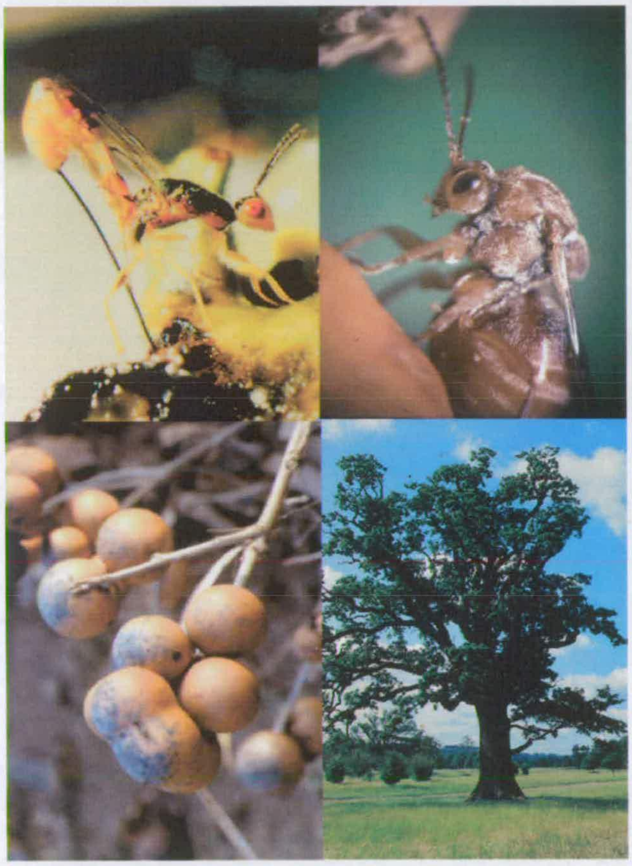
The evolution and ecology of oak gall wasp communities

Alex Hayward



Thesis submitted for the degree of Doctor of Philosophy

The University of Edinburgh

2005

Declaration

This thesis has been composed by me and is the result of my own work. It contains no work done in collaboration except where stated otherwise. The text does not exceed 100,000 words. No part of this thesis has been submitted to any other University in application for a higher degree.

[redacted]

Abstract

This thesis uses DNA sequence data to investigate the evolutionary history and ecology of the insect communities centred on oak cynipid gall wasps (Hymenoptera: Cynipidae). European oak gall wasps show within-species genetic subdivision that can be attributed to confinement to refugia in past ice ages, and which indicates an eastern origin for each species. A major aim of this thesis was to determine whether parasitoids attacking oak gall wasps show similar patterns, compatible with a 'community phylogeography' spanning trophic levels. Mitochondrial sequence data for three chalcid parasitoids found strong evidence of refugial differentiation in one species (*Megastigmus stigmatizans*, Torymidae) but sampling in two others (*Megastigmus dorsalis* Torymidae and *Eurytoma brunniventris* Eurytomidae) provided inadequate geographic resolution. Two species, *M. dorsalis* and *E. brunniventris*, showed patterns consistent with an eastern origin, while data for *M. stigmatizans* were unresolved basally. Western palaearctic oak subspecific richness is also highest in Asia Minor, suggesting that this region may represent a cradle of within-species diversity across three trophic levels in the same community.

Human dispersal of the Turkey oak, *Quercus cerris*, has triggered range expansion from southern Europe into northern Europe by 8 oak gall wasp species over the last 500 years. All of these invaders have recruited communities of parasitoids in their invaded range. I aimed to use mitochondrial sequence and nuclear microsatellite and allozyme data to distinguish among alternative hypotheses for the recruitment of these natural enemies. Lack of nuclear markers restricted analyses to haplotype sequence data, which supported the following conclusions. First, invading gall wasps have allowed parasitoids previously trapped in glacial refugia to expand their ranges northwards. Human planting of Turkey oak has thus triggered a trophic cascade of range expansion. Second, there is a high probability that galls of *Andricus kollari* imported to southern England in the nineteenth century introduced parasitoids alongside the gall wasps.

Traditional morphological taxonomy has long been known to miss important biological diversity. The three parasitoid species named above attack a diversity of oak galls, and an important question is whether each currently recognised parasitoid species is a genuine generalist, or in fact consists of discrete genetic lineages, each attacking a subset of the known host range. Such subdivision is relevant to debates on the adaptive significance of oak gall phenotypes, and could have a major impact on the structure of foodwebs centred on gall wasp hosts. Mitochondrial sequence data revealed no parasitoid population subdivision in response to the oak taxon on which host galls develop, location of host galls on the host, or aspects of gall morphology (spininess, stickiness, and internal gall structure). Limitations to the resolution possible with the markers used are discussed, and the existence of significant host-associated population structure cannot be discounted. *Eurytoma brunniventris* showed significant intraspecific structure unassociated with host traits, compatible with the existence of 4 morphologically indistinguishable subspecies or species. This finding is consistent with published doubts as to the status of species in this genus.

Difficulties associated with identification of some adult parasitoids, and almost all immature stages, triggered the final work in my thesis, on the use of molecular barcodes to identify parasitoids. I discuss the rationale behind the use of DNA sequence data to identify taxa, and show that the approach has considerable promise in oak cynipid communities. Nevertheless, some generalisations on levels of between-species sequence divergence made by leading proponents of molecular barcodes may not apply specifically to chalcids.

Acknowledgements

I feel very grateful to have had the opportunity to study for a PhD, particularly in such a friendly environment as the Ashworth labs.

Firstly, I have to thank my supervisor Graham Stone for all his support. His generosity has been far beyond the responsibilities of a supervisor, and he has always dealt with any problems with humour, good advice and a small libation of hard liquor of some variety. In addition, I am indebted to his Alex for putting up with all the encroachments upon their family time, particular around the date of the submission of this thesis.

I would also like to thank all the staff at Ashworth who have made life easier during my PhD, particularly David and Jayne in accounts, Jill and Andy in the sequencing department, Billy in Stores, and Carol, Kath, Rosy and Pat in the office.

The following have also provided valued assistance at various stages during my PhD:
George Melika, Zoltan Acs, Dick Askew, Gyorgy Csoka, Ebrahim Sadeghi, Juli Pujade Villar, Jose-Luis Nieves Aldrey, Josephine Pemberton, Jean-Yves Rasplus, Johan Nylander, Antonis Rokas, James Cook, Karsten Schonrogge, Olivier Plantard, Ahmed Raza, Lucy Weinert, and anyone else I have neglected to mention.

I would like to thank all my friends and colleges from the Ashworth Laboratories, and give a special thank you to Zee Sulleyman for being my trusty field assistant. I also wish to thank past and present flatmates for keeping life interesting, and particularly my original flatmates Becky Coe and Jules Hernandez for their friendship.

Most of all I am indebted to my parents and family for their unending support, I dedicate this thesis to them.

Finally I would like to thank the Natural Environment Research Council who funded my PhD position, and the James Rennie Bequest, the Genetics Society and the Royal Entomological Society who all provided additional funding for travel.

Contents

Chapter 1

General introduction

[A modified version of this chapter appears as a review article entitled "Oak gall wasp communities: ecology and evolution" in: *Basic and Applied Ecology*, **6**(5), 435-443, by Hayward A, Stone GN (2005)]

1.1 Introduction

Oak cynipid galls are microcosms of intense ecological activity and their study holds much promise for diverse fields in evolution and ecology. They commonly contain at least three trophic levels (gall tissue; gall wasps and inquilines; parasitoids and predators), and much research on them has focussed on their value as model tritrophic systems. Interactions between these levels can be complex and hyperparasitoids and superparasitoids also occur in oak gall communities. Processes structuring and driving the evolution of oak cynipid communities remain areas of ongoing debate. This Chapter provides a brief description of the current understanding of oak cynipid communities and examines key areas requiring further research. It begins with a description of the components of oak cynipid communities, and then provides a summary of what is known about their phylogeography within Europe. The ecology and evolution of the communities are then examined with a focus on issues of relevance to the parasitoids. The Chapter ends with an outline of the questions addressed in this thesis.

Although this thesis focuses on oak gall wasp communities, the issues dealt with are of wider relevance outside the group. Chapters 3 and 4 address the issues of community phylogeography, the possible existence of host-races in apparently generalist parasitoids, and the tracking of parasitoid species following a recent range expansion by a subset of their hosts. The study of phylogeography has largely been restricted to individual species, and a need exists to extend the discipline to include the analysis of whole communities. This is necessary in order to identify whether communities in different geographic locations represent truly independent data points, or alternatively whether they are linked by phylogeny. This is an important consideration when analysing the structure of communities in the search for possible rules of community assembly. A further application is the identification of areas that correspond to centres of diversity across trophic levels for the assessment of conservation priorities. The existence of host-races, or population subsets associated with different host taxa, is of importance for the analysis of foodweb structure. True generalists link many hosts together and have the potential to mediate apparent competition among shared hosts, whereas the existence of host races implies a set of smaller subcompartments within which only a subset of hosts interact. Currently, little is known about the way in which parasitoids locate and discriminate between hosts, but this is of significant importance for their use as successful biocontrol agents. Tracking the origin and colonisation pattern of invading species is of importance for predicting changes in species distributions associated with climate change, evaluating the potential consequences of the introduction of biocontrol agents from different physiological regimes, and controlling the spread of invasive species. Chapter 5 involves an examination into the evolution of trophic shifts in parasitic wasps. Shifts between entomophagy and various forms of phytophagy are apparent within several parasitoid families. The ease and direction of such transitions is unknown, but a more complete understanding could provide important clues regarding the evolutionary pressures acting upon parasitoids. Chapter 6 consists of an assessment into the utility of DNA barcoding in oak gall wasp communities. Molecular taxonomy and the use of DNA barcodes hold much promise, but the specifics of their application are currently hotly debated.

1.2 Components of oak cynipid communities

The oak cynipids form the Cynipini, one of six tribes in the gall wasp subfamily (Hymenoptera: Cynipidae: Cynipinae). With approximately 1000 species in 27 genera the Cynipini is by far the largest tribe in the subfamily (Ronquist 1999), and the true figure may actually be much larger (Nordlander 1984). Oak cynipids develop as obligate parasites in galls that they initiate on *Quercus* and related genera in the family Fagaceae.

Oak cynipid galls also support rich communities characterised by a high diversity of generalist natural enemies, predominantly chalcid hymenopteran parasitoids, and a distinct set of inquilines (see below) (Stone *et al* 2002). Most of these are exclusively restricted to oak cynipid hosts, making it possible to consider the interactions that occur within oak galls in isolation and their communities as effectively closed (Askew 1984, Braune 1992, Gilbert *et al* 1994, Stone *et al* 2002). The high diversity present in oak cynipid communities is due in part to the abundance and diversity of inquilines, which act as additional host resources for parasitoids (Schönrogge *et al* 1994, Schönrogge *et al* 1996, Stone *et al* 2002). The term 'inquiline' derives from the Latin *inquilinus* meaning tenant or lodger, and is used to refer to an organism that lives commensally in the home of another species. Inquilines of oak cynipid galls include gall midges, moths and beetles, but by far the most numerous in terms of both number of species and abundance are the cynipid inquilines (Csóka *et al* 2005). These belong to the tribe Synergini, which are thought to be derived from gall-inducing cynipids within the basal paraphyletic tribe Aylacini (Ronquist 1999, Nylander 2004). There are 128 species recorded from oak cynipid galls, the vast majority of which are specific to these host galls and belong to the genera *Ceroptres* (Pujade-Villar & Nieves-Aldrey 1993, Nieves-Aldrey 2001), *Saphronecrus* (Pujade-Villar & Nieves-Aldrey 1990), *Synergus* (Nieves-Aldrey & Pujade-Villar 1985, 1986, Ritchie 1984), and *Synophrus* (Stone *et al* 2002). Relationships among the inquiline genera remain unclear. While morphology-based analyses suggest monophyly (Ronquist 1994, 1999, Vårdal 2003, Vårdal *et al* 2004), the most recent sequence-based analyses suggest 3 separate evolutions of inquilinism from gall induction (Nylander 2004, Nylander *et al* 2004). The Synergini have lost the ability to induce their own galls, but remain able to induce the development of larval chambers lined with nutritive tissue inside the galls of their hosts (Shorthouse 1973, Ronquist 1994). In oak cynipid galls inquiline species either induce chambers in the outer gall tissues, or in the host larval chamber. Although the Synergini are apparently wholly phytophagous, species that initiate chambers in the host larval chamber often cause the death of the host, and are termed lethal inquilines. Lethal inquilines can result in local extinction of the gall former (Washburn & Cornell 1981) and cannot be considered commensals in the true sense, instead being better classed as natural enemies (Wiebes Rijks & Shorthouse 1992).

All known parasitoids that attack cynipid hosts are wasps, including members of the Ichneumonidae, Braconidae, and most importantly in terms of species richness and level of mortality inflicted, the Chalcidoidea (Askew 1984, Wiebes-Rijks & Shorthouse 1992, Stone *et al* 2002). Six chalcid families attack European oak cynipid communities: Pteromalidae, Eurytomidae, Eupelmidae, Eulophidae, Ormyridae, and Torymidae (Askew 1965, 1984, Nieves-Aldrey 1983a, 1983b, 1983c, 1984a, 1984b, Pujade-Villar 1989, 1992a, Plantard *et al* 1996, Plantard & Hochberg 1998, Schönrogge & Crawley 2000, Schönrogge & Askew, in press). These parasitoids feed either externally (ectobionts) or internally (endobionts) to the host, either as the host continues developing (koinobionts), or when development has been arrested by stinging prior to oviposition (idiobionts). The majority are solitary idiobiont ectoparasitoids whose larvae feed suctorially (Askew 1984, Csóka *et al* 2005). There are relatively few endoparasitoids in cynipid galls, but examples include *Pediobius lysis* and *Sycophila biguttata* (Askew 1984). Gregarious parasitoids, in which many wasps develop on a single host, are also rare (examples include members of the genus *Baryscapus* (Schönrogge *et al* 1995)). Some species such as *Torymus cyaneus* and *Eurytoma brunniventris* display facultative phytophagy and are able to feed on gall tissue in addition to host larvae (Askew 1984). Indeed, *E. brunniventris* is able to develop to maturity by feeding entirely on gall tissue, although viability of the adult is low (Askew 1999). Some ectoparasitoids such as *T. auratus* and *E. brunniventris* can act as facultative hyperparasitoids and cannibals, and will attack any larva they encounter.

Parasitoids vary in their host specificity and oak cynipid communities are characterised by large numbers of polyphagous species and limited numbers of parasitoids with a narrower host range (Askew 1984). However, almost all parasitoids reared from cynipid galls are specific to these hosts. Exceptions include *Eupelmus urozonus* and *E. vesicularis*, which are also secondary parasitoids of a range of Lepidoptera, Diptera and Coleoptera larvae (Noyes 1998). These species also have the broadest range of cynipid hosts and are able to develop as ectoparasitoids on a great range of other parasitoids (Schönrogge & Askew, in press). Individual parasitoid species generally attack only galls induced by members of a single cynipid gall wasp tribe, with the exception of *E. urozonus, S. biguttata, S. flavicollis, Mesopolobus sericeus* and *Aulogymnus skianeuros,* which have been recorded from both oak and rose cynipid (Cynipini and Diplolepidini) (Askew 1984). However, different tribes are commonly attacked by the same parasitoid genera (Schönrogge & Askew, in press). All of the endoparasitoids and some of the larger ectoparasitoids only ever attack the gall inducer rather than the inquilines, and gall inducer and cynipid inquilines often support distinct parasitoid communities.

