1), 39-59.

Speight, M. C. (1989). *Saproxylic invertebrates and their conservation*. Council of Europe.

Stefanelli, M. S., Della Rocca, F., & Bogliani, G. (2014). Saproxylic beetles of the Po plain woodlands, Italy. *Biodiversity data journal*, (2).

Stephens, M., & Scheet, P. (2005). Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *The American Journal of Human Genetics*, *76*(3), 449-462.

Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics*, *68*(4), 978-989.

Stork, N. E., McBroom, J., Gely, C., & Hamilton, A. J. (2015). New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *Proceedings of the National Academy of Sciences*,*112*(24), 7519-7523.

Strange, J. P., Knoblett, J., & Griswold, T. (2009). DNA amplification from pin-mounted bumble bees (Bombus) in a museum collection: effects of fragment size and specimen age on successful PCR. *Apidologie*, *40*(2), 134-139.

Strangi, A., Sabbatini Peverieri, G. & Roversi, P. F. (2013) Managing outbreaks of the citrus long-horned beetle *Anoplophora chinensis* (Forster) in Europe: molecular diagnosis of plant infestation. *Pest Management Science* **69** 627-634

Su, X. Z., Wu, Y., Sifri, C. D., & Wellems, T. E. (1996). Reduced extension temperatures required for PCR amplification of extremely A+ T-rich DNA. *Nucleic acids research*, *24*(8), 1574-1575.

Summers, M. M., Al-Hakim, I. I., & Rouse, G. W. (2014). Turbo-taxonomy: 21 new species of *Myzostomida* (Annelida). *Zootaxa*, *3873*(4), 301-344.

Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology & Evolution*,*15*(5), 199-203.

Svenning, J. C., Gravel, D., Holt, R. D., Schurr, F. M., Thuiller, W., Münkemüller, T., Schiffere, K. H., Dullinger, S., Edwards Jr., T. C., Hickler, T., Higgings, S. I., Nabel, J. E. M. S., Pagel, J. & Higgins, S. I. (2014). The influence of interspecific interactions on species range expansion rates. *Ecography* *37*(12), 1198-1209.

Svenning, J. C., Normand, S., & Kageyama, M. (2008). Glacial refugia of temperate trees in Europe: insights from species distribution modelling. *Journal of Ecology*, *96*(6), 1117-1127.

Svensson, G. P., Sahlin, U., Brage, B. & Larsson, M. C. (2011) Should I stay or should I go? Modelling dispersal strategies in saproxylic insects based on pheromone capture and radio telemetry: a case study on the threatened hermit beetle *Osmoderma eremita*. *Biodiversity and Conservation* **20** 2883-2902

Sverdrup-Thygeson, A., Skarpaas, O., & Ødegaard, F. (2010). Hollow oaks and beetle conservation: the significance of the surroundings. *Biodiversity and conservation*, *19*(3), 837-852.

Swift, M. J. (1977). The ecology of wood decomposition. *Science Progress (1933-)*, 175-199.

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular ecology*, *21*(8), 2045-2050.

Taberlet, P., Fumagalli, L., Wust-Saucy, A. G., & Cosson, J. F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Molecular ecology*, *7*(4), 453-464.

Taberlet, P., & Luikart, G. (1999). Non-invasive genetic sampling and individual identification. *Biological journal of the Linnean society*, *68*(1-2), 41-55.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585-595.

Tamura K., Stecher G., Peterson D., Filipski A., and Kumar S. (**2013**). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*30: 2725-2729.

Tanahashi, M., & Hawes, C. J. (2016). The presence of a mycangium in European *Sinodendron cylindricum* (Coleoptera: Lucanidae) and the associated yeast symbionts. *Journal of Insect Science*, *16*(1), 76.

Tanahashi, M., Kubota, K., Matsushita, N., & Togashi, K. (2010). Discovery of mycangia and the associated xylose-fermenting yeasts in stag beetles (Coleoptera: Lucanidae). *Naturwissenschaften*, *97*(3), 311-317.

Tauzin, P. (2004a.) Iconographie de quelques espèces du genre *Aleurostictus* Kirby, 1827 et nouvelle precision sue leur distribution. (Coleoptera, Cetoniidae, Trichiinae, Trichiini). *Cetoniimania* **2** 77-94

Tauzin, P. (2004b.). Quelques localités connues pour les espèces françaises d'*Aleurostictus* (Coleoptera, Cetoniidae, Trichiinae, Trichiini). *Cetoniimania, volume1*, 33-52.

Tauzin, P. (2005a.). Ethologie et distribution de *Cetonischema aeruginosa* Drury 1770 en France (Coleoptera, Cetoniidae, Cetoniinae, Cetoniini).*Cetoniimania, 1*(930), 2008.

Tauzin, P. (2005b.). Ethology and distribution of the "Hermit beetle" in France (Coleoptera, Cetoniidae, Trichiinae, Osmodermatini). *Cetoniimania*, *4*, 131-153.

Telnov, D., Gailis, J., Kalniņš, M. , Napolov, A., Piterāns, U., Vilks, K., & Whitehead, P. F. (2005). Contributions to the knowledge of Latvian Coleoptera. 4. *Latvijas entomologs*, *42*, 18-47.

Templeton, A. R., Robertson, R. J., Brisson, J., & Strasburg, J. (2001). Disrupting evolutionary processes: the effect of habitat fragmentation on collared lizards in the Missouri Ozarks. *Proceedings of the National Academy of Sciences*, *98*(10), 5426-5432.

Templeton, A. R., Shaw, K., Routman, E. & Davis, S. K. (1990) The genetic consequences of habitat fragmentation. *Annals of the Missouri Botanical Garden***77** 13-27

Tezcan, S. & Pahlivan, E. (2001) Evaluation of the Lucanoidea and Scarabaeoidea (Coleoptera) fauna of ecological cherry orchards in İzmir and Manisa provinces of Turkey. Ege Üniversitesi Ziraat Fakültesi Dergisi **38** 31-37

Thomaes, A., Kervyn, T., Beck, O., & Cammaerts, R. (2008). Distribution of *Lucanus cervus* (Coleoptera: Lucanidae) in Belgium: surviving in a changing landscape. *Revue d'écologie*, 147-152.

Thomaes, A., Crevecoeur, L., & Wijnants, M. (2015). Tree cavity beetles in Haspengouw and pays De herve: *Crepidophorus mutilatus* (Elateridae) new for the Belgian fauna and rediscovery of *Gnorimus variabilis* (Cetoniidae) (Insecta: Coleoptera). *Bull Soc Roy Belg Entomol*, *151*, 40-51.

Thomas, C. D. (2000) Dispersal and extinction in fragmented landscapes. *Proceedings of the Royal Society B* **267** 139-145

Thomas, C. D., Franco, A. M., & Hill, J. K. (2006). Range retractions and extinction in the face of climate warming. *Trends in Ecology & Evolution*,*21*(8), 415-416.

Thomsen, P. F., Elias, S., Gilbert, M. T. P., Haile, J., Munch, K., Kuzmina, S., Froese, D. G., Sher, A., Holdaway, R. N. & Willerslev, E. (2009). Non-destructive sampling of ancient insect DNA. *PLoS One*, *4*(4), e5048.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA–an emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4-18.

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*(22), 4673-4680.

Toews, D. P., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*,*21*(16), 3907-3930.

Tournant, P., Joseph, L., Goka, K., & Courchamp, F. (2012). The rarity and overexploitation paradox: stag beetle collections in Japan. *Biodiversity and Conservation*, *21*(6), 1425-1440.

Trizzino, M., Bisi, F., Morelli, C. E., Preatoni, D. G., Wauters, L. A. & Martinoli, A. (2013) Spatial niche partitioning of two saproxylic sibling species (Coleoptera, Cetoniidae, genus *Gnorimus*). *Insect Conservation and Diversity* **7** 223-231

Tscharntke, T., Steffan-Dewenter, I., Kruess, A., & Thies, C. (2002). Characteristics of insect populations on habitat fragments: a mini review. *Ecological research*, *17*(2), 229-239.

Tyler, M. (2008) British Oaks; A Concise Guide. Published by The Crowood Press Ltd, Ramsbury, Wiltshire.

Tzedakis, P. C., Emerson, B. C., & Hewitt, G. M. (2013). Cryptic or mystic? Glacial tree refugia in northern Europe. *Trends in Ecology & Evolution*, **28**(12), 696-704.

UK Biodiversity Action Plan (2008) UK Biodiversity Action Plan Priority Habitat Descriptions: Traditional Orchards.

Valière, N., Fumagalli, L., Gielly, L., Miquel, C., Lequette, B., Poulle, M. L., Weber, J.-M., Arlettaz, R. & Taberlet, P. (2003). Long-distance wolf recolonization of France and Switzerland inferred from non-invasive genetic sampling over a period of 10 years. *Animal Conservation*, *6*(1), 83-92.

Van Houdt, J. K. J., Breman, F. C., Virgilio, M., & De Meyer, M. (2010). Recovering full DNA barcodes from natural history collections of Tephritid fruitflies (Tephritidae, Diptera) using mini barcodes. *Molecular ecology resources*, *10*(3), 459-465.

van Oppen, M. J., Rico, C., Turner, G. F., & Hewitt, G. M. (2000). Extensive homoplasy, nonstepwise mutations, and shared ancestral polymorphism at a complex microsatellite locus in Lake Malawi cichlids. *Molecular Biology and Evolution*, *17*(4), 489-498.

Vera, F. W. M. (2000). *Grazing ecology and forest history*. CABI publishing.

Verma, M. (2016). Modelling the effect of rarity value on the exploitation of a wildlife species subjected to the Allee effect.. *Natural Resource Modeling*, *29*(3), 470-494.

Vernon, P., & Vannier, G. (2001). Freezing susceptibility and freezing tolerance in Palaearctic Cetoniidae (Coleoptera). *Canadian journal of zoology*, *79*(1), 67-74.

Vernon, P., & Vannier, G. (2002). Evolution of freezing susceptibility and freezing tolerance in terrestrial arthropods. *Comptes rendus biologies*,*325*(12), 1185-1190.

Vieira, F. G., Sánchez-Gracia, A., & Rozas, J. (2007). Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. *Genome biology*, *8*(11), 1.

Vila, M., Auger-Rozenberg, M. A., Goussard, F. & Lopez-Vaamonde, C. (2009) Effect of non-lethal sampling on life-history traits of the protected moth *Graellsia isabelae* (Lepidoptera: Saturniidae). *Ecological Entomology* 34 356-362

Vila, M., & Björklund, M. (2004). The utility of the neglected mitochondrial control region for evolutionary studies in Lepidoptera (Insecta).*Journal of Molecular Evolution*, *58*(3), 280-290.

Vila, M., Lundhagen, A. C., Thuman, K. A., Stone, J. R. & Björklund, M. (2006) A new conservation unit in the butterfly *Erebia triaria* (Nymphalidae) as revealed by nuclear and mitochondrial markers. *Annales Zoologici Fennici* 43 72-79

Vila, M., Vidal-Romaní, J. R., & Björklund, M. (2005). The importance of time scale and multiple refugia: incipient speciation and admixture of lineages in the butterfly *Erebia triaria*. (Nymphalidae). *Molecular phylogenetics and evolution*, *36*(2), 249-260.