Oak cynipids are predominantly Holarctic in distribution, with the greatest diversity present in the Nearctic, where there are approximately 700 species (Melika & Abrahamson 2003). However, their communities have been most thoroughly studied in Europe, where community diversity consists of about 200 species of gall former, 104 species of hymenopteran parasitoid (mostly chalcids), and 38 species of inquiline cynipid (see below) (Schönrogge *et al*, in prep; Csóka *et al* 2005). Within Europe, oak cynipid community diversity is highest in the south, with diversity decreasing along a latitudinal gradient northward (Stone *et al* 2001, 2002, Rokas *et al* 2003a).

1.3 Phylogeographic patterns in oak cynipid communities

A major influence on the phylogeography of European species is the legacy of ice age refugia. During the Quaternary (the last 2.4 Myr) repeated glacial periods of about 100 kyr are thought to have restricted the ranges of many European plants and animals to regions (termed refugia) in the south that escaped coverage by ice sheets or tundra (Hewitt 1996, 1999). The last glacial period ended around 10,000-12,000 years ago (Godwin 1975). However, limited gene flow between populations in different refugia has produced a refugial pattern of population structure still visible at the present time for many European plants and animals (Hewitt 1999, 2000, Taberlet *et al* 1998). The impact of glacial cycles on population structure is probably determined by three general factors: (a) the length of time for which populations within refugia were genetically isolated from each other. (b) The extent to which populations originating in different refugia interbreed in more northerly regions during interglacials. (c) The extent to which interbred northern populations contributed to the genetic composition of southern refugia in successive glaciations. The following section provides a brief summary of European refugial structure and post-glacial range expansion for oaks and gall wasps. The section will end with a discussion of the possible impacts that these patterns may have had on the population structure of European oak gall wasp parasitoids. While much has been done on oaks and gall wasps, this study is the first look at phylogeographic structure in the parasitoids.

1 *Patterns of oak refugia*

Oak gall wasps predominantly gall oaks within *Quercus* subgenus *Quercus*. This subgenus is divided into four sections on the basis of DNA sequence data (Manos *et al* 1999): Lobatae, Protobalanus, Quercus *sensu stricto,* and Cerris. Of these, only the sections Quercus *sensu stricto* and Cerris are native to the Western Palaearctic.

Spatial patterns of genetic variation present within species have only been studied in detail for a few western palaearctic oak species of importance here. These are *Q. robur*, *Q. petraea* (Kremer & Petit 1993, Zanetto & Kremer 1995, Dumolin-Lapegue *et al* 1998, Ferris *et al* 1998, Petit *et al* 2002a), and *Q. suber* (Tuomi & Lumaret 1998). On the basis of fossil pollen and genetic data (DNA microsatellites & chloroplast haplotypes) three main refugia for oaks in Europe have been suggested: the Iberian Peninsula, Italy, and the Balkans (Huntley & Birks 1983, Hewitt 1996) (see Figure 1.1). The locations of refugia outside Europe are not so well known. However, current distributions of endemic oaks suggest that Turkey may have had two separate refugia, one in the southwest and one in the southeast. Fossil pollen is subject to the limitation that it does not currently allow discrimination between closely related species within an oak section, or sometimes even between Quercus *sensu stricto* and deciduous members of the section Cerris (van Benthem *et al* 1984).

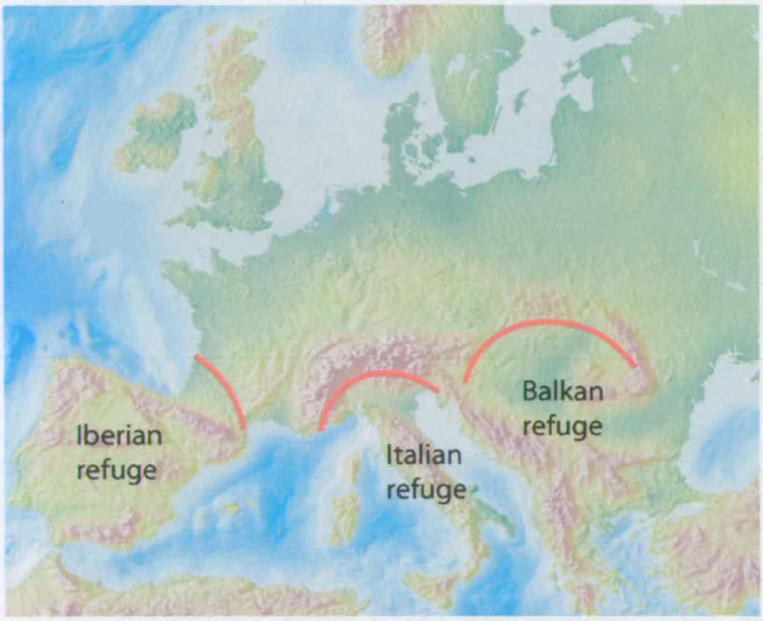


Figure 1.1 Putative refugia for oaks within Europe

Estimates of chloroplast DNA evolution rate (about 0.11% per million years) suggest refugial populations probably diverged at least three million years ago (Ferris *et al* 1998, Tzedakis 1993). Thus, despite the fact that lineages of the same oak species from different refugia probably interbred during interglacials, the resultant 'inter-refugial hybrids' appear to have been unable to disperse southwards again. This is probably due to the rapid fall in temperatures at the end of each interglacial (10-12% over 5-10 years), which left little time for northern populations to disperse southwards. An additional factor may have been the east-west orientation of mountains in southern parts of the Western Palaearctic (the Pyrenees, Alps, Carpathians, Taurus Mountains and Caucasian ranges), since passes at higher altitude may already have been too cold to allow further southwards dispersal by the time northern populations were forced against their north-facing slopes,

Only a very limited number of western palaearctic oaks have managed to escape glacial refugia in the present Holocene interglacial. However, natural range expansion has been augmented in many places by human introduction of species outside their native range (see below). The northwards spread of oaks has been reconstructed mainly from analyses of fossil pollen from lake sediment cores (Huntley and Birks 1983). Pollen data is subject to the limitations discussed above, however, we know from current distributions that records north of glacial refugia probably indicate the spread of *Q. petraea* and *Q. robur*, with minor contributions from *Q. pubescens* and *Q. pyrenaica* in the west. Estimates of the rate of tree spread since the last ice age in Europe range from approximately 500 metres per year (Bennet 1986) to 2000 metres per year (Huntley & Birks 1983).

Ferris *et al* (1998), Dumolin-Lapegue *et al* (1997) and Petit *et al* (2002b) made use of chloroplast DNA haplotypes specific to the three European refugia to investigate the origins of current northern populations of *Q. petraea* and *Q. robur* (Figure 1.2). Both species are now found across Europe as far north as Scotland in the west and the Central Russian Uplands in the east (Jalas & Suominen 1987). Data to date suggest that Iberian populations expanded into western France, the United Kingdom and southern Norway. Italian populations expanded into central Europe, and a clear north-south contact zone between populations derived from the Iberian and Italian refugia now runs through central Europe (Ferris *et al* 1993). Balkan populations expanded over a broad front, and lineages originating there are now found throughout central and western Europe. A feature of populations in northern Europe is a dramatic loss of population genetic diversity within species with distance from refugia (Petit *et al* 2002a). Such a pattern is consistent with the founding of new populations by small numbers of individuals and a 'stepping-stone' model of range expansion.

The only other European species to have expanded its range significantly beyond glacial refugia is *Q. pubescens*, which is found throughout much of eastern and central Europe and southern France to the Pyrenees. Although it is widespread in France, ecological preferences (well-drained lime-rich soils and sensitivity to frost) have restricted it to localised areas, resulting in an increasingly patchy distribution with increasing latitude.

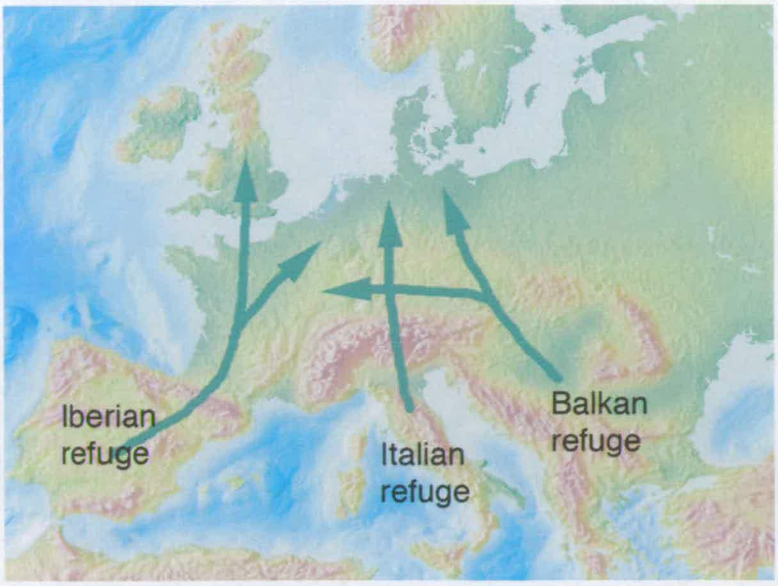


Figure 1.2 Section Quercus oak putative refugia and pattern of post-glacial range expansion (modified from Petit *et al* 2002)

The deciduous members of the section *Cerris* show a longitudinal series of distributions within Europe from *Q. suber* (the cork oak) in the Iberian Peninsula, to *Q. cerris* (the Turkey oak) in Italy, the Balkans, Greece and western Turkey, and *Q. libani* and *Q. ithaburensis* in eastern Turkey. Section Cerris oaks appear to have been unable to escape from their refugia and have a natural distribution limited to the south. *Q. suber*, though predominantly an Iberian species, is also found in Italy. Italian populations show low genetic variation that represents a subset of that found in Iberia, and it is possible that most or all of the Italian populations result from ancient human planting. Post-glacial range expansion in this species is restricted to isolated populations in south-western France.

2 *Patterns of gall wasp refugia*

Phylogeographic patterns have been investigated within Europe for several oak cynipid species in the genera *Andricus* and *Biorhiza,* using mitochondrial haplotypes (derived from the cytochrome *b* gene) and allozyme markers. The pattern of oak gall wasp refugial structure differs from that of the oak tree hosts (described above), which includes a separate refugium in each of Italy and the Balkans (Brewer *et al* 2002, Petit *et al* 2002)(see Figure 1.1). In cynipid species, the pattern of refugial population structure suggests the existence of at least three refugia within Europe: an Iberia refugium, a central European refugium consisting of Italy and the Balkans, and a Turkish refugium (see Figure 1.3). What is labelled as the Turkish refugium in Figure 1.3 may actually consist of multiple refugia as discussed above for the oaks. Oak gall wasp refugia are naturally delineated by the Alps, the Tatras and the Carpathians in central Europe, and in Iberia apparently by the northern most distribution limits of *Q. suber*. Populations sampled from Iberia usually appear most derived and form a single discrete monophyletic clade (Stone *et al* 2001, Stone *et al* 2002). In contrast, populations from the Balkans and Italy do not constitute refuge-specific clades but form a paraphyletic group, encompassing the Iberian clade. Turkish populations are divided, with haplotypes from some individuals sampled from Turkey falling in the Italian/Balkan clade, and others forming a basal paraphyletic group encompassing all other groupings (Stone *et al* 2001, Rokas *et al* 2003a). The relative genetic isolation of Iberian populations suggests that divergence between the common ancestor of all Iberian sequences and that of Italian and Balkan sequences is substantial, implying that Iberian populations have been isolated from other European populations for a relatively long time. Coalescence times for Iberian lineages are far more recent than those elsewhere in Europe and intraspecific diversity is low, suggesting that Iberian populations may also have a recent history of genetic bottlenecks (Stone *et al* 2001, Rokas *et al* 2003a, Atkinson *et al*, in press).

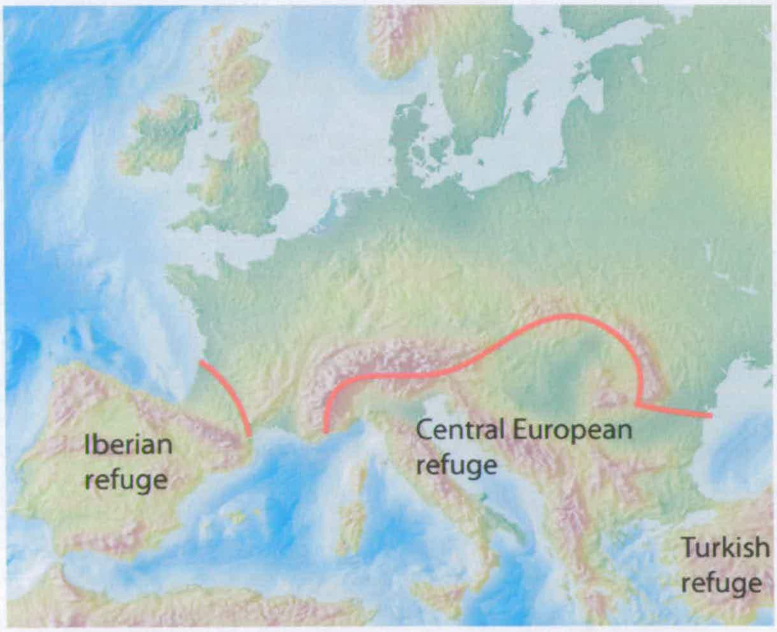


Figure 1.3 Putative refugia for oak gall wasps within the western palaearctic

The refugial pattern of population differentiation uncovered for the cynipids is accompanied by a longitudinal decrease in genetic diversity from Anatolia in the east to Iberia in the west. This is compatible with the idea that oak cynipids as a guild may be ancient invaders to Europe from a centre of origin in the east. This idea has prompted the formulation of the 'Out of Anatolia' hypothesis for an origin of oak gall wasp communities as a whole in the east (Rokas *et al* 2003a, Atkinson *et al*, in press).

Although there have been investigations into the phylogeography of cohorts of species (Taberlet *et al* 1998, Hewitt 1999) the analysis of 'community phylogeography' - tracing the geographic origin and subsequent spread of communities – is new. In contrast to the cohorts of species reported by Hewitt and Taberlet *et al*, the sort of communities that must be considered in order to pick up community phylogeography are tightly linked and ecologically closed communities. The lack of studies on 'community phylogeography' may be due to the difficulties involved in amassing data for whole series of species that exist within a community across a wide geographical scale. Oak gall wasps are well suited to examining the question of community phylogeography due to the closed nature of their communities, and the expected tight association between members in the community as a result of this. The groundwork for examining the community phylogeography in oak gall wasp communities is already in place (Csóka *et al* 2005) and the system offers a promising study ground with which to launch the field.