Villani, M. G., Allee, L. L., Diaz, A., & Robbins, P. S. (1999). Adaptive strategies of edaphic arthropods. *Annual review of entomology*, *44*(1), 233-256.

Vogler, A. P., & Monaghan, M. T. (2007). Recent advances in DNA taxonomy. *Journal of Zoological Systematics and Evolutionary Research*, *45*(1), 1-10.

Vollrath, F. (1999) Biology of spider silk. *International Journal of Biological Macromolecules* **24** 81-88

Vollrath, F. (2000) Strength and structure of spiders' silks. *Reviews in Molecular Biotechnology* **74** 67-83

Voolma, K. & Randveer, T. (2003) *Gnorimus variabilis* (Linnaeus, 1758) (Coleoptera; Scarabaeidae), a new record from Estonia. *Baltic Journal of Coleopterology* **3** 15-52

Vosshall, L. B., & Hansson, B. S. (2011). A unified nomenclature system for the insect olfactory coreceptor. *Chemical senses*, bjr022.

Vrezec, A. (2007). Status brazdarja (Rhysodes sulcatus) v Sloveniji (Coleoptera: Rhysodidae): dosedanje poznavanje in raziskovalne perspektive. *Acta entomologica slovenica*, *15*(1), 51-56.

Wade, M. J., Shuster, S. M. & Stevens, L. (1996) Inbreeding: its effect on response to selection for pupal weight and the heritable variance in fitness in the flour beetle, *Tribolium castaneum*. *Evolution* **50** 723-733

Wagener, W. W., & Davidson, R. W. (1954). Heart rots in living trees. *The Botanical Review*, *20*(2), 61-134.

Waits, L. P., & Paetkau, D. (2005). Noninvasive genetic sampling tools for wildlife biologists: a review of applications and recommendations for accurate data collection. *Journal of Wildlife Management*, *69*(4), 1419-1433.

Wakeley, J. (2008). An introduction to coalescent theory. *Roberts & Co*, Colarado

Wakeley, J., & Aliacar, N. (2001). Gene genealogies in a metapopulation. *Genetics*, *159*(2), 893-905.

Walter, K. S., Brown, J. E., & Powell, J. R. (2014). Microhabitat Partitioning of *Aedes simpsoni* (Diptera: Culicidae). *Journal of medical entomology*,*51*(3), 596-604.

Wandeler, P., Hoeck, P. E., & Keller, L. F. (2007). Back to the future: museum specimens in population genetics. *Trends in Ecology & Evolution*, *22*(12), 634-642.

Wang, J. (2005). Estimation of effective population sizes from data on genetic markers. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1459), 1395-1409.

Wang, S., Lorenzen, M. D., Beeman, R. W., & Brown, S. J. (2008). Analysis of repetitive DNA distribution patterns in the *Tribolium castaneum* genome. *Genome biology*, *9*(3), 1

Waples, R. S. & Do, C. (2010) Linkage disequilibriums estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution. *Evolutionary Applications* **3** 244-262

Warren, D. L., Glor, R. E., & Turelli, M. (2010). ENMTools: a toolbox for comparative studies of environmental niche models. *Ecography*, *33*(3), 607-611.

Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. *Ecological Applications*, *21*(2), 335-342.

Warren, M. S., & Key, R. S. (1991). Woodlands: past, present and potential for insects. *The conservation of insects and their habitats*, 155-211.

Warren, M. S., Hill, J. K., Thomas, J. A., Asher, J., Fox, R., Huntley, B., Roy, D. B., Telfer, M. G., Jeffcoate, S., Harding. P., Jeffcoate, G., Willis, S. G., Greatorex-Davies, J. N., Moss, D & Thomas, C. D. (2001). Rapid responses of British butterflies to opposing forces of climate and habitat change. *Nature,414*(6859), 65-69.

Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takcmura, T., Okajima, H., Nozawa, T., Kawase, H., Abe, M., Yokohata, T., Ise, T., Sato, H., Kato, E., Takata, K., Emori, S. & Ise, T. (2011). MIROC-ESM 2010: model description and basic results of CMIP 5-20 c 3 m experiments. *Geoscientific Model Development*,*4*(4), 845-872.

Watts, P. C., Thompson, D. J., Allen, K. A., & Kemp, S. J. (2007). How useful is DNA extracted from the legs of archived insects for microsatellite-based population genetic analyses? *Journal of Insect Conservation*, *11*(2), 195-198.

Watts, P. C., Thompson, D. J., Daguet, C., & Kemp, S. J. (2005). Exuviae as a reliable source of DNA for population-genetic analysis of odonates. *Odonatologica*, *34*(2), 183-187.

Waugh, J. (2007). DNA barcoding in animal species: progress, potential and pitfalls. *BioEssays*, *29*(2), 188-197.

Whitehead, P. F. (2003) The noble chafer *Aleurostictus nobilis* (L., 1758) (Col., Scarabaeidae) in Britain. *Proceedings of the second pan-European conference on Saproxylic Beetles*. People's Trust for Endangered Species.

Whiteley, A. R., Spruell, P. & Allandorf, F. W. (2004) Ecological and life history characteristics predict population genetic divergence of two salmonids in the same landscape. *Molecular Ecology* **13** 3675-3688

Whitlock, M. C. (2011). *GST* and *D* do not replace *FST*. *Molecular ecology*,*20*(6), 1083-1091.

Wickman, P. O., & Karlsson, B. (1989). Abdomen size, body size and the reproductive effort of insects. *Oikos*, 209-214.

Wielstra, B., Crnobrnja-Isailović, J., Litvinchuk, S. N., Reijnen, B. T., Skidmore, A. K., Sotiropoulos, K., Toxopeus, A. G., Tzankov, N., Vukov, T. & Arntzen, J. W. (2013). Tracing glacial refugia of *Triturus* newts based on mitochondrial DNA phylogeography and species distribution modeling. *Frontiers in zoology*, *10*(1), 1.

Wild, A. L., & Maddison, D. R. (2008). Evaluating nuclear protein-coding genes for phylogenetic utility in beetles. *Molecular Phylogenetics and Evolution*, *48*(3), 877-891.

Wildt, D. E., Bush, M., Goodrowe, K. L., & Packer, C. (1987). Reproductive and genetic consequences of founding isolated lion populations. *Nature*, *329*, 24.

Williams, B. L. (2002). Conservation genetics, extinction, and taxonomic status: a case history of the regal fritillary. *Conservation Biology*, *16*(1), 148-157.

Williams, H. W., Cross, D. E., Crump, H. L., Drost, C. J., & Thomas, C. J. (2015). Climate suitability for European ticks: assessing species distribution models against null models and projection under AR5 climate. *Parasites & vectors*,*8*(1), 1.

Winter, S., & Möller, G. C. (2008). Microhabitats in lowland beech forests as monitoring tool for nature conservation. *Forest Ecology and Management*,*255*(3), 1251-1261.

Wood, S.O., (2006) Wood and Tree Fungi: Biology, Damage, Protection, and Use. Springer, Berlin

Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97-159.

Wright, S. (1943) Isolation by distance. *Genetics* **28** 114-38

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

Xu, J., & Fonseca, D. M. (2011). One-way sequencing of multiple amplicons from tandem repetitive mitochondrial DNA control region. *Mitochondrial DNA*, *22*(5-6), 155-158.

Xu, C. C., Yen, I. J., Bowman, D., & Turner, C. R. (2015). Spider web DNA: a new spin on noninvasive genetics of predator and prey. *PloS one*, *10*(11), e0142503.

Yang, Y., Krieger, J., Zhang, L., & Breer, H. (2012). The olfactory co-receptor Orco from the migratory locust (*Locusta migratoria*) and the desert locust (*Schistocerca gregaria*): identification and expression pattern. *Int J Biol Sci*, *8*(2), 159-170.

Yen, A. L. (1995). Australian spiders: An opportunity for conservation. *Records of the Western Australian Museum Supplement*, *52*, 39-47.

Young, R. M. (1989). Euchirinae (Coleoptera: Scarabaeidae) of the world: distribution and taxonomy. *The Coleopterists' Bulletin*, 205-236.

Zabransky, P. (1998). Der Lainzer Tiergarten als Refugium für gefährdete xylobionte Käfer (Coleoptera). *Z. Arb. Gem. Öst. Ent*, *50*(3/4), 95-117.

Zapponi, L., Cini, A., Bardiani, M., Hardersen, S., Maura, M., Maurizi, E., Redolfi De Zan, L., Audisio, P., Bologna, M. A., Roversi, P. F., Sabbatini Peverieri, G., Mason, F. & Campanaro, A. (2016). Citizen science data as an efficient tool for mapping protected saproxylic beetles. *Biological Conservation*.

Zeh, D. W., Zeh, J. A., & Bonilla, M. M. (2003). Phylogeography of the giant harlequin beetle (*Acrocinus longimanus*). *Journal of biogeography*, *30*(5), 747-753.

Zelditch, M. L., Swiderski, D. L., & Sheets, H. D. (2012). *Geometric morphometrics for biologists: a primer*. Academic Press.

Zhan, A., Hulak, M., Sylvester, F., Huang, X., Adebayo, A. A., Abbott, C. L., Adamowicz, S. J., Heath, D. D., Cristescu, M. E. & MacIsaac, H. J. (2013). High sensitivity of 454 pyrosequencing for detection of rare species in aquatic communities. *Methods in Ecology and Evolution*,*4*(6), 558-565.

Zhang, Y., Gu, S., Li, C., Sang, M., Wu, W., Yun, X., Hu, X., & Li, B. (2014). Identification and characterization of novel ER-based hsp90 gene in the red flour beetle, Tribolium castaneum. *Cell Stress and Chaperones*, *19*(5), 623-633.

Zhang, D. X., & Hewitt, G. M. (1997). Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochemical Systematics and Ecology*,*25*(2), 99-120.

Zhang, J., Rosenberg, H. F., & Nei, M. (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proceedings of the National Academy of Sciences*, *95*(7), 3708-3713.

Zhao, H., Gao, P., Zhang, C., Ma, W., & Jiang, Y. (2013). Molecular identification and expressive characterization of an olfactory co-receptor gene in the Asian honeybee, *Apis cerana cerana*. *Journal of Insect Science*, *13*(1), 80.

Zinetti, F., Dapporto, L., Vanni, S., Magrini, P., Bartolozzi, L., Chelazzi, G., & Ciofi, C. (2013). Application of molecular genetics and geometric morphometrics to taxonomy and conservation of cave beetles in central Italy. *Journal of insect conservation*, *17*(5), 921-932.

Zink, R. M. (2004) The role of subspecies in obscuring avian biological diversity and misleading conservation policy. *Proceedings of the Royal Society B* **271** 561-564

Zink, R. M., & Barrowclough, G. F. (2008). Mitochondrial DNA under siege in avian phylogeography. *Molecular ecology*, *17*(9), 2107-2121.