3 *Impacts of oak distributions on gall wasp refugial structure*

Most oak cynipids are cyclically parthenogenetic (heterogonous), with a springtime sexual generation and a summer/autumn parthenogenetic generation. In Europe two genera (*Andricus, Callirhytis*) include species that gall different oak host taxa in their two generations. Host-alternating *Andricus* gall oaks in the section *Cerris* in their sexual generation, and oaks in the section *Quercus* in their parthenogenetic generation. The phylogeography of two oak host alternating cynipid species, *A. kollari* and *A. quercustozae,* has been studied in particular detail (Stone *et al* 2001, Rokas *et al* 2003a). Both display a particularly marked population divide between populations sampled in Iberia up to southern France just north of the Pyrenees, and those sampled in the rest of Europe. This dividing line corresponds exactly with the dividing line between two region specific section Cerris oak hosts, *Quercus suber* and *Q. cerris* (see Figure 1.4)(Pujade 1992b, 2003, Stone *et al* 2001), suggesting that rarity of host shifts between these two oaks may maintain the genetic isolation of Iberia. This impact of host specificity is seen more generally in the failure of many host alternating gall wasps to expand from central Europe into Iberia (Rokas *et al* 2003a, Atkinson *et al*, in press). Such biological structuring processes may well have impacts extending to other components of gall wasp communities.

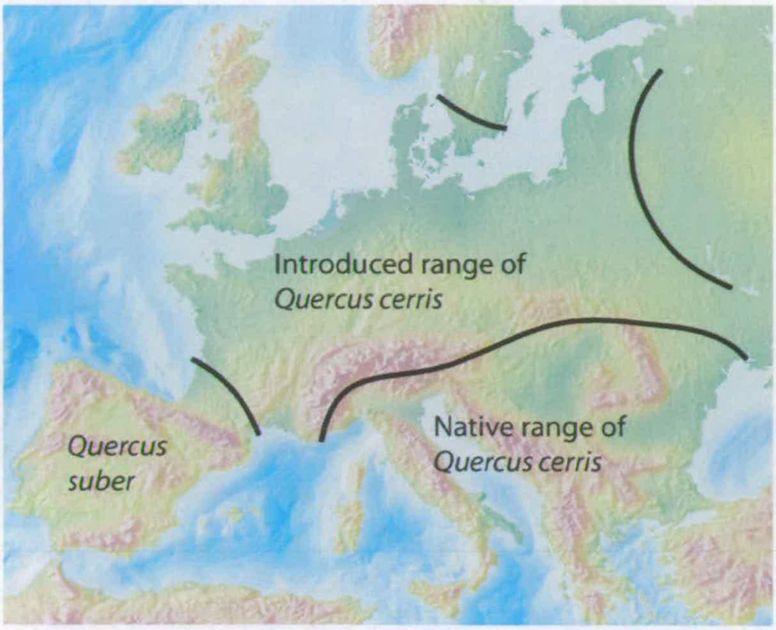


Figure 1.4 Native and introduced range of *Q. cerris* and *Q. suber* (Modified from Stone *et al* 2001). The northern limit to *Q. suber* represents the southern limit to the introduced range of *Q. cerris*.

4 *Impacts of recent human dispersal of Q. cerris*

As stated above, section Cerris oaks have not been able to escape their refugial distribution in the south. However, over the last 400 years, *Q. cerris* has been planted extensively north and west of its native range and has since become naturalised across much of northern and central Europe (Jalas & Suominen 1987)(see Figure 1.4). In parts of its range it has become so successful that it has acquired pest status in native oak woodlands. This anthropogenic range expansion has facilitated a recent range expansion by at least 8 host alternating *Andricus* species (Stone & Sunnucks 1993, Schönrogge *et al* 1998, Kwast 2001, Stone *et al* 2001, Walker *et al* 2002).

Phylogenetic analyses on gall wasp populations in the invaded range support a central European origin (see Figure 1.5, Stone & Sunnucks 1993, Stone *et al* 2001, Rokas *et al* 2003a) and range expansion is associated with qualitative and quantitative changes in the community associated with the galls (Schönrogge *et al* 1994, Stone *et al* 1995). Community diversity and the total mortality inflicted on the gall wasp by parasitoids both decrease with distance from the invasion origin. Strikingly, over time, communities and mortalities in the invaded range converge on those in the native range, suggesting a strong impact of deterministic host-specific factors on parasitoid community composition (Schönrogge *et al* 1994, 1995).

Figure 1.5 Hypothetical pattern of range expansion for three of the invading species: *A. lignicola*, *A. corruptryx*, *A. quercuscalicis* (Stone & Sunnucks 1993, Csóka *et al* 1998). The same invasion route may have been used by *A. kollari*.

5 *Potential impacts on parasitoids*

It is not clear what the refugial structure of oak gall wasp parasitoid populations will be. Parasitoid refugia could potentially mirror either gall wasp refugial structure, with Italian populations consisting of a subset of the Balkans genetically, or show an oak-like structure with Italy and the Balkans distinct. Alternatively, refugial structure might be absent altogether. If the "Out of Anatolia" hypothesis for the origin of oak gall wasp communities is correct, we would expect to see a concordant phylogeographic pattern in the parasitoids and their gall wasp hosts. However, an alternative possibility for community development is that members have discordant origins, with different species having been recruited by host shifts from existing communities. Individual parasitoid species may show separate phylogeographic origins, such as "Out of Iberia" for instance. It is also not clear what pattern of recruitment will be evident for parasitoid populations present in invading hosts in Northern Europe. As stated above, parasitoids are present in these hosts, however, they could have been recruited from a number of possible origins. These origins include host shifts from native communities (if the parasitoid is native to the invaded range) and pursuit from one or more of the putative refugia. Given the closed nature of gall wasp communities, pursuit of hosts is quite likely *a* *priori*, with generalists free to follow the first gall wasps, and specialists restricted until their hosts make the move. In addition, although the gall wasp hosts have been shown to originate from the Balkan refugia, parasitoids from a different refugium may have 'hitchhiked' onto invading hosts. In the UK introduction represents a further possibility as a consequence of the mass importation of *A. kollari* galls to the country during the 19th century.

1.4 Ecology of oak cynipid communities

In addition to looking at spatial population structure an important issue in parasitoid community biology is host-mediated population structure. Here subsets of populations of individual parasitoids are potentially divided among alternative hosts. This section provides a background to ecological interactions in oak galls and explains why consideration of host mediated population structure is appropriate for this study.

Oak cynipid communities are highly diverse and possess the greatest community richness amongst the cynipid gall wasps (Askew 1984, Csóka *et al* 2005). The fact that a high diversity of gall formers and parasitoids often occur in sympatry, and the abundance of generalist parasitoids in oak cynipid communities, raise two important questions:

(A) Do sympatric gall wasp species compete, either directly or indirectly? Direct bottom-up competition could occur for limiting resources such as oviposition sites, or for nutrients where multiple galls develop on the same plant part. Evidence for both types of direct competition exists (Atkinson *et al* 2002, 2003, Gilbert *et al* 1994), but the frequency of such effects in oak galls remains unclear. Indirect competition could arise as a result of top-down impacts imposed by shared natural enemies. Theoretical treatments of such indirect effects, termed apparent competition, predict that a shared natural enemy will drive all but one of its prey hosts extinct (Holt 1977, Holt & Lawton 1994). This prediction is also supported in simple experimental systems, and apparent competition has recently been shown to be an important factor in real communities (Settle & Wilson 1990, Bonsall & Hassell 1997, Morris *et al* 2004). Most oak cynipids share natural enemies (Askew 1984, Askew & Shaw 1986, Hails & Crawley 1992, Schönrogge & Askew, in press), so how then is host cynipid diversity maintained?

(B) How do multiple parasitoid species competing for the same resource coexist without driving each other extinct? The great number of ecologically similar parasitoid species with overlapping host preferences seems inconsistent with the theory of competitive exclusion.

Clearly mechanisms exist that permit the sympatric coexistence of high diversities of oak cynipids and parasitoids. One answer to the above questions is that coexistence may depend upon the partitioning of parasitoid species or populations within species by gall traits. Askew (1984) states that the main factors influencing parasitoid community composition are:

1. Host oak taxa
2. Gall location on the host plant
3. Gall structure: galls of a similar size that develop on the same plant part at the same time of year have similar communities (Askew 1961, 1984).
4. Phenology.

These potential structuring factors are discussed in turn below:

Host oak taxa.

Very little is known about the cues used by oak cynipid parasitoids to locate their hosts, but host oak taxa may be an important factor. Parasitoids are known to use plant volatiles as a means of detecting hosts (Olson *et al* 2003, Birkett *et al* 2003, Pettersson 2001). Galls are composed entirely of plant tissue, and may have characteristic volatile signatures used by parasitoids as foraging cues. Indeed, preferences for volatiles from different gall species have been demonstrated in oak cynipid parasitoids (Schönrogge *et al*, unpublished data), and these may influence their attack rates on different gall species in the field.

Location

As a group oak cynipids induce galls on buds, stems, catkins, leaves and roots, and it is possible that host-associated cues used by parasitoids vary with gall location. Since odour, shape, and size are all likely to be important during the initial stages of parasitoid searching (Stone & Schönrogge 2003), gall location may play an important role in structuring parasitoids.

Structure

Galls of similar size but divergent morphologies often share very similar associated parasitoid communities (Schönrogge & Askew, in press). This observation appears to conflict with theories suggesting that enemy escape may have driven the evolution of novel gall morphologies (see 'Evolution of oak gall communities' below). However, as discussed above changes in size and structure during development of a given gall type may constrain when a given parasitoid can attack the gall. Many galls are large enough when mature that only parasitoids with very long ovipositors can attack the gall former. There are periods of time, termed 'windows of opportunity' in which parasitoids with shorter ovipositors are able to attack a gall former before it becomes physically out of reach (Craig *et al* 1990, Stone & Schönrogge 2003). Should these windows occur at different times for different gall species, they could drive selection for synchronisation of gall development and parasitoid flight periods and hence generate temporal partitioning of parasitoids. It is quite possible that patterns of development in other gall traits, such as sticky coatings or a covering of hairs, may have similar community impacts.

Phenology

This factor combines 2 closely interrelated issues which I group together - the phenology of gall development, and the quantity of resource provided by the growing gall wasp larva within the gall. Phenology is also partially confounded with host oak section, since the different sections have different phenologies that dictate when gall induction, and hence parasitism can occur. Most oak cynipids have two generations a year, a spring sexual generation and a summer/autumn parthenogenetic generation. The gall structures and locations of the two generations are usually very different. It is possible that such alternation of gall phenotypes was initially favoured by selection because it allowed escape from mortality inflicted by enemies able to recognise only one of the two phenotypes in the lifecycle (Pujade-Villar *et al* 1999). Support for the enemy-escape hypothesis is provided by the fact that there are relatively few cases of a single parasitoid attacking both generations of the same host. Where a parasitoid attacks both generations (for example, *Mesopolobus amaenus* attacks both generations of *Andricus quercuscalicis*) it may well attack different hosts (gall former or inquiline) in the two generations (Stone *et al* 1995). However, both generations now commonly experience heavy parasitoid attack, albeit by different parasitoid faunas.

A more important difference between generations may be resource availability, with spring generations and their galls usually smaller or much smaller than those of their corresponding summer generations. Without specific adaptations (see below), resource limitation prevents large idiobiont parasitoids from attacking galls early in their development, and host eggs or young larvae only provide enough resources for small idiobiont parasitoids or koinobionts. Larger parasitoids such as *M. stigmatizans, Ormyrus nitidulus* and *Torymus erucarum* only have one generation a year, and may be excluded from sexual generation spring galls because they provide insufficient resources for successful development of the parasitoid larva. These species do not eat gall tissue, and in contrast to some discussed below, can only feed on the host resource available at the time of attack (Stone *et al* 1995). For such parasitoids, resource maximisation requires attacking their hosts late in development, and this group includes the species with the largest ovipositors seen in oak gall communities (Stone *et al* 2002). A similar pattern is also apparent at the seasonal level, and three *Torymus* species with two generations a year display seasonal polymorphism in ovipositor length (with longer ovipositors for the summer/autumn generation) (Askew 1965). An alternative strategy to possessing a long ovipositor is to lay eggs early in gall development that lie dormant in the host-larval chamber. This permits attack when the gall is small but allows the host larva to be harvested later when it provides a larger resource. Such a strategy is displayed by *Sycophila biguttata* (Schönrogge *et al* 1995, Csóka *et al* 2005). Alternatively, *Torymus cyaneus* attacks early, but its eggs hatch quickly and the larva feeds initially on gall tissue, switching to the cynipid larva later.

Although the impact of gall traits on parasitoid population structure remains unclear, an intriguing possibility is that apparently generalist parasitoids in fact consist of morphologically indistinguishable host-races or ecotypes associated with particular subsets of the available gall community. The existence of host-races would reduce the potential for apparent competition between gall wasps and for potential resource competition amongst parasitoids. The evolution of specialists from generalists is predicted by mosaic models of coevolution (Thompson 1999), and is also predicted by food-web theory to increase foodweb stability (Kondoh 2003).

The extent to which these factors are predicted to drive parasitoid community structure depends on other factors including the impact of bottom up and stochastic effects on the relative abundance of different hosts (Stone & Schönrogge 2002). Intraspecific variation in gall traits allows the teasing apart of these complex effects (*eg* Plantard *et al* 1996, Plantard &

Hochberg 1998) for single species, and the major challenge remains examining the same effects at a community level. Work in this field is at an early stage, but initial results by Acs *et al* (2001) support population subdivision associated with different oak and gall wasp hosts in the parasitoid *Eurytoma brunniventris*. These results provide a tantalizing suggestion of the possible existence of host races in oak gall wasp parasitoids. However, the sample size used in this study was very small and further examination is necessary. *E. brunniventris* is one of three oak gall wasp parasitoid species examined in larger sample sizes across host oak taxa, gall locations and gall morphology in Chapter 4 of this thesis with the aim of clarifying this issue.

1.5 Evolution of oak gall communities

An outstanding question in the evolution of oak gall communities is what selective forces drive the diversification of gall phenotypes (Cornell 1983, Price *et al* 1987, Stone & Schönrogge 2003). Each species of oak cynipid produces a gall of characteristic morphology, many of which display great complexity (Stone & Cook 1998). Gall traits are under the control of the gall wasp and thus represent the extended phenotypes of gall inducer genes (Stone & Schönrogge 2003), and any gall structure that enhances gall wasp survival should be under strong selection. The natural enemies of oak cynipids can only reach their hosts by oviposition through gall tissue and the current favoured hypothesis, the enemy hypothesis, states that gall diversification has been driven by selection to avoid mortality imposed by these enemies. Cynipid galls possess a wide variety of potentially defensive traits such as surface spines and hairs, increased hardness and thickness of gall tissue, the presence of false chambers and the secretion of ant-attracting nectar (Stone & Schönrogge 2003). Selective advantage for these gall traits is also supported by the fact that several of them have evolved convergently (Cornell 1983, Price *et al* 1987, Stone & Cook 1998). However, the diversity of generalist parasitoids able to attack divergent gall morphologies appears to contradict the enemy hypothesis. Galls provide some degree of protection against nonspecialist predators and pathogens, but are far from being enemy-free space. In fact oak cynipids often experience a higher rate of mortality due to natural enemies than free-feeding insects (Price *et al* 1987, Stone & Schönrogge 2003).