Supplementary Material

Chapter 3: *Trichius* phylogeography and taxonomy

Supplementary Table 1: the Trichius samples used in the study. The sex codes for Male (M) and Female (F), whilst both the Cytochrome Oxidase I (COI ID) and assumed morphological ID (Morph ID) code for *T. fasciatus* (F), *T. gallicus* (G) and *T. sexualis* (S). Samples with a numeric code and no COI sequence are museum samples

Locality	**Latitude**	**Longitude**	**Code**	**Sex**	**COI ID**	**Morph ID**	**Congruence?**
Adamello, Lago di Bissina, Italy	46.08	10.50	IA1	F	S	G	**No**
Bohemia, Krkonose, Vit Picek 1gt.	50.75	15.62	BK1	M	F	S	**No**
Bohemia, Vodnany	49.13	14.14	VB1	M	F	F	Yes
Bulgaria, Melnik	41.53	23.40	BM2	M	F	F	Yes
Bulgaria, Melnik	41.53	23.40	BM3	M	F	F	Yes
Bulgaria, Melnik	41.53	23.40	BM1	M	F	S	**No**
Croatia, Gospie, Velebit Mtn. Samoylenko	44.53	15.21	CG6	F	F	S	**No**
Croatia, Gospie, Velebit Mtn. Samoylenko	44.53	15.21	CG7	F	F	S	**No**
Croatia, Gospie, Velebit Mtn. Samoylenko	44.53	15.21	CG11	F	F	S	**No**
Croatia, Gospie, Velebit Mtn. Samoylenko	44.53	15.21	CG12	F	S	S	Yes
Croatia, Gospie, Velebit Mtn. Samoylenko	44.53	15.21	CG1	M	S	S	Yes
Croatia, Gospie, Velebit Mtn. Samoylenko	44.53	15.21	CG10	M	S	S	Yes
Croatia, Gospie, Velebit Mtn. Samoylenko	44.53	15.21	CG2	M	F	S	**No**
Croatia, Gospie, Velebit Mtn. Samoylenko	44.53	15.21	CG3	M	F	S	**No**
Croatia, Gospie, Velebit Mtn. Samoylenko	44.53	15.21	CG4	M	F	S	**No**
Croatia, Gospie, Velebit Mtn. Samoylenko	44.53	15.21	CG5	M	S	S	Yes
Croatia, Gospie, Velebit Mtn. Samoylenko	44.53	15.21	CG8	M	F	S	**No**
Croatia, Gospie, Velebit Mtn.	44.53	15.21	CG9	M	S	S	Yes

Samoylenko							
Drenovac, Golemo, Vranje, Serbia	42.63	21.85	SV1	M	F	S	**No**
Eaux Chaudes, France	44.53	15.21	TrFaFC1	F	F	F	Yes
France,Lac d'Issarles (Ardeche)	44.82	4.07	TrFa1	F	F	F	Yes
France,Lac d'Issarles (Ardeche)	44.82	4.07	TrFa2	M	F	F	Yes
Greece, Saloniki	40.64	22.81	GT1	M	F	S	**No**
Hautes-Pyrénées, France	42.95	-0.44	7865	F	-	S	-
Italy, BG, Roncobello, Bordognia	45.94	9.73	RB1	M	F	F	Yes
Italy, BG, Roncobello, Bordognia	45.94	9.73	RB2	M	S	F	**No**
Italy, BG, Roncobello, Bordognia	45.94	9.73	RB3	M	F	F	Yes
Italy, BG, Roncobello, Bordognia	45.94	9.73	RB4	M	S	F	**No**
Kadamzhai District, Kyrgyzstan	39.94	72.20	7849	F	-	G	-
Kadamzhai District, Kyrgyzstan	39.94	72.20	7852	M	-	G	-
Kimry dist. Tver reg., Shchelkovo vill. Russia	56.87	37.49	TK3	F	F	F	Yes
Kimry dist. Tver reg., Shchelkovo vill. Russia	56.87	37.49	TK7	F	F	F	Yes
Kimry dist. Tver reg., Shchelkovo vill. Russia	56.87	37.49	TK5	M	F	F	Yes
Kimry dist. Tver reg., Shchelkovo vill. Russia	56.87	37.49	TK2	F	F	G	**No**
Kimry dist. Tver reg., Shchelkovo vill. Russia	56.87	37.49	TK1	M	F	F	Yes
Kimry dist. Tver reg., Shchelkovo vill. Russia	56.87	37.49	TK4	M	F	G	Yes
Kimry dist. Tver reg., Shchelkovo vill. Russia	56.87	37.49	TK6	M	F	G	Yes
Krasnodar Ondracek	45.07	39.06	CK1	M	F	G	Yes
Ljubeli, Slovenia	46.43	14.27	SL1	F	F	F	**No**
Manastirski potok, southern Serbia	42.63	21.85	7862	M	-	F	-
Melnishki piramidi, southern Bulgaria	41.53	23.40	7859	M	-	S	-
Nizke Tatry, Ohniste, Slovakia	48.95	19.54	SN1	M	F	G	**No**
Norway, Eis 28, BO, Drammen, Konnerud	59.72	10.16	ND1	M	F	F	Yes
Poland, Barwinek	49.43	21.69	PB1	F	F	G	**No**
Poland, Barwinek	49.43	21.69	PB2	M	S	G	**No**

Poland, Mikow	49.29	22.12	PM1	F	F	G	**No**
Romania, Vâlcea	45.22	24.08	TrFaRV1	M	F	F	Yes
Romania, Vâlcea	45.22	24.08	TrFaRV2	M	F	F	Yes
Russia, NW Caucasus, Krasnodar region, near Sochi	43.62	39.78	TrFaRS1	M	F	F	Yes
Russia, NW Caucasus, Krasnodar region, near Sochi	43.62	39.78	TrFaRS2	M	F	F	Yes
Russia, 60 km SE Irkutsk. Baikal lake. Listvyanka.	51.86	104.88	RIr1	M	F	G	**No**
S. Kyrgyzstan, Alai Mts., Isfairam-Sai Valley	39.94	72.20	TrFaKA 1	M	F	F	Yes
S. Kyrgyzstan, Alai Mts., Isfairam-Sai Valley	39.94	72.20	TrFaKA 2	M	F	F	Yes
Thessaloniki, Greece	40.64	22.81	7857	F	-	F	-
USSR, Abchazia Cebelda/Suchumi/R.Rous 1gt.	43.03	41.01	AC1	M	F	F	Yes

Supplementary Table 2. Eigenvalues and the percentage of variation explained by the first three principle components in the *Trichius* dataset with all individuals but no sex-specific characteristics

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2.445	34.935	34.935
2	1.993	28.473	63.408
3	.865	12.359	75.767

Supplementary Table 3. Eigenvalues and the percentage of variation explained by the first three principle components in the male *Trichius* dataset including sex-specific characteristics

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2.854	35.673	35.673
2	2.502	31.280	66.954
3	.756	9.449	76.403

Supplementary Table 4. Eigenvalues and the percentage of variation explained by the first three principle components in the male *Trichius* dataset not including sex-specific characteristics

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2.663	38.037	38.037
2	2.058	29.403	67.440
3	.727	10.379	77.819

Supplementary Table 5. Eigenvalues and the percentage of variation explained by the first three principle components in the female *Trichius* dataset including sex-specific characteristics

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	3.440	42.999	42.999
2	1.934	24.169	67.168
3	.786	9.821	76.989

Supplementary Table 6. Eigenvalues and the percentage of variation explained by the first three principle components in the female *Trichius* dataset not including sex-specific characteristics

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	2.778	39.687	39.687
2	1.933	27.617	67.304
3	.754	10.765	78.069

Chapter 4: Geometric morphometrical and morphological variation across European *Gnorimus*

Supplementary Table 7. *Gnorimus* samples used in the morphological and genetic studies. The COI and CR haplotypes are given for each successfully sequenced individual. Inclusion in the HSP, *Orco*, Meristic and Geomorphometric analyses are designated by a "Y". British Museum of Natural History (BMNH) samples are listed at the end.

Species	Locality	Longitude	Latitude	Collection Date	COI Haplogroup	CR Haplogroup	HSP	Orco	Meristic	Geomorpho
G. nobilis	Velebit, Croatia	44.52929	15.23223	17/07/1985	1	3			Y	
G. nobilis	Bologna Lizzano, Italy	44.15995	10.88766	08/07/1990	7				Y	Y
G. nobilis	Mt. Čabulja, Mostar Prov, Bosnja/Hercegowina	43.47946	17.62246	20.6.2012	1	3	Y		Y	Y
G. nobilis	Mt. Čabulja, Mostar Prov, Bosnja/Hercegowina	43.47946	17.62246	20.6.2012	1	3	Y		Y	Y
G. nobilis	Mt. Čabulja, Mostar Prov, Bosnja/Hercegowina	43.47946	17.62246	20.6.2012	17	3	Y	Y	Y	Y
G. nobilis	Badia Pratalgia, Italy	43.80183	11.87334	26/07/1958	1		Y		Y	Y
G. nobilis	Mt. Čabulja, Mostar Prov, Bosnja/Hercegowina	43.47946	17.62246	20.6.2012	17	3	Y			
G. nobilis	Belluno, Vel Tovghella, Italy	46.16216	12.21101	08/07/1988	1		Y		Y	Y
G. nobilis	Badia Pratalgia, Italy	43.80183	11.87334	26/07/1958	1				Y	Y
G. nobilis	Val. Bartolo, Udine Tarvisio, Italy	46.50826	13.58178	18-22/7/2005	1				Y	Y
G. nobilis	Badia Pratalgia, Italy	43.80183	11.87334	26/07/1958	1				Y	Y
G. nobilis	Altopiano Montassio, Udine Tarvisio, Italy	46.41462	13.43089	20/07/2004	1		Y		Y	Y
G. nobilis	Val. Bartolo, Udine Tarvisio, Italy	46.50826	13.58178	18-22/7/2005	1		Y		Y	Y
G. nobilis	Leoncel, Ft. de Lente, Drome, France	44.91063	5.197819	7.7.2011	7		Y	Y	Y	Y
G. nobilis	Leoncel, Ft. de Lente, Drome, France	44.91063	5.197819	7.7.2011	7	5	Y	Y	Y	Y
G. nobilis	Leoncel, Ft. de Lente, Drome, France	44.91063	5.197819	7.7.2011	1					Y
G. nobilis	Lungoz, Ospitale, Fanano, Italy	46.32394	12.30022	04/06/2003	1				Y	Y
G. nobilis	Cave di Predil, Udine Tarvisio, Italy	46.4376	13.56824	31/07/1995	1				Y	Y
G. nobilis	Lungoz, Ospitale, Fanano, Italy	46.32394	12.30022	04/06/2003	3				Y	Y

G. nobilis	Acero di Pollino, Polenza, Italy	39.88555	16.20127	04/07/2009	8	12	Y	Y	Y	Y
G. nobilis	Aiguines, France	43.77385	6.238695	7.2009	1		Y			
G. nobilis	Aiguines, France	43.77385	6.238695	7.2009	18		Y	Y	Y	Y
G. nobilis	Aiguines, France	43.77385	6.238695	07/01/1900	1	3	Y			
G. nobilis	Aiguines, France	43.77385	6.238695	17.7.2012	18		Y	Y	Y	Y
G. nobilis	Aiguines, France	43.77385	6.238695	17.7.2012	18	3	Y	Y		
G. nobilis	Wels, Austria	48.16417	14.02978	31/05/1905	1	3	Y	Y	Y	Y
G. nobilis	Parc Paradiso, Belgium	50.58506	3.886501	10.07.2008	1			Y	Y	Y
G. nobilis	Melnik, Bulgaria	41.52715	23.39701	28/05/2013	1	10		Y	Y	Y
G. nobilis	Bedovice, Czech Republic	50.19903	16.01847	2.6.2011	1	3	Y	Y	Y	Y
G. nobilis	Stirovaca, Croatia	44.69115	15.05247	8.2006	1	3	Y		Y	Y
G. nobilis	Stirovaca, Croatia	44.69115	15.05247	8.2006	1		Y		Y	Y
G. nobilis	Stirovaca, Croatia	44.69115	15.05247	8.2006	1	11	Y			
G. nobilis	Velebit, Gospie, Croatia	44.53444	15.37281	23/07/2014	1	3				Y
G. nobilis	Velebit, Gospie, Croatia	44.53444	15.37281	23/07/2014	1	19	Y			Y
G. nobilis	Velebit, Gospie, Croatia	44.53444	15.37281	23/07/2014	1		Y	Y		Y
G. nobilis	Velebit, Gospie, Croatia	44.53444	15.37281	23/07/2014	1		Y			Y
G. nobilis	Velebit, Gospie, Croatia	44.53444	15.37281	23/07/2014	1		Y			Y
G. nobilis	Velebit, Gospie, Croatia	44.53444	15.37281	23/07/2014	1		Y			Y
G. nobilis	Ardeche, France	44.73131	4.394912	6.1993	1	3	Y		Y	Y
G. nobilis	Ardeche, France	44.73131	4.394912	6.1993	1		Y		Y	Y
G. nobilis	Eaux Chandes, France	42.95225	-0.44058	01/07/1993	1		Y	Y		
G. nobilis	Eaux Chandes, France	42.95225	-0.44058	01/07/1993	1			Y		
G. nobilis	N.W. Gresigne, (Ft de la Gawigne?) France	44.04049	1.734833	16-23/7/2005	1	3				
G. nobilis	N.W. Gresigne, (Ft de la Gawigne?) France	44.04049	1.734833	16-23/7/2005	1					
G. nobilis	N.W. Gresigne, (Ft de la Gawigne?) France	44.04049	1.734833	16-23/7/2005	1			Y	Y	Y

G. nobilis	N.W. Gresigne, (Ft de la Gawigne?) France	44.04049	1.734833	16-23/7/2005	1			Y	Y	Y
G. nobilis	Panteleimonas, Pieria, Greece	39.98653	22.5712	12.6.2013	1		Y	Y		
G. nobilis	Panteleimonas, Pieria, Greece	39.98653	22.5712	12.6.2013	16	3	Y	Y		
G. nobilis	Taibon, Italy	46.30179	12.01865	26/06/2010	1	3			Y	Y
G. nobilis	Dundaga parish, Slitere NP, Davidplava, Latvia	57.62145	22.29222	19/07/2012	1	17	Y	Y	Y	Y
G. nobilis	Dundaga parish, Slitere NP, Davidplava, Latvia	57.62145	22.29222	19/07/2012	1	18		Y	Y	Y
G. nobilis	Dundaga parish, Slitere NP, Davidplava, Latvia	57.62145	22.29222	19/07/2012	1	14	Y	Y	Y	Y
G. nobilis	Sirolo, Marche, Italy	43.52113	13.60748	07/08/2013	1	13	Y		Y	Y
G. nobilis	Pelvoux, France	44.89913	6.441514	01/06/2009	1	3	Y	Y		Y
G. nobilis	Pelvoux, France	44.89913	6.441514	01/06/2009	1			Y		Y
G. nobilis	Mikow, Poland	49.28965	22.12298	17/07.2013	2	3	Y		Y	Y
G. nobilis	Mikow, Poland	49.28965	22.12298	17/07.2013	3	3		Y		Y
G. nobilis	Mikow, Poland	49.28965	22.12298	17/07.2013	2	3		Y	Y	Y
G. nobilis	Oltenia, Mehedinti, Bahna, Romania	44.72123	22.47112	06/07/2011	1	3			Y	Y
G. nobilis	Oltenia, Mehedinti, Bahna, Romania	44.72123	22.47112	06/07/2011	1		Y		Y	Y
G. nobilis	Barzawa, Romania	46.1156	21.99899	15/05/2014	1		Y			
G. nobilis	Barzawa, Romania	46.1156	21.99899	16/05/2014	1		Y		Y	Y
G. nobilis	Vilaceau, Romania	45.1046	24.40051	07/07/2013	1	3	Y	Y	Y	Y
G. nobilis	Salguero de Juanos, Spain	42.30163	-3.4879	20/06/2007	4	6	Y	Y	Y	Y
G. nobilis	Palencia, Olleros, Spain	41.96742	-4.55713	13/06/2007	4	6	Y	Y	Y	Y
G. nobilis	Palencia, Olleros, Spain	41.96742	-4.55713	13/06/2007	5	8	Y	Y	Y	Y
G. nobilis	Palencia, Olleros, Spain	41.96742	-4.55713	13/06/2007	4	6	Y	Y	Y	Y
G. nobilis	Palencia, Olleros, Spain	42.74537	-4.28149	13/06/2007	6	7		Y	Y	Y
G. nobilis	Palencia, Olleros, Spain	42.74537	-4.28149	13/06/2007	4		Y	Y	Y	Y
G. nobilis	Remetske Hamre, Vihorlatske, Slovakia	48.91005	22.08185	05/07/2003	1			Y	Y	Y
G. nobilis	Selva de Oza, Hecho, Spain	42.83427	-0.71019	18/07/1983	1	3				
G. nobilis	Bujanuelo, Torla, Spain	42.69464	-0.11072	27/07/1983	1		Y	Y		

G. nobilis	Omberg, Sweden	58.3293	14.65301	01/07/2014	1	3	Y	Y	Y	Y
G. nobilis	Omberg, Sweden	58.3293	14.65301	01/07/2014	1	14	Y		Y	Y
G. nobilis	Chiornyj Les, Kirovograd, Ukraine	48.60805	32.29514	31/05/2007	10	21	Y		Y	Y
G. nobilis	Chiornyj Les, Kirovograd, Ukraine	48.60805	32.29514	31/05/2007	11		Y		Y	Y
G. nobilis	Chiornyj Les, Kirovograd, Ukraine	48.60805	32.29514	31/05/2007	11	21	Y	Y	Y	Y
G. nobilis	Chiornyj Les, Kirovograd, Ukraine	48.60805	32.29514	31/05/2007		21			Y	Y
G. nobilis	Chiornyj Les, Kirovograd, Ukraine	48.60805	32.29514	31/05/2007	12	22	Y	Y	Y	Y
G. nobilis	Chiornyj Les, Kirovograd, Ukraine	48.60805	32.29514	31/05/2007	10	23			Y	Y
G. nobilis	Chiornyj Les, Kirovograd, Ukraine	48.60805	32.29514	31/05/2007					Y	Y
G. nobilis	Chiornyj Les, Kirovograd, Ukraine	48.60805	32.29514	31/05/2007	13	23	Y	Y		Y
G. nobilis	Koctylyovka, Rakhov, Uzhgorod, Ukraine	47.98932	24.19526	17/07/2003	1		Y		Y	Y
G. nobilis	Koctylyovka, Rakhov, Uzhgorod, Ukraine	47.98932	24.19526	17/07/2003	1	3	Y	Y	Y	Y
G. nobilis	Koctylyovka, Rakhov, Uzhgorod, Ukraine	47.98932	24.19526	17/07/2003	1	3	Y	Y	Y	Y
G. nobilis	Koctylyovka, Rakhov, Uzhgorod, Ukraine	47.98932	24.19526	17/07/2003	1	19		Y	Y	Y
G. nobilis	Koctylyovka, Rakhov, Uzhgorod, Ukraine	47.98932	24.19526	17/07/2003	1	19	Y	Y	Y	Y
G. nobilis	Tiddesley, Worcestershire	52.11315	-2.10353	15/06/2013	1	3	Y			
G. nobilis	Tiddesley, Worcestershire	52.11315	-2.10353	14/06/2013	1		Y	Y		
G. nobilis	Tiddesley, Worcestershire	52.11315	-2.10353	14/06/2013	1					
G. nobilis	Bologna Lizzano, Italy	44.15995	10.88766	10/07/1991	1				Y	Y
G. nobilis	Acceltura, Matera, Italy	40.6745	16.61781	22/05/2005	9	12	Y	Y		
G. nobilis	Iwade, Kent	51.37715	0.731342		1	1	Y	Y		
G. nobilis	Dębica, Poland	50.04165	21.41068	05/06.2013		15				
G. nobilis	Dębica, Poland	50.04165	21.41068	05/06.2013	1	16	Y			
G. nobilis	Dębica, Poland	50.04165	21.41068	05/06.2013	2	16	Y			
G. nobilis	Dębica, Poland	50.04165	21.41068	05/06.2013	1	16	Y		Y	Y
G. nobilis	Dębica, Poland	50.04165	21.41068	05/06.2013		16		Y		
G. nobilis	Dębica, Poland	50.04165	21.41068	05/06.2013	1	3				
G. nobilis	Dębica, Poland	50.04165	21.41068	05/06.2013	2			Y	Y	Y