One possible explanation is that gall traits represent the 'ghost of parasitism past', in that they were once useful as means against parasitoid attack, but have since lost their utility due to the subsequent adaptation of parasitoids (Price & Pschorn-Walcher 1988, Stone & Cook 1998).

The potential ability of parasitoids to exploit windows of opportunity before protective gall traits have developed, as discussed above, provides one means of circumventing morphological defences. It is also possible that whilst not fully excluding the majority of natural enemy species, gall traits nonetheless reduce the mortality inflicted by specific parasitoids on specific hosts. The probability of successful attack by parasitoids may be reduced in a number of ways. A hairy covering makes it difficult for parasitoids to explore a gall surface with their antennae and insert their ovipositors, while hard galls with thick-walled parenchyma are drilled into more slowly, leading to a higher probability of predation or ovipositor damage than when dealing with soft galls. Multiple larval cells (Jones 1983) and secretion of nectar to recruit ant-guards (Washburn 1984, Abe 1992, Fernandes *et al* 1999) have both been shown to reduce parasitoid attack. However, correlations between variation in many gall attributes and parasitoid mortality remain to be experimentally demonstrated (Stone *et al* 2002).

What evolutionary processes might have led to generation of the high diversity present in oak cynipid communities? Host shifting may be of central importance to the diversification of herbivores (Bush 1969, 1975), and associated parasitoid taxa (Cronin & Abrahamson 2001). Where natural enemy search strategies depend upon host-plant specific cues (Brown *et al* 1995), host shifts may be favoured because they result at least temporarily in escape from natural enemies (Lawton 1986, Price *et al* 1986). If parasitoids make the same host oak shift, radiation of natural enemies may eventually follow, in a process termed 'sequential radiation' by Abrahamson *et al* (2003), perhaps first by host race formation as outlined above, and then later by speciation. Phylogenetic evidence to date suggests that host shifts have been rare in oak gall wasps (Cook *et al* 2002), but similar effects could result from parasitoid specificity in any other aspect of host gall selection, such as gall location on a given host. To our knowledge, there has been no specific study of sequential radiation within oak cynipid communities. A challenge here is that until the basis of host choice by parasitoids is understood, potential cases of sequential radiation will be hard to identify. A number of parasitoid genera have enough species in oak gall communities for sequential radiation to be possible. An example is *Torymus* (Torymidae), a genus which includes species attacking a wide range of concealed hosts, including cynipids on host plants other than oaks. Members of this genus associated with oak galls are closely related, and display a similar basic bauplan. It is possible to imagine a situation where a putative ancestor, originally associated with a different host or mode of feeding entered the oak cynipid system through an oviposition mistake, and subsequently descendant lineages radiated through the available diversity of oak cynipid hosts. Indeed, some in-built propensity for making oviposition mistakes or plasticity in oviposition events could be seen as being of long-term value for the persistence of parasitoid lineages. New polyphagous parasites may be continually being added to communities (Askew, 1984). Askew & Shaw (1986) suggest that parasitoid communities may initially consist of polyphagous parasitoids that evolve increasing specialism over time. Their argument is that although generalists have a larger host range, being a specialist conveys benefits of synchronisation and searching efficiency that offer competitive advantage in interspecific competition.

1.6 Aims and objectives of thesis

This discussion highlights a range of unanswered questions in oak gall wasp community ecology and evolution. This thesis aims to investigate several of the key questions relating to oak gall wasp parasitiods amongst these.

Chapter 2 details the methods used in this thesis. It consists of a review of phylogenetic techniques and justification of the methods applied. Also included is a rational of marker selection, and specific details of the laboratory techniques and computer analyses used.

In Chapter 3 the extent to which different oak hosts exert an effect on the population structure of parasitoids will be examined. As described above, the major divide in oak hosts that occurs in southern France appears to exert a strong impact upon the population structure of the gall wasp *Andricus kollari*. However, it is currently unclear whether host oak effects will be transmitted up a trophic level to parasitoids. The discovery of a similar pattern could have important implications for the understanding of structuring processes that operate in oak gall wasp communities.

In Chapter 4 the phylogeographic histories of several parasitoid species are examined in order to establish whether these are concordant with what has been found for their gall wasp hosts. In addition, the existence of host races is examined with reference to host oak section, gall location, and gall phenotype. As described above, the existence of host races in oak gall wasp parasitoids has important connotations for our understanding of the structuring processes that occur within the communities. Also addressed in this chapter is an analysis of the origin of parasitoids attacking invading oak gall wasps in Northern Europe. The consideration of the above questions are combined into one Chapter because they are all potentially amenable to investigation using the same genetic marker and method of analysis.

In Chapter 5 the evolution of trophic shifts in a parasitoid family is examined. The family includes both parasitic entomophagous species, and phytophagous species in the form of gall inducers and seed feeders. We currently understand little about switches between these alternate feeding strategies. Of particular interest is the ease with which switches occur from one trophic niche to another. If such switches are rare it implies that they are difficult to evolve, and there are barriers to such changes that we do not yet understand. The existence of entomophagous species that also feed on gall tissue implies one possible direction by which switches in trophic niche may evolve, however, the actual transition of change remains to be demonstrated.

Chapter 6 provides an investigation into the possible utility of DNA barcoding in oak gall wasp communities. Given the various difficulties associated with rearing oak galls and identifying their inhabitants, a standardised method with which to assay gall inhabitants would be immensely useful.

Chapter 2

Methods

2.1 Phylogenetic techniques

Much of the analysis used in this thesis relies on the use of a molecular phylogenetic framework. Chapters 3 and 4 are based on phylogeographic analysis, which requires the establishment of phylogenetic components of population structure, and Chapter 5 is an investigation into the relationship between European oak gall attacking torymids. Therefore, a review of current molecular phylogenetic methods is provided below, with justification for the methods applied in this thesis. In addition to building phylogenies, my analyses also involve mapping of characters across phylogenies. This latter aspect is discussed in detail in Chapter 4.

2.11 What is molecular phylogenetics?

Phylogenetics is the establishment of inferred relationships of descent and common ancestry between taxa. The true evolutionary history of a group of taxa is usually unknown. Phylogenetics is concerned with determining a best estimate. The term originates from the Greek words *phylon* meaning ‘tribe’, and *genesis* meaning ‘origin’. After acceptance of the theory of evolution, it was clear that the establishment of evolutionary relationships between organisms was of great importance, both to the development of our understanding of evolution and as a tool for addressing other questions of interest within biology.

Until recently, inferences of evolutionary relationship were limited to the analysis of morphological characters. However, the last fifty years has seen the discovery of Deoxyribose Nucleic Acid (DNA) as the hereditary material and the adoption of molecular phylogenetic techniques. The field of phylogenetics is undergoing rapid development, driven not only by the availability of molecular data, but also by the development of powerful statistical methods coupled with the availability of high-powered computing with which to analyse the data.

The phylogenetic method is based upon the use of hypotheses of homology as a means to establish relationships between organisms. A homology in the evolutionary framework is defined as a shared derived character (synapomorphy). That is, a character inherited from a common ancestor. The idea is that evolutionary history can be reconstructed by tracking changes in these characters. In practice, the picture is usually clouded by homoplasy: characters appearing alike due to convergent evolution and not because of shared ancestry. However, the method works because in theory there should be more characters alike due to shared ancestry, than because of independent evolution in related organisms.

The interpretation of morphological homology can be subjective and can lead to artificial groupings. Molecular data is sometimes considered superior since character states can be objectively defined (*eg* A, T, G, C), and are thus more easily comparable across studies by different investigators. Molecular data is also relatively easy to obtain, and large data sets can be assembled relatively quickly. For these reasons molecular phylogenetics has largely replaced studies using morphological data, and is the method utilised in this thesis. Although it has a number of benefits over morphological analyses, subjectivity can be introduced into molecular phylogenetics and assumptions regarding sequence evolution must be carefully considered. Furthermore, morphological phylogenetics can still play an important role, especially when used in conjunction with molecular data, or for use with fossil organisms where molecular methods are not possible.

Hypotheses of homology in molecular phylogenetic analyses are determined in a sequence alignment, and relationships amongst taxa are inferred on the basis of these homologies. The first important assumption in such studies is that the genes under consideration are homologous in different species. A second is that the respective sites are homologous to each other. Differences in the sequences indicate genetic divergence as a result of molecular evolution over time. However, separating homology from analogy in DNA sequences can be difficult, since there are only four possible character states at any given position along the sequence. A third assumption is that there is independence among characters. That is, a change in one character will not affect the likelihood of a change in a neighbouring character.

A further assumption of phylogenetics in general, is a bifurcating tree-like model of evolution, in which each ancestor gives rise to two daughter lineages. Under this model, for any given number of taxa there are $[(2n-3)!]/[2^{n-2}(n-2)!]$ possible unrooted networks (Cavalli-Sforza and Edwards 1967), each one of which represents a hypothesis of evolutionary relationship. Thus, when $n = 10$, there are more than thirty-four thousand possible hypotheses to choose between, only one of which represents the true relationship between the taxa in question. A number of methods have been developed, with which to select from competing hypotheses of relationship. The particular method chosen has a substantial effect on the result of phylogeny estimation. A brief review of the most popular methods, and a justification of those used in this thesis follow. This section is synthesised from an understanding gained from multiple sources, both primary literature and textbooks. The following sources were particularly important: Felsenstein (1981, 2004), Page & Holmes (1998), Hillis *et al* (1996), Huelsenbeck & Ronquist (2001), Huelsenbeck *et al* (2002).

A key assumption of phylogenetics is that sampled sequences do not include internal nodes but only represent tips on the tree. It is however possible for real populations to include both ancestral and descendent sequences. Such relationships can be incorporated in network-based approaches such as Templeton's nested clade analysis (Templeton 1983, 1998). The statistical basis of hypothesis testing in nested clade analysis and other related approaches remains an issue of debate, and lacks the objective rigor of Bayesian phylogenetic approaches (see below). Thus, having noted the existence of alternative methods, I apply only phylogenetic approaches in this thesis.

2.12 Methods of phylogeny estimation

Maximum parsimony

The term parsimony originates from the Latin *parcere* 'to spare', and is used to express the principle of 'Ockham's Razor'. The principle, which is attributed to the 14th century Franciscan friar and logician William of Ockham, states: *Pluralitas non est ponenda sine necessitate* 'plurality should not by posited without necessity', or put more simply: 'the simplest explanation is usually the best'. Ockham's Razor cuts away unnecessary complexity. In the context of phylogeny, the most parsimonious tree is that which requires the least number of evolutionary gains or losses to explain the data. The use of maximum parsimony as a phylogenetic method was adopted during the 1960's (Camin & Sokal 1965, Eck & Dayhoff

1966, Edwards & Cavalli-Sforza 1967, Fitch & Margoliash 1967), and was one of the earliest attempts at making phylogeny estimation more objective.

Concerns about the use of maximum parsimony have been raised, since the criterion assumes that evolutionary change is improbable *a priori* (Felsenstein 1973, 1979). It is also evident that the evolutionary process does not always proceed in the most parsimonious manner. However, despite these discrepancies parsimony has proven a useful means of providing a good working estimate of phylogeny, and its proponents claim that its use keeps error lower than with methods based on other assumptions.

More significant problems have become apparent recently, with the replacement of morphological data by sequence data as the standard means of inferring phylogeny. Unlike morphological adaptations, DNA sequences possess just four possible character states, making the potential for homoplasy high. Thus, as the number of changes per sequence increases, the potential for homoplastic groupings becomes an increasing problem. Indeed, Felsenstein (1978) has shown that the use of maximum parsimony in analyses that contain taxa with long-branches can result in misleading results. Furthermore, parsimony is concerned solely with minimising the number of evolutionary steps necessary to explain a data set, and takes no account of the process of sequence evolution. For example, it ignores the possibility of a change at a given site followed by reversion to the original state (a 'multiple hit'), or the higher likelihood of a transition than a transversion.

Over recent years, there has been much progress in understanding the way in which substitutions occur, and numerous models have been formulated with which to approximate sequence evolution. Proponents of parsimony argue that we should not use models to approximate the process of molecular evolution, since our knowledge is incomplete and their use clouds the objectivity of an analysis; although parsimony itself is an assumption, it is the criterion that requires us to make the least assumptions about the evolutionary process. However, although maximum parsimony lacks a set of explicit evolutionary assumptions, it clearly contains implicit assumptions. Maximum parsimony is not truly 'model-free', and furthermore its use can be positively misleading (Felsenstein 1978). Maximum parsimony can be a fast method, and in some cases can provide a reasonable indicator of relationship when dealing with molecular data. However, the use of models allows more informed choices between hypotheses of relationship to be made, and provides scientists with a means of developing their understanding of the underlying mechanisms of sequence evolution. For these reasons, model-based analyses have become more popular, and are used as the basis for analysis within this thesis.

Distance-based methods

Distance methods differ from other methods of phylogeny estimation in that they do not use the sequence alignment directly to construct the tree. Instead, the sequence alignment is converted into a matrix of pairwise-distances between sequences. Thus, since a summary of the genetic data is used instead of the full range of information present in the alignment, some information is lost. The assumptions of distance-based methods are that the less two sequences differ, the more closely they are related, and that as time passes, the more divergent two sequences that evolved from a common ancestor will become. Saturation can cause a problem for distance methods, since more recent mutations can mask earlier ones if multiple hits occur, thus breaking the second assumption. However, models of sequence evolution may be applied in order to help correct for this. The most popular distance method is Neighbour-joining (NJ)(Saitou & Nei 1987), which calculates the divergence of each taxon to every other taxon in the data set to produce a distance matrix, from which the closest pair of taxa is identified. A new matrix is then calculated with these taxa represented as a single node, and the process is repeated until all taxa have been joined. Distance methods usually provide a reasonable estimate of relationship and are very fast. Because of this they are frequently used as the starting point for other analyses.

Maximum likelihood

Unlike maximum parsimony and distance methods, maximum likelihood is a method based on statistical inference. It was developed by the geneticist and statistician Sir Ronald A. Fischer, but its first application to phylogeny estimation was suggested by Edwards & Cavalli-Sforza (1964), and it was only fully developed for nucleotide sequence data with Felsenstein's implementation (1981).

The likelihood is the probability of the data given the hypothesis: P(D|H). In other words, it is the probability that a certain hypothesis confers upon the data. When considering likelihood, it is important to understand that a hypothesis can confer high likelihood on the data, but may be highly improbable itself. That is, the probability of getting the observed data given a certain hypothesis may be very high, but the actual probability of the hypothesis can be low.