G. nobilis	Dębica, Poland	50.04165	21.41068	05/06.2013						
G. nobilis	Dębica, Poland	50.04165	21.41068	05/06.2013						
G. nobilis	Banat, Caras, Severin, Bistra Valley, Romania	45.51118	22.35302	23.7.2011	1	3			Y	Y
G. nobilis	Banat, Caras, Severin, Bistra Valley, Romania	45.51118	22.35302	23.7.2011	1	3			Y	Y
G. nobilis	Banat, Caras, Severin, Bistra Valley, Romania	45.51118	22.35302	23.7.2011	1	3				
G. nobilis	Banat, Caras, Severin, Bistra Valley, Romania	45.51118	22.35302	23.7.2011	1	3			Y	Y
G. nobilis	Banat, Caras, Severin, Bistra Valley, Romania	45.51118	22.35302	23.7.2011	1					
G. nobilis	Banat, Caras, Severin, Bistra Valley, Romania	45.51118	22.35302	23.7.2011	1				Y	Y
G. nobilis	Banat, Caras, Severin, Bistra Valley, Romania	45.51118	22.35302	23.7.2011	1		Y		Y	Y
G. nobilis	Banat, Caras, Severin, Bistra Valley, Romania	45.51118	22.35302	23.7.2011	1			Y	Y	Y
G. nobilis	Cleeve Prior, Worcestershire	52.14288	-1.87	03/06/2013	1					
G. nobilis	Cleeve Prior, Worcestershire	52.14288	-1.87	04/06/2013	1	2		Y		
G. nobilis	Cleeve Prior, Worcestershire	52.14288	-1.87	05/06/2013	1	2	Y			
G. nobilis	Cleeve Prior, Worcestershire	52.14288	-1.87	06/06/2013	1	2	Y	Y		
G. nobilis	Cleeve Prior, Worcestershire	52.14288	-1.87	07/06/2013	1	2	Y	Y		
G. nobilis	Cleeve Prior, Worcestershire	52.14288	-1.87	08/06/2013	1	2	Y			
G. nobilis	Cleeve Prior, Worcestershire	52.14288	-1.87	09/06/2013	1	2	Y			
G. nobilis	Cleeve Prior, Worcestershire	52.14288	-1.87	10/06/2013	1	2	Y	Y		
G. nobilis	Cleeve Prior, Worcestershire	52.14288	-1.87	11/06/2013	1	2		Y		
G. nobilis	Banat, Caras, Severin, Bistra Valley, Romania	45.51118	22.35302	23.7.2011	1					
G. nobilis	Banat, Caras, Severin, Bistra Valley, Romania	45.51118	22.35302	23.7.2011	1					
G. nobilis	Banat, Caras, Severin, Bistra Valley, Romania	45.51118	22.35302	23.7.2011	1					
G. nobilis	Banat, Caras, Severin, Bistra Valley, Romania	45.51118	22.35302	23.7.2011	1					
G. nobilis	The Cherries, Worcestershire, UK	52.35658	-2.37337	?	1					
G. n. bolshakovi	Orlovo, Stschekino, Tula, Russia	53.95903	37.16011	6.7.2003	14		Y	Y	Y	Y
G. n. bolshakovi	Zaseki Forest, Orlovo, Tula, Russia	53.99082	36.8747	26/06/2005	10	20	Y	Y	Y	Y
G. n. bolshakovi	Zaseki Forest, Orlovo, Tula, Russia	53.99082	36.8747	27/06/2005		20	Y		Y	Y
G. n.	Between Ribnica and Nistrovo, Macedonia	41.72651	20.62399	8.6.2014	15	21			Y	Y

macedonicus										
G. n. macedonicus	Skopje, Skopska Crna Mts. Macedonia	42.16538	21.41583	9.7.1997	1					
G. n. macedonicus	Skopje, Skopska Crna Mts. Macedonia	42.16538	21.41583	9.7.1997	1					
G. subopacus	Sovetskoye, S. Sakhalin, East Russia			3.7.2013	-	-		Y		
G. subopacus	Kamenshka, Ussuriysk, Primorsky Kray, Russia			12.06.2012	-	-				
G. variabilis	Civate, Italy	45.83056	9.334333	24.7.2012	West	-			Y	Y
G. variabilis	Civate, Italy	45.83056	9.334333	24.7.2012	West	-			Y	Y
G. variabilis	Brive, France			07/01/1900	West	-	Y	Y	Y	Y
G. variabilis	Verignon, France	43.66042	6.276318	01/06/2009	West	-		Y	Y	Y
G. variabilis	Verignon, France	43.66042	6.276318	01/06/2009	West	-	Y		Y	Y
G. variabilis	Verignon, France	43.66042	6.276318	01/06/2009	West	-	Y	Y	Y	Y
G. variabilis	Verignon, France	43.66042	6.276318	01/06/2009	West	-	Y	Y	Y	Y
G. variabilis	Verignon, France	43.66042	6.276318	01/06/2009	West	-	Y		Y	Y
G. variabilis	Verignon, France	43.66042	6.276318	01/06/2009	West	-	Y		Y	Y
G. variabilis	Remet, Slovakia			15.6.1972	East	-	Y	Y	Y	Y
G. variabilis	Belaya Krinitsa, Radomyshi, Zhitomir, Ukraine	50.63266	29.47394	14.07.2011	East	-	Y	Y	Y	Y
G. variabilis	Belaya Krinitsa, Radomyshi, Zhitomir, Ukraine	50.63266	29.47394	16.06.2013	East	-		Y	Y	Y
G. variabilis	Belaya Krinitsa, Radomyshi, Zhitomir, Ukraine	50.63266	29.47394	16.06.2013	East	-		Y		Y
G. variabilis	Belaya Krinitsa, Radomyshi, Zhitomir, Ukraine	50.63266	29.47394	16.06.2013	East	-	Y		Y	Y
G. variabilis	Aiguines, France	43.76261	6.259295	01/07/2009	West	-	Y	Y	Y	Y
G. variabilis	Windsor Castle	51.48916	-0.59789		West	-	Y	Y		Y
G. variabilis	Windsor Castle	51.48916	-0.59789		West	-	Y		Y	Y
G. variabilis	BMNHE_887812									Y
G. variabilis	BMNHE_887813									Y
G. variabilis	BMNHE_887814									Y
G. variabilis	BMNHE_887816									Y
G. variabilis	BMNHE_887817									Y

G. variabilis	BMNHE_887819								Y
G. variabilis	BMNHE_887821								Y
G. variabilis	BMNHE_887822								Y
G. variabilis	BMNHE_887823								Y
G. variabilis	BMNHE_887824								Y
G. variabilis	BMNHE_887825								Y
G. variabilis	BMNHE_887827								Y
G. variabilis	BMNHE_887828								Y
G. variabilis	BMNHE_887830								Y
G. variabilis	BMNHE_887834								Y
G. variabilis	BMNHE_887835								Y
G. variabilis	BMNHE_887836								Y
G. variabilis	BMNHE_887837								Y
G. variabilis	BMNHE_887838								Y
G. variabilis	BMNHE_887839								Y
G. variabilis	BMNHE_887840								Y
G. variabilis	BMNHE_887841								Y
G. variabilis	BMNHE_887842								Y
G. variabilis	BMNHE_887843								Y
G. variabilis	BMNHE_887844								Y
G. variabilis	BMNHE_887845								Y
G. variabilis	BMNHE_887847								Y
G. variabilis	BMNHE_887848								Y
G. variabilis	BMNHE_887849								Y
G. nobilis	BMNHE_887850								Y
G. nobilis	BMNHE_887851								Y
G. nobilis	BMNHE_887852								Y
G. nobilis	BMNHE_887854								Y

G. nobilis	BMNHE_887855								Y
G. nobilis	BMNHE_887856								Y
G. nobilis	BMNHE_887857								Y
G. nobilis	BMNHE_887858								Y
G. nobilis	BMNHE_887859								Y
G. nobilis	BMNHE_887860								Y
G. nobilis	BMNHE_887861								Y
G. nobilis	BMNHE_887863								Y
G. nobilis	BMNHE_887864								Y
G. nobilis	BMNHE_887865								Y
G. nobilis	BMNHE_887866								Y
G. nobilis	BMNHE_887867								Y
G. nobilis	BMNHE_887868								Y
G. nobilis	BMNHE_887869								Y
G. nobilis	BMNHE_887870								Y
G. nobilis	BMNHE_887871								Y
G. nobilis	BMNHE_887873								Y
G. nobilis	BMNHE_887875								Y
G. nobilis	BMNHE_887878								Y
G. nobilis	BMNHE_887881								Y
G. nobilis	BMNHE_887882								Y
G. subopacus	BMNHE_884126							Y	
G. subopacus	BMNHE_884127							Y	
G. subopacus	BMNHE_884128							Y	
G. subopacus	BMNHE_884129							Y	
G. subopacus	BMNHE_884130							Y	
G. subopacus	BMNHE_884131							Y	

Supplementary Table 8. Discriminant analysis showing the ability of the three PCA values to discriminate between the four *Gnorimus* haplotypes

		Taxon	Predicted Group Membership				Total
			Western	Eastern	Subopacus	Variabilis	
Cross-validated[b]	Count	Western	13	16	2	2	33
		Eastern	3	1	0	0	4
		Subopacus	0	0	7	0	7
		Variabilis	1	0	0	14	15
	%	Western	39.4	48.5	6.1	6.1	100.0
		Eastern	75.0	25.0	.0	.0	100.0
		Subopacus	.0	.0	100.0	.0	100.0
		Variabilis	6.7	.0	.0	93.3	100.0

Supplementary Table 9. ANOVA results for significant differences in group means for *Gnorimus*

ANOVA

		Sum of Squares	df	Mean Square	F
REGR factor score 1 for analysis 1	Between Groups	68.046	3	22.682	40.049
	Within Groups	84.954	150	.566	
	Total	153.000	153		
REGR factor score 2 for analysis 1	Between Groups	43.902	3	14.634	20.120
	Within Groups	109.098	150	.727	
	Total	153.000	153		
REGR factor score 3 for analysis 1	Between Groups	56.923	3	18.974	29.624
	Within Groups	96.077	150	.641	
	Total	153.000	153		

Supplementary Table 10. Tukey post-hoc test showing between-group differences based on the regression scores (PC1-3)

Dependent Variable	(I) Taxon	(J) Taxon	Mean Difference (I-J)	Std. Error	Sig.
REGR factor score 1 for analysis 1	Western	Eastern	.06510742	.21475014	.990
		Subopacus	-1.53664979*	.22187716	.000
		Variabilis	1.19274608*	.16321686	.000
	Eastern	Western	-.06510742	.21475014	.990
		Subopacus	-1.60175721*	.28986197	.000
		Variabilis	1.12763866*	.24785123	.000
	Subopacus	Western	1.53664979*	.22187716	.000
		Eastern	1.60175721*	.28986197	.000
		Variabilis	2.72939587*	.25405134	.000
	Variabilis	Western	-1.19274608*	.16321686	.000
		Eastern	-1.12763866*	.24785123	.000
		Subopacus	-2.72939587*	.25405134	.000
REGR factor score 2 for analysis 1	Western	Eastern	.00881398	.24336148	1.000
		Subopacus	.22030475	.25143803	.817
		Variabilis	-1.37162654*	.18496238	.000
	Eastern	Western	-.00881398	.24336148	1.000
		Subopacus	.21149077	.32848051	.918
		Variabilis	-1.38044052*	.28087265	.000
	Subopacus	Western	-.22030475	.25143803	.817
		Eastern	-.21149077	.32848051	.918
		Variabilis	-1.59193129*	.28789880	.000
	Variabilis	Western	1.37162654*	.18496238	.000
		Eastern	1.38044052*	.28087265	.000
		Subopacus	1.59193129*	.28789880	.000
REGR factor score 3 for analysis 1	Western	Eastern	.29917271	.22837674	.558
		Subopacus	-1.67515403*	.23595599	.000
		Variabilis	-1.12401740*	.17357351	.000
	Eastern	Western	-.29917271	.22837674	.558
		Subopacus	-1.97432673*	.30825466	.000
		Variabilis	-1.42319010*	.26357820	.000
	Subopacus	Western	1.67515403*	.23595599	.000
		Eastern	1.97432673*	.30825466	.000
		Variabilis	.55113663	.27017173	.178
	Variabilis	Western	1.12401740*	.17357351	.000
		Eastern	1.42319010*	.26357820	.000
		Subopacus	-.55113663	.27017173	.178

A values 1-3) for *Gnorimus*

Supplementary Table 11. Discriminant analysis showing the ability of the three PCA values to discriminate between the four female *Gnorimus* haplotypes