In the case of phylogenetics, the data is the sequence alignment, and the hypothesis is the tree and model of sequence evolution. Thus, the maximum likelihood method seeks to find the tree whose topology maximises the probability of observing the information in the sequence alignment, under a certain model of sequence evolution. The data are fixed, but we can modify the hypothesis to observe the effect this has upon the probability of the data, and thus seek out the hypothesis that maximises the probability of achieving the observed sequence alignment. In practice, the likelihood is taken as a function of the hypothesis (the tree) as opposed to a function of the data, and is usually expressed in terms of log likelihood because the numbers involved in likelihood calculation are so small that computers have problems handling them. Maximum likelihood estimation generally proceeds as follows:

1. Construction of a particular unrooted network for n taxa
2. Arbitrary rooting
3. Calculation of the likelihood for each site for that tree
4. Calculation of the sum of site likelihoods to find the likelihood of the entire tree
5. Search for trees with a higher likelihood

A new tree is either constructed by joining the taxa in the alignment according to some specific algorithm, or an initial tree is constructed using an alternative faster method such as neighbour-joining or sometimes parsimony. This provides an initial estimate for the maximum likelihood process to improve upon. The tree is rooted, but the specific location of the root does not change the likelihood of the tree, as described by Felsenstein's Pulley principle (Felsenstein 1981).

Due to the assumption of the independence of sites, it is possible to calculate the probability of a given data set on a particular tree site by site, with the product calculated at the end. In this way, the joint probability of observing all site patterns is the product of the individual site patterns. For a given topology the likelihood is thus:

$$L_N = \sum_{i=1}^{N} L_i$$

Where,
N = the number of sites in an alignment
i = each site i

However, computation of the likelihood of a particular topology for a given site is not straightforward. The calculation consists of the sum of the probabilities of every possible reconstruction of ancestral states along each branch of the particular tree (given a certain model of base substitution) that would give rise to the observed pattern in the alignment for that site. For n taxa, there are n-1 internal nodes (branching points in the tree), each of which can have one of four states for DNA sequence data. Thus, it is necessary to calculate 4^{n-1} terms per site for the sum over all possible assignments of bases to nodes on a tree (Felsenstein 1981). This number of calculations quickly becomes intractable. For instance, for an alignment consisting of 30 taxa and 500 nucleotide sites, calculation of 8.6×10^{18} terms would be necessary for a single given topology and set of branch lengths, and adjustment of a single branch length would require recalculation of them all. Fortunately, the probability summation can be rearranged into another form; that of Felsenstein's "pruning" algorithm (Felsenstein 1973, 1981), which makes use of the conditional likelihood of a subtree (a subtree being a subsection of another tree). Using this algorithm, the likelihood of a tree for a particular site under consideration is given by:

$$L_i = \sum_{x} \pi_x L_0^{(i)}(s)$$

Where,
i = each site i
x = a node on a tree, for site i
s = the state (base) at a node
π = the prior probabilities of finding each of the four bases at point x on the tree

Since an evolutionary steady state in base composition is assumed, the prior probabilities of finding each of the four bases at point x on the tree reflects the overall base composition in the group under study. The conditional likelihood of a subtree can be represented by the following term: $L_x^{(i)}(s)$. It is the probability of everything that is observed from node x on the tree upwards, given that node x has condition s. There are always four of these quantities, corresponding to the four different possible values of s. The basis of the pruning algorithm is that once these values have been computed, it is not necessary to compute them again. Under the algorithm, a set of four $Lx(i)(s)$ are calculated for each tip, and progress is made down the tree, until the bottom node (0) is reached. In this way, the four conditional likelihoods given the possible states of each site at point x are computed, thus substantially cutting down on the number of calculations necessary.

Ideally, the likelihood would be maximised over all possible evolutionary trees, and the tree with the highest likelihood selected as the best representation of the relationship between the taxa. However, exhaustive searches are impractical in most cases due to limited processing power, and some form of heuristic searching is employed instead. To understand the rationale of tree searches it is easiest to visualise the surface of the distribution of possible trees as a landscape in which trees occupy a certain height in proportion to their likelihood. Trees of high likelihood form 'hills', whilst those of low likelihood form 'valleys'. The tree occupying the apex of the highest hill represents the tree that maximises the likelihood of the data and trees lying closer together on the surface are more similar. A heuristic tree search moves around the landscape of trees by starting with an initial tree, usually an existing estimate of phylogeny generated by a relatively fast method such as neighbour-joining or parsimony, and then adjusting it and recalculating the likelihood. Changes are usually only accepted if they result in an increase in log likelihood. 'Nice' landscapes are hill-shaped and every search quickly converges upwards to the global optimum. 'Bad' landscapes look like a wide flat plain that contain one or more narrow spikes of different heights. In this case a heuristic search must move across the plain in order to find the single highest spike.

Maximum likelihood is known to be consistent under a large range of circumstances. In addition, calculation of the likelihood provides a measure of fit between the data and model with which alternatives can be quantified, and a range of evolutionary hypotheses investigated. However, the pruning algorithm is computationally demanding, and analyses can be prohibitively slow even on powerful machines.

Bayesian Inference

Bayesian inference is one of the oldest forms of statistical inference, dating to the 18th century. It is based on Bayes' theorem (Bayes 1763), named after the British mathematician and Presbyterian minister, the Reverend Thomas Bayes. The theorem is a result in probability theory, which gives the conditional probability distribution of a random variable A given B (the posterior), in terms of the conditional probability distribution of variable B given A (the likelihood) and the marginal probability distribution of A alone (the prior). In essence, the theorem states that a hypothesis is confirmed by a body of data if the existence of that data implies that truth in the hypothesis is probable.

Bayes theorem:

$$P(H_0|D) = \frac{P(D|H_0)\,P(H_0)}{P(D)}$$

Where,
H_0 = a hypothesis, either new or induced from a preceding set of observations
D = the data

Thus,
$P(D|H_0)$ = the conditional probability of observing D, given that the hypothesis H_0 is true (As a function of H_0 given D, it is called the likelihood function).
$P(H_0)$ = the prior probability of H_0
$P(D)$ = the marginal probability of D. It is a normalising constant and can be calculated as the sum of all mutually exclusive hypotheses: $\sum P(D|H_i)P(H_i)$
$P(H_0|D)$ = the posterior probability of H_0 given D

Bayesian inference was proposed as a method of phylogeny estimation relatively recently (Li 1996, Mau 1996, Rannala & Yang 1996). The posterior probability of the *i*th tree is defined as:

$$f(\tau_j|X) = \frac{f(X|\tau_i)\,f(\tau_i)}{\sum_{J=1}^{B(s)} f(X|\tau_j)\,f(\tau_j)}$$

Where,
X = an alignment of DNA sequences
τ_i = the posterior probability of the *i*th phylogenetic tree

(After Huelsenbeck & Ronquist 2001)

Bayesian inference is similar to maximum likelihood in that it is a method of statistical inference, and makes use of the likelihood function. However, it differs in several key ways. Firstly, it sets out the problem of finding the optimum tree in a more logical framework, by seeking to find the tree that maximises the probability of the data, as opposed to *vice versa.* A more significant difference is the use of the prior. The use of priors in statistical inference is a contentious issue, and application of a prior probability distribution on trees can be viewed as either a strength or weakness of the method. A prior is based on belief, and as such its validity cannot be tested statistically. Statisticians are divided between those who support the use of degrees of belief as a statistical practice (Bayesians), and those who believe that statistical procedures only make sense when one uses the relative frequency concept (frequentists). With respect to phylogenetics, maximum likelihood represents the frequentist viewpoint, and Bayesian inference that of the Bayesian. The choice between frequentist and Bayesian approaches is a philosophical issue; is it right to specify prior beliefs, or is it better to ignore background knowledge? On the one hand, the use of priors allows an investigator to incorporate their prior knowledge about the phylogeny of a group in an analysis. However, in the absence of strong beliefs care must be taken, as under certain circumstances it is possible for the prior to have more influence over the posterior than the data (Huelsenbeck *et al* 2002). The usual solution to the allocation of priors is to assume that all possible trees are equally likely, unless there is strong evidence to the contrary, and assign 'flat-priors', which place an equal probability on all possible values within a given range likely to contain the true value, to all model parameters. A further key difference is the use of Markov Chain Monte Carlo (MCMC) (Metropolis *et al* 1953, Hastings 1970) or 'random walk' methods, which are used to sample from an approximation of the posterior probability distribution.

As with maximum likelihood, the tree on its own does not provide enough information to allow calculation of the likelihood. A substitution model is specified to describe how the characters evolve on the tree, as well as information on the value of parameters and branch lengths on the tree. Calculation of posterior probabilities require a summation over all $B(s)$ trees that are possible for s taxa, in addition to the integration over all combinations of branch lengths and substitution-model parameters necessary to determine the likelihood for each tree. These summations and integrals cannot be evaluated analytically (Huelsenbeck & Ronquist 2001, Huelsenbeck *et al* 2002), presenting a problem for the implementation of Bayesian inference. A solution to this problem is the use of MCMC methods. These are currently the only numerical methods that can approximate the posterior probability of trees. MCMC algorithms take steps (known as generations) of varying length, taking a different direction after each step. They always go uphill if the option is available, and only go downhill if the height difference is small. In this way, they visit points in proportion to their height, (posterior probability density). If run for long enough, a MCMC chain is guaranteed to visit tree topologies and parameter values in proportion to their posterior probabilities (or probability densities). Thus, to approximate the volume under a surface in the likelihood landscape, all that is necessary is to count the number of steps that fall within the area and divide by the total number of steps taken. The first few steps when the algorithm is not walking very close to the global likelihood maximum (which equates to the centre of the probability distribution) are known as the 'burnin' and are discarded, and the longer the algorithm is allowed to walk the better the approximation is. The usual procedure is to begin with a random tree, evaluate its posterior density according to the model specified, make a change, and evaluate the new tree. If the change results in an uphill move on the likelihood surface it is automatically accepted. If not, it is accepted or rejected depending on the ratio of the two posterior densities, with the density corresponding to the proposed change on top. If only going downhill slightly this ratio will be close to one and the change will often be accepted. Every so many generations the current tree and its likelihood are recorded, and then summarised (minus the burnin) at the end of the analysis. These trees are usually summarised in a consensus tree (either strict or majority rule). The MCMC approach allows all model parameters to be effectively integrated out to obtain the posterior probability of just one parameter, the tree topology.

Standard MCMC can be prone to entrapment in local optima, with resultant poor 'mixing' of the chain. Mixing refers to the ability of the chain to explore the posterior distribution. Poor mixing means that a parameter did not vary much from its starting value, so the MCMC did not do a good job of marginalizing over that particular parameter. Poor mixing of the chain can result from not taking steps of the appropriate size, or from the presence of isolated ‘peaks’ in the posterior density surface. In the case of isolated peaks it is possible for the chain to become stuck on one of these and never visit other peaks. Whenever the chain takes a step to get off the peak, it lands in the valley between peaks, and such proposals are always rejected because the posterior is much lower in the valleys. Metropolis Coupled MCMC ($(MC)^3$) is a variant of MCMC that allows multiple peaks in the landscape of trees to be more readily explored, reducing the likelihood of poor mixing. It is analogous to certain heuristic searching algorithms employed in maximum likelihood tree searches. It uses several 'heated' chains in addition to the 'cold' chain. The posterior density of the current location of the cold chain, which is the only chain used for the summary statistics, is always raised to the power 1. The posterior of the heated chain is raised to a power between 0 and 1. If a chains power were 0 the chain would be exploring a completely flat terrain (since any value raised to the power 0 equals 1). A heated chain therefore sees the landscape of trees as flattened relative to the ‘cold’ chain, and can cross valleys more readily. After all the chains each take a step, two chains are chosen at random to attempt a swap. A swap of two chains is successful if the difference in their heights is not too great. The heated chains thus act as scouts to identify peaks for the cold chain.

The downside to using heated chains is that each additional heated chain effectively doubles the time required to complete a given number of steps. The use of heated chains can increase the execution time of a run considerably, since each iteration within a chain involves the calculation of a potentially expensive likelihood function. Thus, it is necessary to balance the need for heated chains to avoid poor mixing, with the need to run the cold chain as long as possible. However, since analyses based on the MCMC algorithm are computationally more efficient than the standard methods of establishing support using maximum likelihood, the total time taken is usually considerably less than with these methods. This also makes it possible to analyse more complex and realistic evolutionary models using Bayesian inference. One problem with the MCMC process is that it can be difficult to decide whether the chain has been run for long enough to converge upon its stable distribution. However, multiple independent runs will often show conspicuous differences if the chains have not been run long enough, and comparisons between the standard deviation of clade support values between different runs give a good indication of convergence. Due to the power of the MCMC approach, the logical support statistics discussed below, and the significant time savings that can be gained by using Bayesian inference, this is the method of choice employed in this thesis.

2.13 Which model of sequence evolution?

As stated above, maximum likelihood and Bayesian approaches to phylogeny estimation are based on statistical inference. In order to draw inferences such methods require that a probabilistic model of sequence evolution be specified. It could be argued that modelling sequence evolution for the purpose of phylogeny estimation is circular, since only with good estimates of phylogeny will it be possible to fully understand the process of sequence evolution itself. However, it is difficult to envisage an alternative solution to the problem, and such an approach appears to be the only means by which understanding can currently advance. Hennig (1966) referred to the situation as "reciprocal illumination", since a good estimate of phylogeny is relevant to the development of good models, and improving models will lead to better trees.

A large number of models of sequence evolution have been proposed. They all provide a means of modelling the evolutionary process of nucleotide substitution, but differ in their complexity and in the assumptions implied by their usage. Generally, increasing the number of parameters in a model will improve its fit to a particular data set. However, the addition of extra parameters increases the variance around the estimate, since each individual parameter must be estimated and is subject to error itself. Therefore, a model complex enough to explain the data, but no more so is generally advocated. In addition, complex models are more computationally difficult and so their usage increases the computing time required for an analysis.

Models of sequence evolution generally consist of two parts, one that estimates composition and one that estimates process. The composition part of the model estimates the proportion of the four nucleotides present in the sequences. For instance, all nucleotides present in equal proportions, or perhaps double the amount of adenines and thymines, than cytosines and guanines. The process part of the model estimates the probability with which the nucleotides change from one to another. For example all substitutions equally likely, or alternatively transitions more probable than transversions. A brief overview of the most commonly used models of sequence evolution follows.

Models of DNA sequence evolution

The Jukes-Cantor model (1969) is the simplest model of DNA evolution. The model assumes that all nucleotides are present in equal frequencies and that all possible nucleotide substitutions occur at equal rates.

Kimura's two-parameter model (1980) is similar to the Jukes-Cantor model, but allows the rate of transitions and transversions to vary.

Models such as the Felsenstein 1984 (1989), Hasegawa-Kishino-Yano (1985) and Tamura-Nei (1993) model relax the assumption that the bases have equal expected frequencies and allow the transition-transversion ratio to vary.

The general time-reversible model (Lanave *et al* 1984, Rodriguez *et al* 1990) represents the most complex model type commonly used. It relaxes the assumption that bases have equal expected frequencies and allows all possible substitutions to occur at their own rate.