		Haplogroup	Predicted Group Membership				Total
			Western	Eastern	Subopacus	Variabilis	
Original	Count	Western	19	11	1	2	33
		Eastern	1	3	0	0	4
		Subopacus	0	0	7	0	7
		Variabilis	1	0	0	14	15
	%	Western	57.6	33.3	3.0	6.1	100.0
		Eastern	25.0	75.0	.0	.0	100.0
		Subopacus	.0	.0	100.0	.0	100.0
		Variabilis	6.7	.0	.0	93.3	100.0
Cross-validated[b]	Count	Western	13	16	2	2	33
		Eastern	3	1	0	0	4
		Subopacus	0	0	7	0	7
		Variabilis	1	0	0	14	15
	%	Western	39.4	48.5	6.1	6.1	100.0
		Eastern	75.0	25.0	.0	.0	100.0
		Subopacus	.0	.0	100.0	.0	100.0
		Variabilis	6.7	.0	.0	93.3	100.0

Supplementary Table 12. ANOVA results for significant differences in group means for female *Gnorimus*

		Sum of Squares	df	Mean Square	F	Sig.
REGR factor score 1 for analysis 1	Between Groups	17.092	3	5.697	6.741	.000
	Within Groups	76.908	91	.845		
	Total	94.000	94			
REGR factor score 2 for analysis 1	Between Groups	63.534	3	21.178	63.256	.000
	Within Groups	30.466	91	.335		
	Total	94.000	94			
REGR factor score 3 for analysis 1	Between Groups	33.766	3	11.255	17.004	.000
	Within Groups	60.234	91	.662		
	Total	94.000	94			

Supplementary Table 13. Tukey post-hoc test showing between-group differences based on the regression scores (PCA values 1-3) for female *Gnorimus*

Tukey HSD

Dependent Variable	(I) Taxon	(J) Taxon	Mean Difference	Std. Error	Sig.
REGR factor score 1 for analysis 1	Western	Eastern	-.17934636	.31579432	.941
		Subopacus	-1.86243968*	.24820723	.000
		Variabilis	.96720650*	.18574128	.000
	Eastern	Western	.17934636	.31579432	.941
		Subopacus	-1.68309332*	.37385892	.000
		Variabilis	1.14655286*	.33565382	.006
	Subopacus	Western	1.86243968*	.24820723	.000
		Eastern	1.68309332*	.37385892	.000
		Variabilis	2.82964618*	.27302795	.000
	Variabilis	Western	-.96720650*	.18574128	.000
		Eastern	-1.14655286*	.33565382	.006
		Subopacus	-2.82964618*	.27302795	.000
REGR factor score 2 for analysis 1	Western	Eastern	.18789959	.43807671	.973
		Subopacus	-.68429286	.34431843	.205
		Variabilis	-1.33521262*	.25766432	.000
	Eastern	Western	-.18789959	.43807671	.973
		Subopacus	-.87219245	.51862518	.343
		Variabilis	-1.52311220*	.46562624	.010
	Subopacus	Western	.68429286	.34431843	.205
		Eastern	.87219245	.51862518	.343
		Variabilis	-.65091975	.37875028	.324
	Variabilis	Western	1.33521262*	.25766432	.000
		Eastern	1.52311220*	.46562624	.010
		Subopacus	.65091975	.37875028	.324
REGR factor score 3 for analysis 1	Western	Eastern	-.01691784	.47258071	1.000
		Subopacus	-1.24162561*	.37143780	.008
		Variabilis	-.87623704*	.27795860	.014
	Eastern	Western	.01691784	.47258071	1.000
		Subopacus	-1.22470777	.55947337	.139
		Variabilis	-.85931920	.50230011	.328
	Subopacus	Western	1.24162561*	.37143780	.008
		Eastern	1.22470777	.55947337	.139
		Variabilis	.36538857	.40858158	.808
	Variabilis	Western	.87623704*	.27795860	.014
		Eastern	.85931920	.50230011	.328
		Subopacus	-.36538857	.40858158	.808

Supplementary Table 14. Discriminant analysis showing the ability of the three PCA values to discriminate between the four male *Gnorimus* haplotypes

Classification Results[a,c]

		Taxon	Predicted Group Membership				Total
			Western	Eastern	Subopacus	Variabilis	
Cross-validated[b]	Count	Western	51	12	4	0	67
		Eastern	1	9	0	0	10
		Subopacus	1	0	5	0	6
		Variabilis	0	1	0	11	12
	%	Western	76.1	17.9	6.0	.0	100.0
		Eastern	10.0	90.0	.0	.0	100.0
		Subopacus	16.7	.0	83.3	.0	100.0
		Variabilis	.0	8.3	.0	91.7	100.0

Supplementary Table 15. ANOVA results for significant differences in group means for male *Gnorimus*

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
REGR factor score 1 for analysis 1	Between Groups	17.092	3	5.697	6.741	.000
	Within Groups	76.908	91	.845		
	Total	94.000	94			
REGR factor score 2 for analysis 1	Between Groups	63.534	3	21.178	63.256	.000
	Within Groups	30.466	91	.335		
	Total	94.000	94			
REGR factor score 3 for analysis 1	Between Groups	33.766	3	11.255	17.004	.000
	Within Groups	60.234	91	.662		
	Total	94.000	94			

Supplementary Table 16. Tukey post-hoc test showing between-group differences based on the regression scores (PCA values 1-3) for male *Gnorimus*

Tukey HSD

Dependent Variable	(I) Taxon	(J) Taxon	Mean Difference (I-J)	Std. Error	Sig.
REGR factor score 1 for analysis 1	Western	Eastern	-.07450995	.31165484	.995
		Subopacus	-1.50187472*	.39175479	.001
		Variabilis	.54837518	.28817177	.234
	Eastern	Western	.07450995	.31165484	.995
		Subopacus	-1.42736476*	.47473394	.018
		Variabilis	.62288513	.39362859	.394
	Subopacus	Western	1.50187472*	.39175479	.001
		Eastern	1.42736476*	.47473394	.018
		Variabilis	2.05024989*	.45965916	.000
	Variabilis	Western	-.54837518	.28817177	.234
		Eastern	-.62288513	.39362859	.394
		Subopacus	-2.05024989*	.45965916	.000
REGR factor score 2 for analysis 1	Western	Eastern	-.60900380*	.19615403	.013
		Subopacus	.13446291	.24656855	.948
		Variabilis	-2.45695961*	.18137390	.000
	Eastern	Western	.60900380*	.19615403	.013
		Subopacus	.74346671	.29879522	.069
		Variabilis	-1.84795581*	.24774791	.000
	Subopacus	Western	-.13446291	.24656855	.948
		Eastern	-.74346671	.29879522	.069
		Variabilis	-2.59142252*	.28930722	.000
	Variabilis	Western	2.45695961*	.18137390	.000
		Eastern	1.84795581*	.24774791	.000
		Subopacus	2.59142252*	.28930722	.000
REGR factor score 3 for analysis 1	Western	Eastern	.40416036	.27580913	.463
		Subopacus	2.27308821*	.34669620	.000
		Variabilis	.89644354*	.25502702	.004
	Eastern	Western	-.40416036	.27580913	.463
		Subopacus	1.86892786*	.42013131	.000
		Variabilis	.49228319	.34835448	.494
	Subopacus	Western	-2.27308821*	.34669620	.000
		Eastern	-1.86892786*	.42013131	.000
		Variabilis	-1.37664467*	.40679039	.006
	Variabilis	Western	-.89644354*	.25502702	.004
		Eastern	-.49228319	.34835448	.494
		Subopacus	1.37664467*	.40679039	.006

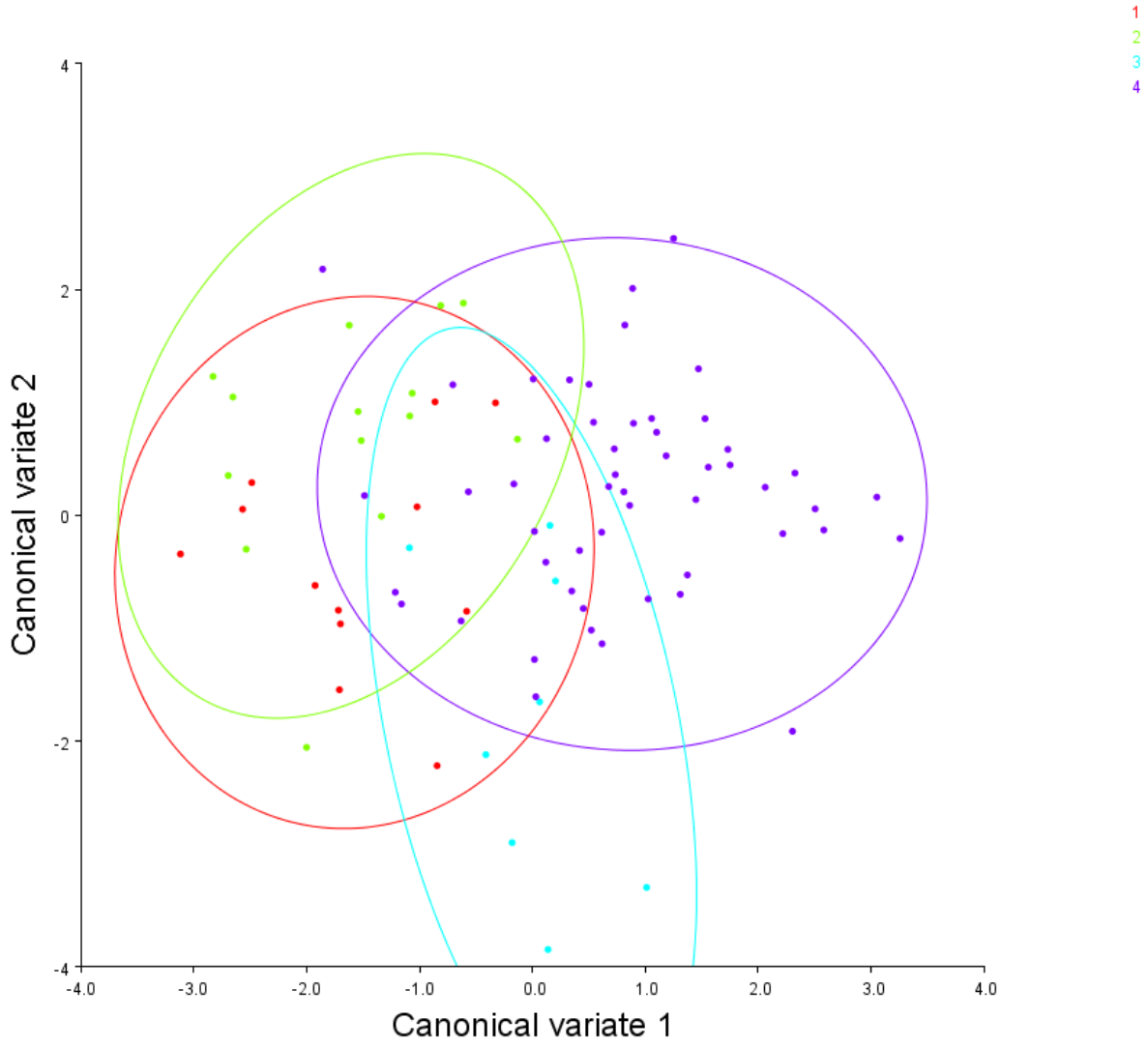
Supplementary Table 17. Mahalanobis distance between metatibial groups for the dorsal analysis, and the significance of the Discriminant analysis to differentiate between the groups. Significant differences between groups are marked in bold