The existence of a range of models is necessary, since different genomic regions evolve under different selective pressures and constraints. Therefore, a model that provides a good degree of fit to one data set might not be appropriate for another. Phylogenetic model selection is a rapidly developing field, and choice of which method of model selection is best is a contentious issue. Model selection is of importance, since the model influences the results of an analysis, and so use of the wrong model can result in an incorrect inference of phylogeny.

An overview of the most commonly used methods of model selection in phylogenetics is given below.

The likelihood ratio test

One of the most common means of selecting between models is by use of the likelihood ratio test (LRT). The LRT examines the goodness of fit of the observed data to the predictions of one model in comparison to the goodness of fit under an alternative model. This is usually achieved using the improvement in log-likelihood that the alternative model affords. The likelihood ratio (LR) is defined as:

$$\delta = 2 \cdot \log \Lambda$$

where,

$$\Lambda = \frac{\max\,[L_0\ (\mathit{null\ model} \mid \text{data})]}{\max\,[L_1\ (\mathit{alternative\ model} \mid \text{data})]}$$

or expressed in terms of log likelihood and as usually tested,

$$\mathrm{LR} = -2(Log_1 - Log_2)$$

When the addition of extra parameters significantly improves the likelihood, the more complex model is favoured. However, due to its statistical framework, the LRT is only applicable when the models being examined are nested. That is, when the null hypothesis is a special case of the alternative hypothesis.

The LRT is most commonly implemented as a hierarchical LRT (hLRT). Firstly, a tree is estimated from the data. It has been shown that the tree has little influence on the final model selected (Posada & Crandall 2001) so a fast method of tree construction such as NJ is usually used. The likelihoods of the candidate models are then calculated for the given data set and tree and compared in a hierarchy of LRTs, with the best-fit model among the candidates selected. In practice, the implementation of hLRTs is usually accomplished using the computer program Modeltest (Posada & Crandall 1998).

Information criteria

Information criteria compare all competing models at a time, instead of just comparing two models as is done in the LRT. Each model is penalized by a function of the number of parameters in the model. Two commonly used examples are the Akaike Information Criterion (AIC) (Akaike 1974), and the Bayesian Information Criterion (BIC) (Schwarz 1974).

The AIC measures the expected distance between the true model and the estimated model, taking into account the goodness of fit and the variance of the parameter estimates. It is defined as:

$AIC_i = -2 \log_e L_i + 2 N_i$
N_i = the number of free parameters in the ith model
L_i = the maximum likelihood value of the data under the ith model

The smaller the AIC the better the fit of the model to the data. Unlike the LRT it can be used to compare non-nested models. The AIC is also most commonly implemented in Modeltest.

The BIC measures the relative support that the data gives to different models. It is defined as:

$BIC_i = -2\log_e L_i + N_i \log_e n$
n = sample size (sequence length)

As with the AIC, the smaller the BIC, the better the fit of the model to the data. It can also be used to compare non-nested models.

Bayes factors

Bayes factors are an alternative to classical hypothesis testing, and represent the Bayesian analogue of LRTs. The Bayes factor is simply the ratio of the marginal likelihoods of the two models being compared. Since the marginal likelihood of a model is difficult to calculate, it is usually estimated using the harmonic mean of the likelihood values of an MCMC sample (Newton & Raftery 1994). The evidence provided by the data for two competing models i and j is contrasted as:

$$B_{ij} = \frac{P(D|M_i)}{P(D|M_j)}$$

In practice the difference between the two logarithms of the harmonic means are usually compared using the recommendations developed by Kass and Raftery (Kass & Raftery 1995, Raftery 1996), where evidence for M_i is considered very strong if $B_{ij} > 150$, strong if $12 < B_{ij} < 150$, positive if $3 < B_{ij} < 12$, barely worth mentioning if $1 < B_{ij} < 3$, and negative if $B_{ij} < 1$.

Bayes factors provide a convenient way of comparing models via the Bayesian approach. They also do not require the models compared to be nested. However, the use of Bayes factors for model selection in phylogenetics is a new technique (Nylander *et al* 2004), and requires refinement. Bayes factors usually give very high support for the more parameter rich models used in phylogenetic analyses (Nylander personal com), and it may be problematic to rely solely on their use for model choice (see also Irestedt *et al* 2004).

In this thesis the hLRT was the method chosen for model selection. It is perhaps the most commonly used method in the literature, and has a well-grounded theoretical basis. However, in a recent review of methods of model selection Posada & Buckley (2004) argue that the LRT is not the most optimal method, suggesting instead the use of AIC and Bayesian methods. It is clear that the field of model selection is progressing rapidly and further changes in methodology are likely in the future.

Site rate heterogeneity

The models of sequence evolution described above assume that all sites evolve at the same rate. However, certain sites within an alignment frequently display a faster rate of substitution than others. The reason for this difference appears to be different selective constraints acting on different sites due to the functional and or structural requirements of the gene in question (Yang 1996). Models that incorporate variation in rates across sites can result in a much better fit with the data (Yang 1993, Sullivan & Swofford 1997). Furthermore, it has been found that violation of the assumption of equal rates can have severe consequences for the results of a phylogenetic analysis (Yang 1994, 1996, Gaut & Lewis 1995, Lockhart *et al* 1996). Therefore, rate heterogeneity is often incorporated into substitution models.

The standard approach to characterizing among-site rate variation is to use the gamma distribution to approximate rates at sites (Yang 1993). A certain site is assumed to have an unknown rate determined by its position in the molecule, and the rate of each site is assumed constant across evolutionary lineages. Each site has a certain probability of evolving at any rate contained within a gamma distribution with shape parameter α. When α equals infinity all rates are equal, however, most empirical estimates of the shape parameter fall in the range of 0.1-0.5 indicating substantial rate variation among sites (Huelsenbeck & Crandall 1997). The distribution is usually broken down into several discrete rate categories instead of being implemented in continuous form, for computational ease. Yang (1994) found that four rate categories are usually a good compromise, and this is the default number used in most applications. The likelihood for a given site is obtained by summing over rate categories the likelihoods of the site at each given rate, weighted by the probability that the site is drawn from each category. A rate is assigned to each site and this estimate is then improved over time.

The invariable sites model (Palumbi 1989) is also often used. This model assumes that some sites are incapable of accepting substitutions, perhaps due to strong functional constraint, and the remaining sites all vary at the same rate. The invariable sites and gamma model are often used in combination to describe the rate distribution observed at a certain locus.

If some *a priori* knowledge exists relevant to the assignment of rates within a locus, this can also be incorporated within a model. For instance, given what we know about the genetic code, for protein coding genes sites may be assigned to categories based on codon position and the rate estimated for each category separately. Alternatively, when dealing with ribosomal RNA genes, sites may be assigned to different rate categories depending on whether they are in stem or loop regions. Although this approach assumes that all the sites within a category have the same rate, it can provide a better fit and may be more likely to lead to more accurate estimation of phylogenetic relationships (Goldman & Yang 1994).

2.14 Support statistics

A crucial question when dealing with estimates of phylogeny is the degree of confidence we can have that the results of our analysis are in some sense real, and not just the product of chance.

The most popular means of determining support for estimates of phylogeny derived from maximum likelihood is non-parametric bootstrapping (Felsenstein 1985), which is a test of the internal consistency of the data. The term bootstrapping is borrowed from parametric statistics and means ‘sampling with replacement’. It originates from the story of Baron von Munchausen, who pulled himself out of a swamp with the aid of his own bootlaces. Sites within an alignment are randomly sampled with replacement to construct a pseudoreplicate data set of the same size as the original, and the best tree for this data is found. This process is repeated *n* number of times (usually 100 or 1000) and the degree of conflict between the resultant trees is assessed by use of a consensus tree. Clades are only considered well supported if they are present in 70% or more of the pseudoreplicates (Hillis & Bull 1993). The main problem with bootstrapping is that it requires that all the characters in the original data matrix should represent a random sample of all possible characters. Unless the characters in the data set accurately reflect the larger distribution of all possible characters, bootstrap support will not provide good estimates. However, the characters in an alignment are linked as a series from a certain locus. Furthermore, real data sets are never sufficiently large to meet the statistical requirements of the bootstrap, and as a result only approximate confidence levels can be produced. As such, bootstrap values may be considered approximate guides to the support of the data to certain clades, but at worst their usage could be considered to lack any real justification.

The most frequently used method of summarising support for Bayesian analyses is by use of the posterior probability found for each clade. As mentioned above, during MCMC trees are sampled at fixed intervals and the posterior probability of a tree is approximated as the proportion of time that the chains visited it. A consensus tree can be obtained from these sampled trees and the Bayesian posterior probabilities of individual clades can be viewed as clade credibility values. The posterior probabilities of individual clades are known as split probabilities. Split posterior probabilities are analogous to bootstrap proportions, however, they are a much more natural measure of nodal support. Bootstrap proportions are not interpretable as the probability that a split is in the true tree, but this is the exact meaning of a posterior probability. Bootstrap values and posterior probabilities are commonly provided together in phylogenetic studies. However, Douady *et al* (2003) warn against a direct comparison of the two. Bootstrapping examines the range of variation expected if a different set of nucleotides were sampled from the same distribution, whereas split posterior probabilities examine the probability that a clade is present in the true tree, given the data observed. As such they provide answers to different, but related, questions.

Bayesian support values are higher than those usually obtained for bootstrapping (Wilcox *et al* 2002, Douady *et al* 2003). Non-parametric bootstrap proportions represent highly conservative estimations of phylogenetic accuracy (Hillis & Bull 1993). Wilcox *et al* (2000) found that the higher levels of support derived from posterior probabilities are appropriate and indeed much closer to the estimates of phylogenetic accuracy, even though they found them to be still somewhat conservative. Furthermore, determination of Bayesian support is usually much faster than determining support using bootstrapping. Douady *et al* (2003) found that a standard Bayesian search was sixty times faster than a single bootstrap replication using a 1.8Ghz Pentium 4 computer with simultaneous estimation of all parameters. Since Bayesian inference is the method of choice in this thesis, support values given alongside clades are as Bayesian posterior probabilities unless stated otherwise.

2.2 Choice of molecular marker

A mixture of mitochondrial DNA (mtDNA) and nuclear markers represents an ideal means by which to investigate the questions outlined in this thesis. Allele frequency data for multiple nuclear markers permit high-resolution analyses of demographic impacts on genotype variation. This approach often allows a far more detailed examination of population substructure than is possible with mtDNA alone. Furthermore, nuclear markers help to corroborate the relationships inferred by mtDNA data and rule out the action of various processes that can obscure the true signal (see below). For these reasons allozyme and microsatellite data were sought for use in this thesis. However, the attempts made to generate sufficient informative data using these markers were unsuccessful (see Appendix 1). In addition, the Internal Transcribe Spacer (ITS) regions of the ribosomal rDNA array and the long-wavelength opsin gene were also targeted unsuccessfully (Appendix 1). Ultimately resources of time and money limited the further search for nuclear markers. Therefore, interpretation of patterns derived from mtDNA analyses is structured in such a way to avoid artifacts that can be associated with its usage (see below).

For the reasons outlined above, this thesis relies heavily on the use of mtDNA. Chapters 3 and 4 are concerned with the establishment of between population relationships in order to make inferences about the phylogeographic history of species. A further reason for the use of mtDNA in these chapters is so that work on parasitoids will be directly comparable to what has been done already for the gall wasp hosts. In chapter 5, mtDNA is used as a marker with which to infer species relationships, and chapter 6 examines the utility of mtDNA as a tool for the barcoding of oak gall wasp communities. This section discusses the rational behind the use of mtDNA as a genetic marker, and examines possible problems associated with it.

There is a long history of the use of mtDNA in studies of population genetics, phylogeography and phylogenetics (Brown & Vinograd 1974, Avise *et al* 1979, Avise 1987, Harrison 1989, Simon *et al* 1994, Loxdale & Lushai 1998, Caterino *et al* 2000). This popularity is as a result of several characteristics that make mtDNA a useful marker, and these are now discussed in turn. Transmission of the mitochondria occurs solely through the cytoplasm of the maternal egg. Therefore, mtDNA is maternally inherited and undergoes no, or very little recombination (for a discussion of mtDNA recombination see Wallis 1999, Rokas *et al* 2003b, Kraytsberg *et al* 2004). Lack of recombination means that mtDNA haplotypes have a linear evolutionary history, and can be traced back through time easily. The effective population size of mitochondria is approximately one quarter that of a nuclear autosomal gene, owing to the fact that the mitochondrial genome is haploid and maternally inherited. Furthermore, in comparison to most nuclear genes, mtDNA has a high mutation rate. Thus, mtDNA is useful for investigation of the pattern of relatively recent events. There are also several practical advantages associated with the use of mtDNA that can save considerable amounts of both time and money. Since the mitochondrial genome is haploid, there is no need to clone individual alleles in order to avoid the presence of 'multiple peaks' in sequencing results. The mitochondrial genome is present in high copy numbers and as a result, yields of mtDNA are typically high during DNA extraction and the use of old or damaged specimens is more likely to produce useful DNA. There is also an abundance of well conserved primers available, facilitating the likelihood of trouble free amplification during PCR. However, despite these benefits there are issues involved in the use of mtDNA that must be taken into consideration.

Firstly, when using gene trees to infer species trees, the terminal taxa are in effect different alleles of the gene in question. When using molecular phylogenetics, the hope is that the gene tree will provide a good estimate of the phylogenetic history of the species under investigation. However, this is will not necessarily always be the case. Certain processes can result in gene trees being discordant with species trees. This issue is not limited to the use of mtDNA. However, it does depend to some extent upon choice of the marker used. The factors that can result in mtDNA trees being discordant with species trees are reviewed below.

Lineage sorting

This term refers to the process whereby the relationship between gene and species or population trees becomes distorted as a result of the stochastic sorting of genealogical lineages within a species or population. If the locus under consideration was polymorphic in the ancestor to a group of taxa, it is possible that incongruence between a gene tree and a tree for that group of taxa might arise as a result of sorting of this polymorphism and lineage extinction (Neigel & Avise 1986, Nei 1987, Wu 1991, Doyle 1992). This process is illustrated below in Figure 2.1.

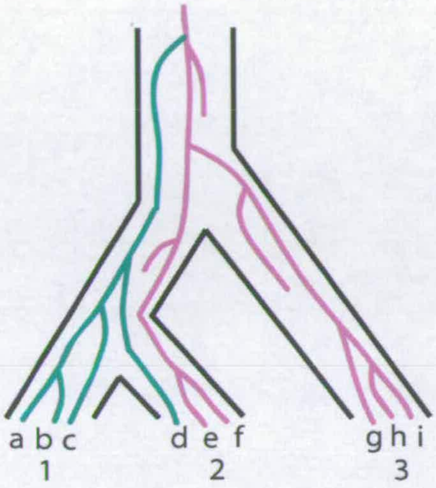


Figure 2.1 Three taxa (1,2,3) containing several different lineages (a-i) derived from ancestral polymorphism. As a result of a short internode and long coalescence times, taxon 2 could be inferred to be more closely related to taxon 1 or taxon 3, depending on which lineages were sampled.

Given long coalescence times and a relatively short internode between bifurcations, the probability of incongruence can be quite high (Nei 1987, Pamilo & Nei 1988, Wu 1991). Thus, it is important to be aware of this issue when using sequence data to infer species trees, and several authors have concluded that it is best to infer species trees from several independent gene trees (Pamilo & Nei 1988, Wu 1991). The problem with the sole usage of mtDNA is that because it is inherited as a single linkage group, it does not provide independent estimates of a species tree. However, Moore (1995) argues that since coalescence time is directly related to effective population size and this is low for mtDNA, mtDNA trees have a substantially greater chance of tracking the species tree correctly in comparison to nuclear autosomal DNA. He calculates that when the internode is so short that inference based on mtDNA is in doubt, a prohibitively large number of independent nuclear-gene trees would be necessary to improve confidence in the inferred species tree.

Pseudogenes

Nuclear copies of mtDNA have been discovered in over 64 animal species (Bensasson *et al* 2001). PCR primers sometimes amplify these nuclear mitochondrial pseudogenes (Numts) in addition to, or in preference to the real mtDNA region of interest (Collura & Stewart 1995, Sorenson & Quinn 1998). If this goes undetected, it may result in the incorrect inference of relationship. However, such problems are not related to the use of mtDNA alone. Gene duplication of nuclear genes occurs (Ohno 1970, Taylor & Raes 2004), and can similarly result in mistaken orthology. Due to the existence of numts and following the recommendations of Zhang & Hewitt (1996) sequences should be checked for unexpected deletions/insertions, stop codons, and frameshifts. Unusual results should be treated with caution, and PCRs should be checked by gel analysis to investigate the possible presence of more than one band. In most instances these precautions should be enough to prevent paralogous gene comparisons being made.

Introgression

Introgression describes the process in which the genes of one species or population infiltrate the gene pool of another. This can occur with mtDNA, whereby the mitochondrial haplotype of one species or population is found in an individual with nuclear markers from another (Powell 1983, Harrison *et al* 1987, Lehman & Wayne 1991). A classic case is that of *Mus musculus* and *M. domesticus* (Ferris *et al* 1983) where mtDNA of one species was found in the second, in an area far from the known contact zone. This has obvious implications for estimation of the species or gene tree, and Nei (1987 p286) states: “once it occurs, the phylogenetic relationship of mtDNAs will be quite different from the phylogenetic tree of the species involved”. Although the occurrence of introgression cannot be identified on the basis of mtDNA alone, since it requires comparison to the distribution of nuclear genes, indication that it might have occurred is possible. For instance, Rokas *et al* (2003b) identified possible evidence of introgression between gall wasp species when they uncovered polyphyly of species and clustering of related species within refugia for phylogenies constructed with partial cytochrome *b* haplotypes. However, for within population studies as implemented in this thesis, it is perhaps unlikely that introgression will be a significant source of error, since gene flow between refugial populations is likely to be limited. It is more likely that introgression could mask the origins of parasitoids within northern Europe, where range-expansion from several refugia may have occurred. Nevertheless no evidence of within species introgression was encountered for the gall wasp hosts using a combination of mtDNA and nuclear markers.

Endosymbionts

The presence of endosymbiotic bacteria in parasitoid populations is a further means by which mtDNA gene trees could potentially be incongruent with population trees. Infection by an endosymbiont can reduce mtDNA diversity through the action of a selective sweep of the mitochondrial genome associated with the infected individual. The best-known endosymbionts that cause selective sweeps on host mtDNA are probably male killing and cytoplasmic incompatibility causing *Wolbachia*. During a selective sweep imposed by an endosymbiont, nuclear diversity is preserved since uninfected males are able to mate with infected females, but due to the maternal inheritance of mitochondria, mtDNA diversity is reduced. Thus, substructure evident in mtDNA can be erased, giving the impression of an alternative history. Due to this, Hurst & Jiggins (2005) conclude in a recent paper that symbionts are common enough to make it unsafe to infer patterns of genome history based on mtDNA sequence data alone, and it was for this reason nuclear markers were targeted in addition to mtDNA in this thesis (see above and Appendix 1). Although informative data for nuclear markers were ultimately not obtained, interpretations are restricted to patterns unlikely to be compatible with symbiont-generated artifacts in this thesis. Furthermore, an important point is that not all *Wolbachia* and other symbionts cause selective sweeps, so detection of *Wolbachia* alone does not imply that a selective sweep has occurred. A sensible precaution is thus to screen target species for *Wolbacchia*. If absent a selective sweep by this endosymbiont can be excluded, and if present it is still necessary to show presence of a single *Wolbachia* strain (see Chapter 3 for *M. stigmatizans*).

2.3 Laboratory techniques

Here I briefly summarise the field and laboratory techniques that were applied in generation of the data included in this thesis.

Specimen collection

Galls were collected during the following periods of fieldwork: Hungary, Austria, Germany, & the Netherlands (October 2001), Spain & France (August 2002), Iran (November 2002), UK (February 2003), Northern Italy, Austria, & Hungary (February 2003), Southern Italy (August 2003). In addition, other specimens were available that had been collected on trips undertaken by other members of the gall wasp research group.

Rearing

Galls were placed in either individual or mass rearings. Emerging adults were preserved either in 99% ethanol and stored at 4°C, or frozen and stored at -80°C until DNA extraction or allozyme screening.

Identification of specimens

Many of the specimens used in this study had been previously identified by Dr RR Askew (University of Manchester, retired), or by Dr C Thuroczy or Dr G Melika (Systematic Parasitoid Laboratory, Korzeg, Hungary). However, all specimens were identified again by the author previous to usage. The main key used to determine specimen identification was Askew & Thuroczy (2005). However, difficult torymid specimens were often identified using Graham & Gijswijt (1998).

DNA extraction

DNA was extracted from single adult females using either the DNAeasy Tissue kit (Qiagen cat. 69504), following the manufacturer's protocol for insect DNA extraction, or with 5% Chelex solution. Chelex resin beads are composed of a styrene divinylbenzene copolymer containing paired iminodiacetate ions, which act as chelating groups for polyvalent metal ions. It has been suggested that the presence of Chelex prevents the degradation of DNA by chelating metal ions that may catalyze the breakdown of DNA (Singer-Sam *et al* 1989, Walsh *et al* 1991). When the DNAeasy kit was used, the whole wasp was macerated during extraction. However, when using chelex, typically just the rear leg or several legs were used so that the rest of the specimen could be preserved. This method is useful in case molecular work for a certain specimen gives an unusual result, and re-examination of morphological characteristics is desirable. An additional advantage is that much tissue remains in case further DNA extraction is required in the future.

Polymerase Chain Reaction (PCR)

A 433 base pair (bp) fragment of the mitochondrial cytochrome *b* (cyt*b*) gene was amplified in two parasitoids (*Eurytoma bruniventris* and *Megastigmus dorsalis* Chapter 4) using the primers CB1/CB2 (Jermin & Crozier 1994). This fragment was chosen for consistency with the gall formers. When amplification of this fragment failed (as was the case for *Megastigmus stigmatizans* Chapters 3 and 4), an overlapping fragment of 643 bp was used from the CP1/CP2 primer pair (Harry *et al* 1998). CP1/CB2 can also be combined to produce a long fragment that appears to amplify very well for a broad range of species. The PCR conditions used to generate these fragments were as follows: 35 cycles of denaturation at 92°C for 60s, annealing at 46°C for 60s, extension at 72°C for 120s. PCR reactions had a total volume of 25µl, with 1µL DNA extract, 2.5 µL PARR buffer, 1µL 50mM $MgCl_2$(final concentration 2mM), 0.5µL 10mM dNTP's, 0.35µL of each primer (each at 20mM), 0.25µL *Taq*, and 19.05µL of ultrafiltered water.

Where the mitochondrial cytochrome oxidase I (COI) gene is used (Chapters 3, 5 and 6), amplification was carried out with using the universal primers LCO1490/HCO2198 (Folmer *et al* 1994) to produce a 654 bp fragment. PCR conditions were as follows: 5 cycles of denaturation at 94°C for 30s, annealing at 45°C for 90s, extension at 72°C for 60s, followed by 35 cycles of denaturation at 94°C for 30s, annealing at 51°C for 90s, extension at 72°C for 60s. PCR reactions had a total volume of 25µl, with 1µL DNA extract, 2.25 µL PARR buffer, 1.25µL 50mM $MgCl_2$(final concentration 2mM), 0.125µL 10mM dNTP's, 0.25µL of each primer (each at 20mM), 0.1µL *Taq*, and 19.75µL of ultrafiltered water.

COI was chosen as an additional mtDNA gene to provide further resolution in Chapter 3 since variation in cytochrome *b* proved very low. Other mtDNA genes were also targeted (ND4, the A-T rich region, and 12s rRNA), however, these either did not amplify, or variation was too low to provide a useful degree of resolution for the purposes required here. COI proved very easy to amplify and had an equivalent level of variation to that of cytochrome *b*. It was also employed in Chapter 6 as a means to attempt to barcode oak gall wasp community members. It is a suitable choice for barcoding purposes since it is present in all animal taxa, contains a suitable level of diversity to differentiate between most species, and because it is a mitochondrial gene it is present in a high copy number per cell. Furthermore, the primers used (LCO1490/HCO2198) are heavily conserved across animal taxa and provide a suitable length of sequence for identification purposes.

A 643 bp fragment of the 28S ribosomal RNA gene was amplified to help resolve higher-level relationships within the species of the Torymidae examined in Chapter 5. This gene evolves relatively slowly and amplifies well across a range of insect taxa and has been shown to be a suitable locus for examining higher-level relationships in insects (Caterino *et al* 2000). The primers rD3A/rD5B1 (Friedrich & Tautz 1995) were used, which amplify the domain 3-5 expansion segments of the 28S gene. PCR conditions were: 35 cycles of denaturation at 92°C for 60s, annealing at 48°C for 60s, extension at 72°C for 120s. PCR reactions had a total volume of 25μl, with 1μL DNA extract, 2.5 μL PARR buffer with $MgCl_2$, 0.5μL 10mM dNTP's, 0.35μL of each primer (each at 20mM), 0.25μL *Taq*, and 20.05μL of ultrafiltered water.

Assessment and clean-up of PCR products

PCR products were checked by running 2μL of PCR reaction through a 1% agarose gel stained with ethidium bromide. Successfully amplified products were cleaned prior to sequencing using shrimp alkaline phosphatase and exonuclease I.

Sequencing

All loci were sequenced in both the forward and backward direction for each specimen. Sequencing was carried out using ABI BigDye 3.1 reagents on an ABI 3730 sequencer.

Wolbachia screening

Megastigmus stigmatizans specimens were screened for *Wolbachia* as part of Chapter 3. The primers used to detect the presence of *Wolbachia* were *wsp* 81F/*wsp* 691R (Braig *et al* 1998) for the *Wolbachia* major surface protein gene. PCR conditions were: 35 cycles of denaturation at 92°C for 60s, annealing at 58°C for 60s, extension at 72°C for 45s. PCR reactions had a total volume of 20μl, with 1μL DNA extract, 2 μL PARR buffer, 2μL 25mM $MgCl_2$(final concentration 2mM), 0.5μL 10mM dNTP's, 0.5μL of each primer (each at 20mM), 0.25μL *Taq*, and 13.5μL of ultrafiltered water. In order to assay for *Wolbachia* PCR products were run on an agarose gel alongside a control that was known to be positive for infection.

2.4 Computer techniques

Sequence chromatograms were initially checked after sequencing using the program Sequence Navigator version 1.0.1. After being checked, forward and backward sequences were aligned using the DNA*II package version 1.64 to provide an additional check for errors. After this they were aligned with other sequences. In this way the error level for the sequences used in this thesis was kept very low. All sequences were of maximum length and there was no missing data for new sequences generated during the lab work associated with the thesis.

Unless otherwise stated, the alignments produced in this thesis were constructed by eye. Most sequences were derived from mtDNA and contained no indels, and alignment was largely routine. Alignments were checked by translation of each sequence using the invertebrate mitochondrial genetic code. The presence of a stop codon within a sequence, or the discovery of sequences of different length might have been an indication of pseudogenes as discussed above. However, no such sequences were discovered, even though they have been reported to be a problem in the gall wasp hosts (Rokas 2003b).

Phylogenetic relationships were estimated using Bayesian inference in MrBayes version 3.0b4 (Huelsenbeck & Ronquist 2001). The best-fit model of sequence evolution was estimated for each data set using hierarchical likelihood ratio tests as implemented in MrModeltest version 2.2 (Nylander 2004). Genetic distances were calculated in PAUP* version 4.0b10 (Swofford 2002). Trees were illustrated using Treeview (Page 2001) version 1.6.6 and Adobe Illustrator version 11.0.0.

Chapter 3

Comparative phylogeography across two trophic levels: the oak gall wasp *Andricus kollari* and its chalcid parasitoid *Megastigmus stigmatizans*

[Submitted in a modified form as a paper to *Molecular Ecology*, for which I was senor author]

3.1 Introduction

This Chapter provides an analysis of the population structure of the parasitoid *Megastigmus stigmatizans* across a known oak divide in southwestern France. This is contrasted with what is already known for its host *Andricus kollari*. A more in-depth look at the population structure of *M. stigmatizans* across Europe will follow in Chapter 4.

There is a growing interest in the impact of insect natural enemy population structure on that of their herbivorous insect hosts. Herbivorous insects are attacked by rich communities of parasitoid wasps, and although these constitute approximately 20% of insect species (LaSalle & Gauld 1991), the factors structuring their populations remain little known. Parasitoid communities centred on a given host group (such as galling or leaf-mining insects) are often both species-rich and complex (Askew & Shaw 1986), and theoretical and empirical studies show that understanding the host specificity and population structure of natural enemies is key to understanding community stability and evolution (Thompson and Cunningham 2002, Kondoh 2003, Stone & Schönrogge 2003). Key questions in this field include the extent to which parasitoid and host insect populations are structured on similar spatial scales, and the extent to which these structures are generated by traits of the insect prey (such as the plants they feed on) or by other features of their environment that could influence enemy foraging decisions (Hayward & Stone 2005). Changes in host distributions over a range of spatial and temporal scales have the potential to leave subsets of natural enemies behind, allowing a host to enter 'enemy free space' (Jeffries & Lawton 1984). Here the ability of a specific natural enemy to track host populations on two temporal and spatial scales – over millenia between glacial refugia in Europe, and over hundreds of years in response to human manipulation of the geographic distribution of a foodplant of the insect host is examined. The insect host is an oak-feeding gallwasp, *Andricus kollari* (Hartig, 1843) (Hymenoptera, Cynipidae), and the enemy is an associated parasitoid wasp, *Megastigmus stigmatizans* (Fabricius, 1798) (Hymenoptera, Torymidae).