Metatibia score	1	2	3
2	3.2749 *P*=0.8370	-	-
3	9.1180 *P*=0.6205	8.2076 *P*=0.7070	-
4	2.6604 ***P*=0.0135**	2.5191 ***P*=0.0102**	2.6811 *P*=0.1161

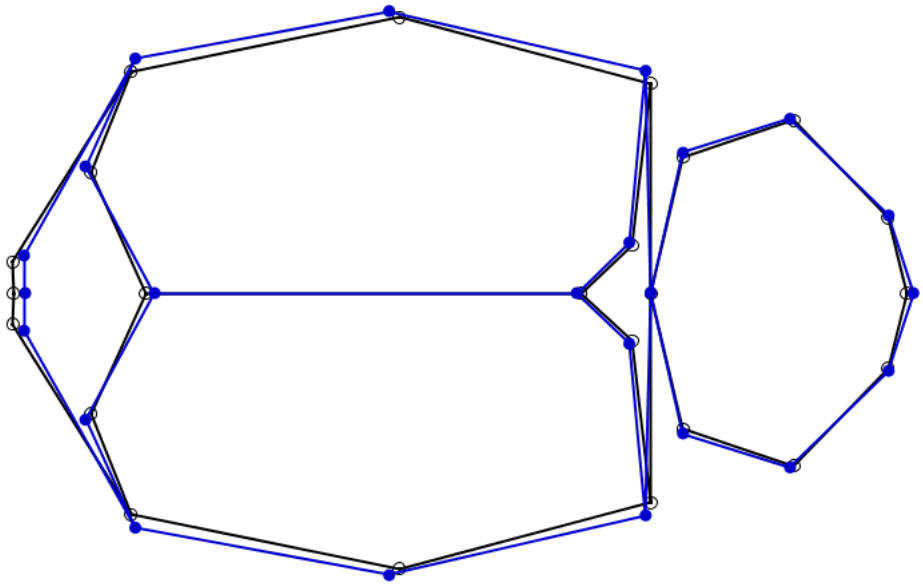
Supplementary Table 18: distance between groups, and the significance of the Discriminant analysis to differentiate between the groups of ventral male *Gnorimus nobilis* based on metatibia score. Significant differences between groups are marked in bold

Metatibia score	1	2	3
2	5.7806 *P*=0.8845	-	-
3	4.1578 *P*=0.9277	4.4238 *P*=0.9374	-
4	4.6976 *P*=0.2191	5.9695 ***P*=0.0166**	8.5395 *P*=0.3906

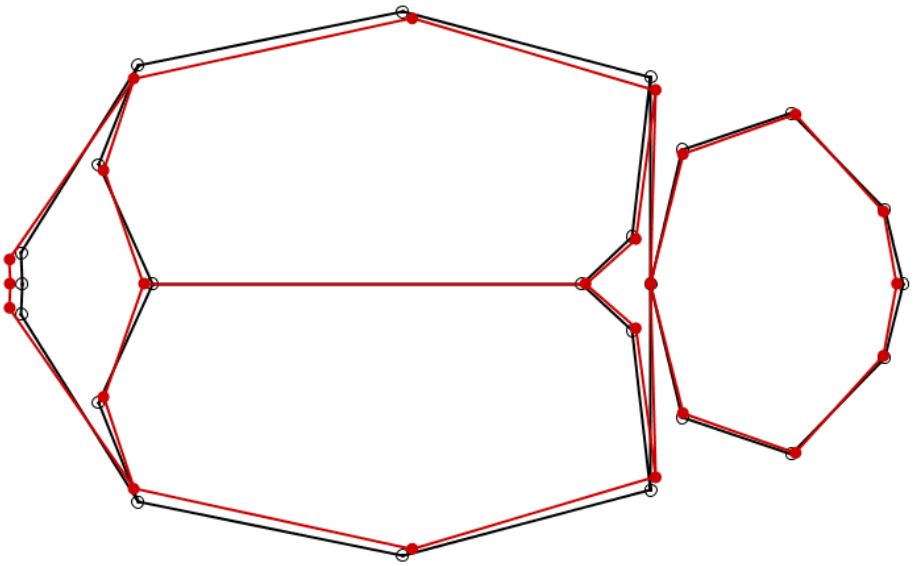
Supplementary Figure 1. Plot of canonical variate 1 against canonical variate 2 for the shape of the male mesotibia dorsal dataset. The confidence ellipses show 95% probability from equal frequency



Supplementary Figure 2. Positive (blue) and negative (red) shifts in canonical variate 1 for the dorsal metatibia analysis. These represent the most extreme values from the data for each variate, showing how the scores vary

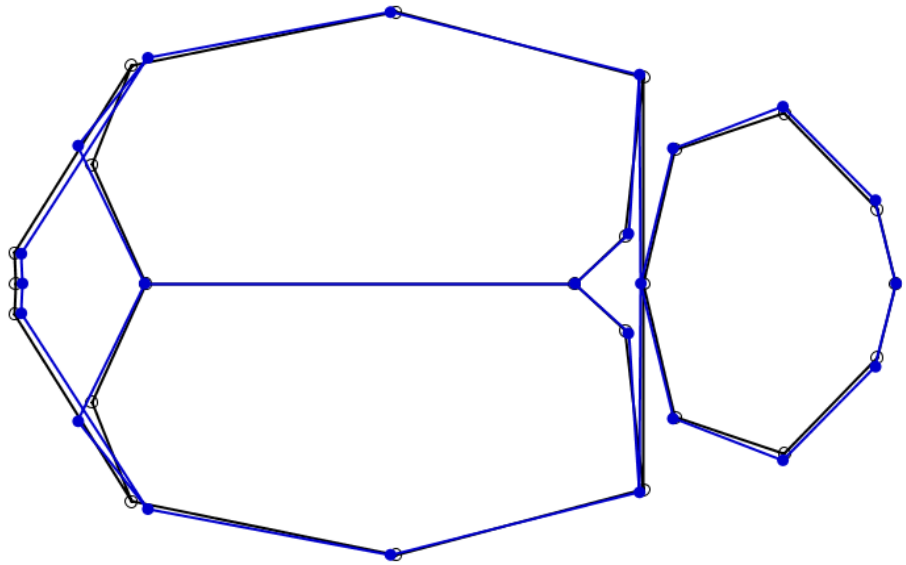


CV1

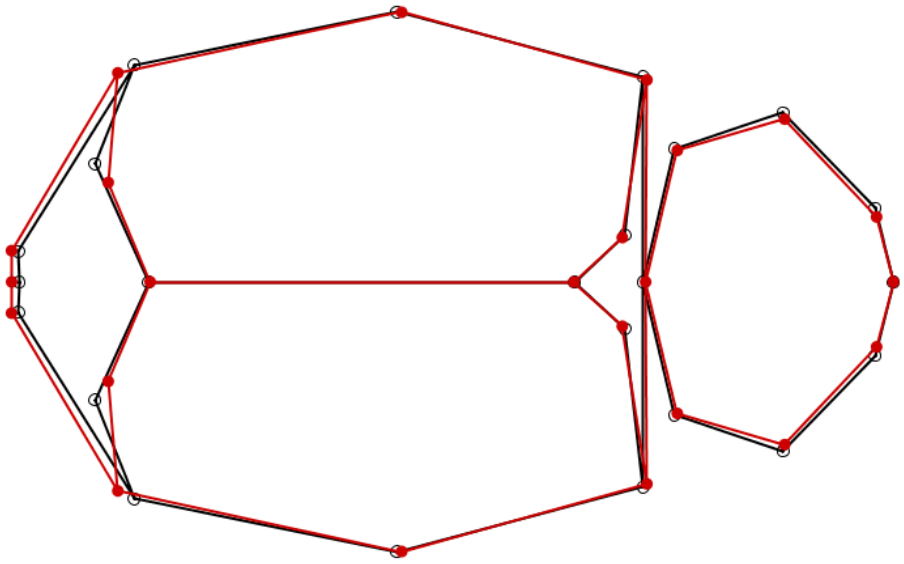


CV1

Supplementary Figure 3. Positive (blue) and negative (red) shifts in canonical variate 2 for the dorsal metatibia analysis. These represent the most extreme values from the data for each variate, showing how the scores vary.

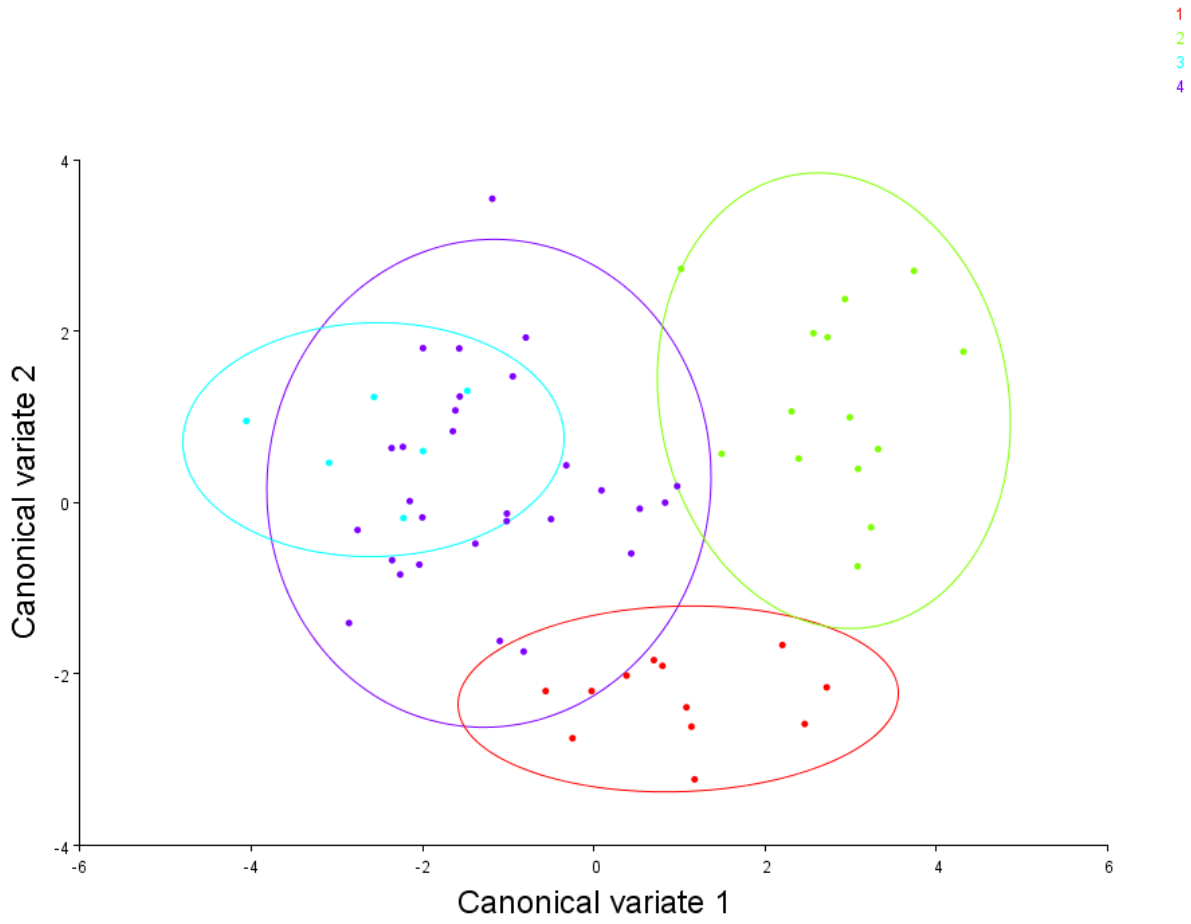


CV2

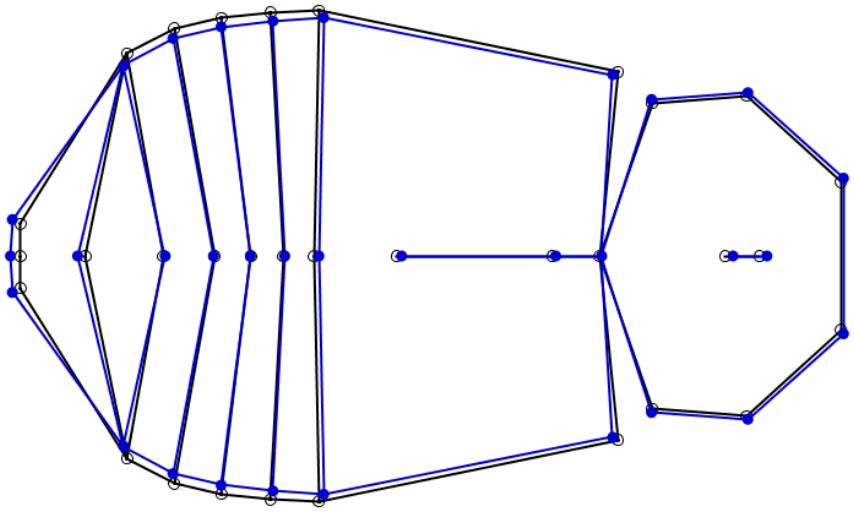


CV2

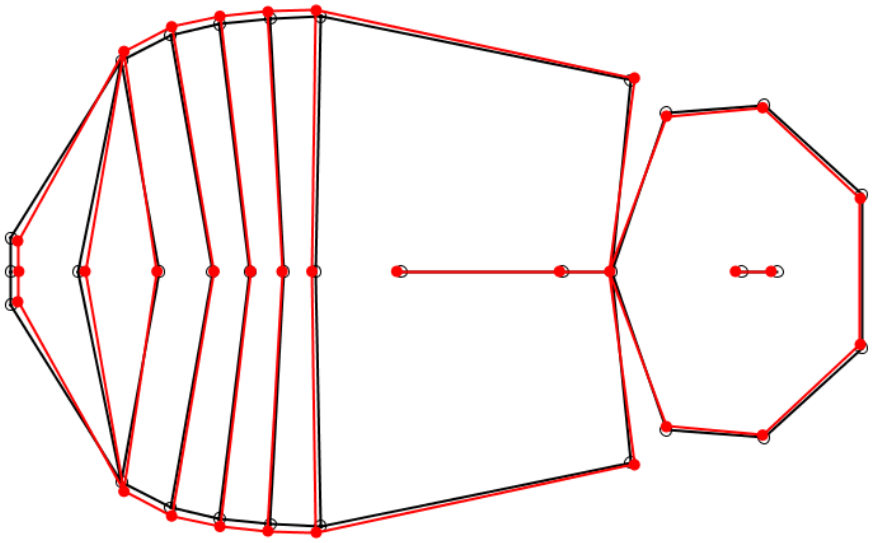
Supplementary Figure 4. Plot of canonical variate 1 against canonical variate 2 for the shape of the male mesotibia ventral dataset. The confidence ellipses show 95% probability from equal frequency



Supplementary Figure 5. Positive (blue) and negative (red) shifts in canonical variate 1 for the ventral analysis of the male mesotibia ventral dataset. These represent the most extreme values from the data for each variate, showing how the scores vary.

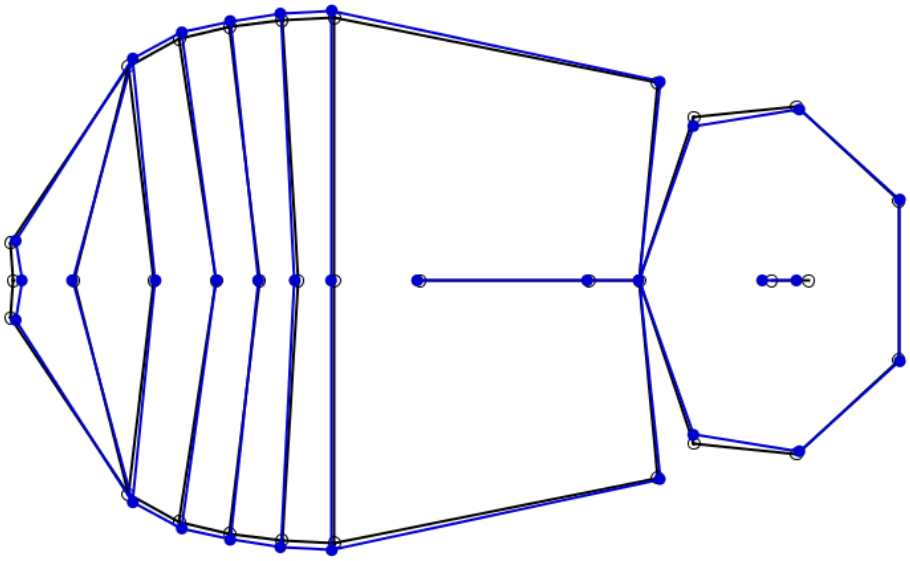


CV1

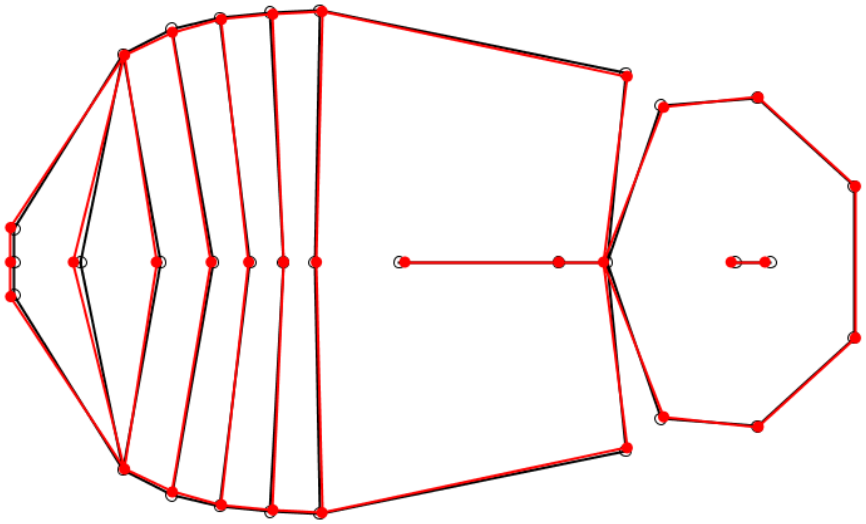


CV1

Supplementary Figure 6. Positive (blue) and negative (red) shifts in canonical variate 2 for the ventral analysis of the male mesotibia ventral dataset. These represent the most extreme values from the data for each variate, showing how the scores vary.



CV2



CV2

Chapter 6: DNA Sampling of rare and / or elusive species

Microsatellite sequences developed for *Gnorimus*

TGTAAAACGACGGCCAGT
pig - gttt

NC-SSR-CT1-F	TGTAAAACGACGGCCAGTGCAGCAACAAATGGTTTCAA
NC-SSR-CT1-R	gtttCAAACTCGAAAGCAAGATATTACC
NC-SSR-CT2-F	gtttAGTGGGGCTTTCATACGATG
NC-SSR-CT2-R	TGTAAAACGACGGCCAGTCGCATACTGAGAAACACACA
NC-SSR-gT-F	gtttCAGGTGTATCTGTCGATGGTAAA
NC-SSR-gT-R	TGTAAAACGACGGCCAGTCAAAACATGTGCCCCACTTA
NC-SSR-ag-F	gtttTGCATCTTTGTAAGAGAGAG
NC-SSR-ag-R	TGTAAAACGACGGCCAGTAGCAGAAGACGGCATACGAG
NC-SSR-at-F	gtttTGCAGCAATACCTTCTTGTTAAT
NC-SSR-at-R	TGTAAAACGACGGCCAGTTTTCAGTTAATTGTGTATTGCATTC
NC-SSR-aat-F	TGTAAAACGACGGCCAGTCAGAGTGGAAATCACCCATTG
NC-SSR-aat-R	gtttGGATGCGAAGGCATGAATTA
NC-SSR-aat2-F	TGTAAAACGACGGCCAGTTTTCTTAAGTCGTCTGCGACA
NC-SSR-aat2-R	gtttGGATGCCAAGGCATGAATTA
NC-SSR-aat3-F	TGTAAAACGACGGCCAGTATCGATATTTACCATTATGATT
NC-SSR-aat3-R	gtttAACCGGTTAATCTGTTGGACA

Chapter 7: Species Distribution Modelling of Saproxylic Scarabs

Reference list for species distributions:

Aberlenc 2006; Alekseev & Nikitsky 2008; Ariana et al. 2011; Arinton 2011; Arnone et al. 2009; Audisio et al. 2009;Avgin et al. 2014; Baraud 1992; Barševskis 2001; Barševskis et al. 2004; Bekchiev & Zlatkov 2010; Bellmann 2002; Brelih et al. 2010; Bunalski 2000; Bunalski et al. 2012; Bussler et al. 2005;Colón 2003; Enyedi 2006; Franc 2002; Georgiev & Doychev 2010; Gokturk & Mihli 2015; Guéorguiev et al. 2011; Kadej et al. 2013Kasic-Lelo et al. 2010; Koren et al. 2011; Král & Malý 1993; López-Colón & Blasco-Zumeta 1997; Nádai & Merkl 1999; Pivotti et al. 2011; Róbert 2004; Rosa 2005; Rozner & Rozner 2009; Sabatinelli 1977; Şenyüz & Şahin 2009; Stefanelli et al. 2014; Tauzin 2004b; Tauzin 2005a; Tauzin 2005b; Telnov et al. 2005; Thomaes et al. 2008; Thomaes et al. 2015; Voolma & Randveer 2003; Vrezec 2007; Young 1989; Zabransky 1998