As summarise in Chapter 1, oak gallwasps have two generations in their annual lifecycle, and *Andricus kollari* also shows obligate alternation between two oak groups. A spring sexual generation always galls oaks in *Quercus* section Cerris, while a summer/autumn asexual generation always galls oaks in *Quercus* section Quercus. The natural distribution of *Andricus kollari* is defined by the restriction of its section Cerris oak hosts to regions of southern Europe corresponding to glacial refugia (Stone *et al.* 2001), while suitable section Quercus hosts (including *Quercus robur* and *Q. petraea*) are native as far north as Scotland and Finland. Natural southern populations of *A. kollari* show refuge-specific associations with different oak species. Iberian populations have sexual generations on the cork oak, *Quercus suber*, and asexual generations primarily on *Q. faginea* and *Q. pyrenaica.* Central European populations (those native to Italy and the Balkans) have their sexual generation on Turkey oak, *Q. cerris*, and an asexual generation on a range of section Quercus oaks. Genetic exchange between eastern (*Q. cerris)* and western (*Q. suber*) refugial populations has been rare enough to allow substantial genetic divergence between them (Stone *et al.* 2001), to the extent that the races probably constitute separate biological species (Pujade-Villar *et al.* 2003). Analysis of mitochondrial haplotype sequences shows the Iberian populations to form a single monophyletic lineage within a wider diversity of central European linages (Stone *et al.* 2001).

Secondary contact between these two gallwasp races was made possible by human planting of *Quercus cerris* as far west as the natural northern limits of *Quercus suber*, near the Gironde estuary in southwestern France (Stone & Sunnucks 1993). Dispersal of Turkey oak was quickly followed by invasion of northern Europe by *Andricus kollari* and other host-alternating gallwasps, and this species is now common as far north as Scotland (Stone *et al.* 2001, Stone *et al.* 2002). A striking feature of range expansion by *Andricus kollari* is that even though Iberian populations are exposed to *Q. cerris* in southwestern France, they were unable to make the host shift onto this oak and so expand their ranges northwards. All colonists reaching northern continental Europe are derived from populations native to Italy and the Balkans, with a long evolutionary history of exploiting *Q. cerris* (Stone *et al.* 2001). Similarly, populations from Italy and the Balkans have been unable to make the southwards host shift to *Q. suber*, and so are excluded from Iberia. The result of this hostplant specificity is a sharp contact zone in southwestern France (Stone *et al.* 2001). This zone provides an opportunity to test the impact of gallwasp population structure on that of their parasitoid natural enemies (Askew 1984, Stone *et al.* 2002). While the population genetic structure and colonisation history of some European oaks and gallwasps have been studied in detail (Stone & Sunnucks 1993, Manos *et al.* 1999, Petit *et al.* 2002, Stone *et al.* 2001, Rokas *et al.* 2003a), very little is known about the population genetics of the next trophic level. One key feature for parasitoids attacking insect herbivores is the plant on which the insect hosts feed (Rott and Godfray 2000, Lill 2002, Wimp *et al.* 2005). Here, the population structure of *Megastigmus stigmatizans*, an important parasitoid of the asexual generation of *Andricus kollari* throughout its native and invaded ranges in continental Europe is analysed. This parasitoid attacks only the asexual generation galls of host-alternating *Andricus* species, and was thus also absent from northern Europe prior to human dispersal of *Quercus cerris*. The *Q. suber* / *Q. cerris* contact zone also represents a transition between alternative host oaks for the asexual generation of *A. kollari*, from *Q. canariensis, Q. faginea* and *Q. pyrenaica* in the south of the Iberian Peninsular to *Q. robur* and *Q. petraea* in northern Europe (Jalas & Suominen1987). For both Iberian and Central European populations of *Megastigmus stigmatizans*, dispersal across the *A. kollari* contact zone thus requires a shift from one group of oaks to another. The little work there has been on oak gall parasitoid population genetics suggests that genetic differentiation can occur within species between sympatric populations on different oak taxa (Ács *et al.* 2002, see Chapter 4). A potentially significant issue for *M. stigmatizans* is that the asexual generation galls induced by the two lifecycle races of *A. kollari* differ in phenology: the galls of the central European race mature earlier, in late July/ early August, than do those of the Iberian race (Stone *et al.* 2001). *Megastigmus stigmatizans* is an idiobiont parasitoid, and has a long drilling ovipositor with which it attacks galls late in their development. The female *M. stigmatizans* stings the gallwasp larva, whose size at this point determines the resources available to the young parasitoid for growth. Close phenological matching between local populations of parasitoid and host could potentially prevent range expansion in either direction: Iberian populations could be prevented from moving northwards by precocious emergence of northern gallwasp hosts, while central European populations could fail to develop on slower-developing Iberian galls because the smaller resource availability prevents successful development.

Here haplotype sequence data for a 643 base pair fragment of the mitochondrial cytochrome b gene and a 654 base pair fragment of the mitochondrial cytochrome oxidase gene is used to examine the population structure of *M: stigmatizans* on either side of the *A. kollari* contact zone. Additional resolution is also provided for this zone in *Andricus kollari.* The following questions are posed:

1. What is the relationship between Iberian populations of *M.stigmatizans* and those further east, beyond the native range of *Q. suber*? Does the intraspecific phylogeny of this species suggest range expansion into Iberia from central Europe and, if so, by one lineage (as for its host) or by many lineages?

2. How does the inferred date of most recent common ancestry of Iberian lineages of *M. stigmatizans* compare with *Andricus kollari*? Did these taxa disperse into Iberia at the same time? If not, did the gallwasp arrive in a region already occupied by the parasitoid, or did the parasitoid pursue its host?

3. How has the anthropogenic contact zone influenced the population structure of *M. stigmatizans*? Are *M. stigmatizans* in northern Europe derived, like the gallwasps, entirely from populations in the native range of *Q. cerris*, or have Iberian populations used the new host to escape their refugial distribution? Have central European *M. stigmatizans* populations been able to invade Iberia?

3.2 Materials and Methods

Sampling

Asexual generation *A. kollari* galls were collected from sites across Iberia and France during August 2002 (Table 1, Fig.1). The sampling scheme covered a north-south transect from northwest France to central Spain, and an east-west transect lying just north of the Pyrenees mountains. The scheme was designed to allow a detailed comparison with the pattern already observed for *A. kollari,* in which a clear divide between populations was apparent in the area surrounding the Gironde estuary. Galls collected at this stage had already been exposed to attack by *M. stigmatizans*, but adults had yet to emerge. Emerging adult *A. kollari* and *M. stigmatizans* were preserved in 99% ethanol and stored at 4°C until DNA extraction.

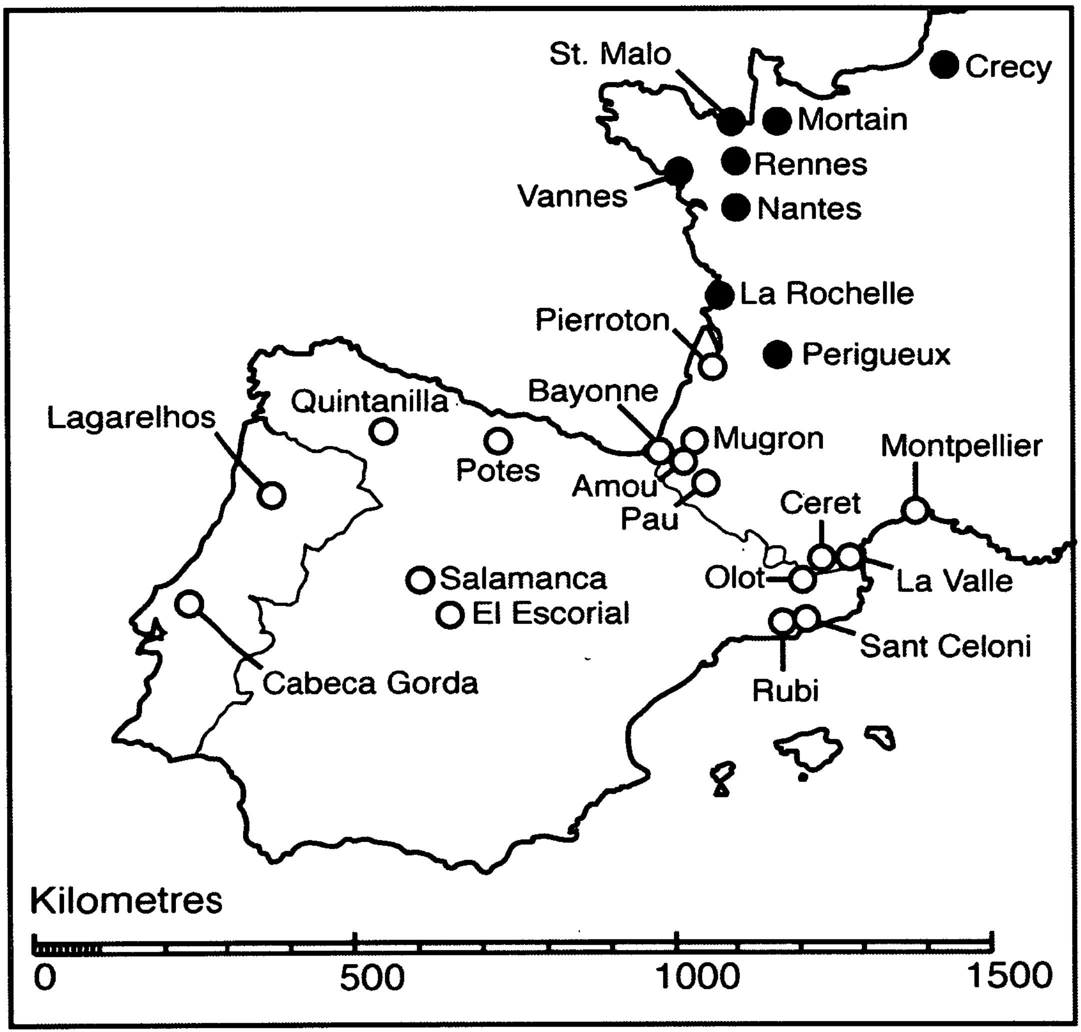


Figure 1. Sampling sites. Filled circles indicate samples collected in the invaded range and clear circles indicate those collected in Iberia, up to and including southwestern France.

Table 1. Sampling locations of Andircus kollari and Megastigmus stigmatizans with haplotypes recovered from each site. The number immediately following 'H' refers to the specific haplotype number, the number in brackets indicates how many individuals were found to possess each haplotype at each site.

Region	Site name	Latitude	Longitude	*A. kollari* Haplotypes	*M. stigmatizans* Haplotypes
NW France	Crecy	50°13'48"N	1°49'48"E	H19(1), H17(1), H15(2), H8(1)	-
NW France	Rennes	48°6'0"N	1°40'12"W	H11(1), H9(1), H2(1)	H10(7), H11(1)
NW France	La Rochelle	46°10'0"N	1°8'60"W	H10(1), H3(1)	-
NW France	Mortain	48°22'0"N	0°55'60"W	H21(1), H19(1)	-
NW France	St Malo	48°38'60"N	2°1'0"W	H20(1), H7(1)	H10(2), H9(1)
NW France	Vannes	47°40'0"N	2°45'0"W	H12(1), H4(1), H1(1)	H9(1)
NW France	Nantes	47°13'48"N	1°34'48"W	H18(1)	H10(2), H9(2)
C France	Perigueux	45°10'60"N	0°43'0"E	-	H9(1)
SW France	Bayonne	43°28'60"N	1°28'60"W	H26(1), H23(2)	H10(1)
SW France	Mugron	43°45'0"N	0°45' 0"W	H23(1), H22(1)	-
SW France	Pierroton	44°43'60"N	0°40'60"W	H32(1), H29(1), H26(1)	-
SW France	Pau	43°17'60"N	0°22'0"W	H30(1), H23(1)	H10(3)
SW France	Ceret	42°28'60"N	2°45'0"E	H36(3)	H12(1)
SW France	Amou	43°36'0"N	0°45'0"W	H24(1), H23(1)	H13(1)
SW France	Montpellier	43°36'0"N	3°52'60"E	H28(1)	-
SW France	La Valle	42°31'60"N	2°57'0"E	-	H10(1)
Spain	Olot	42°10'60"N	2°28'60"E	H26(1), H23(2)	H10(1)
Spain	Rubi	41°28'60"N	2°01'60"E	H35(1) H32(1)	-
Spain	Sant Celoni	41°41'07"N	2°29'57"E	-	H10(1)
Spain	El Escorial	40°34'60"N	4°9'0"W	H33(1), H31(1)	-
Spain	Potes	43°8'60"N	4°37'0"W	H27(1), H23(1)	-
Spain	Salamanca	40°58'0"N	5°39'0"W	H26(1)	H10(2), H14(1)
Spain	Quintanilla	42°46'60"N	5°50'60"W	H25(1)	-
Portugal	Cabeca Gorda	37°55'0"N	7°47'60"W	-	H10(1)
Portugal	Lagarelhos	41°2'60"N	8°4'0"W	H34(1)	H15(1)
Italy	Ruffeno	44°12'0"N	10°55'0"E	H15(1), H16(1)	-
Italy	Greve	43°34'48"N	11°19'12"E	H6(1)	-
Italy	Pruno	43° 58' 60N	10° 13' 0"E	-	H6(1), H7(1)
Italy	Chiusi	43°1'0"N	11°57'0"E	-	H8(1)
Hungary	Godollo	47°36'0"N	19°19'48"E	H5(1), H7(1), H13(1)	H3(1), H4(1)
Hungary	Sopron	47°40'12"N	16°34'48"E	-	H5(1)
Turkey	Antalya	36°52'48"N	30°42'0"E	H14(2)	-
Turkey	Beybesli	41°1'0"N	35°54'0"E	-	H1(1)
Turkey	Refahiye	39°54'0"N	38°45'0"E	-	H2(1)

Molecular methods

DNA was extracted and amplified as described in Chapter 2 of this thesis. For the *A. kollari* cyt*b* data set, 27 new sequences were added to the 27 existing sequences of Stone *et al.* (2001), plus two outgroups in the same *Andricus* species group, *A. conglomeratus* and *A. polycerus* (Cook *et al.* 2002, Rokas *et al.* 2003b). For the *M. stigmatizans* data set 38 new sequences were combined with 3 Torymidae outgroups – *Torymus auratus, Megastigmus dorsalis* and *M. synophri*.

Since analyses are restricted to the use of mtDNA, a consideration of the possible effects of endosymbiont-generated artifacts is necessary. The action of endosymbionts can result in incorrect inferences being drawn for mtDNA studies and it is good practice to use at least two independent markers. However, for the reasons outlined in Chapter 2, this was not successful. As described below, *Wolbachia* were detected in *M. stigmatizans* (see Chapter 2 for methods of *Wolbachia* screening). It is not known whether this infection has a phenotype (*eg* male killing, cytoplasmic incompatibility) likely to cause a selective sweep. Were such a sweep to have taken place over any refugial structure in sequence diversity, we would expect either: (a) much lower mtDNA sequence diversity in regions where *M. stigmatizans* is infected, or (b) if all *M. stigmatizans* are infected, loss of any refugial signal in mtDNA variation. A selective sweep prior to refugial subdivision is expected to reduce diversity in the infected populations as a whole, but refugial substructure should still be visable in subsequent refuge-specific sequence evolution. The patterns observed in *M. stigmatizans* are discussed below in the light of these possible symbiont-mediated outcomes.

Phylogenetic analyses

All sequences were the same length, and were aligned by eye. A partition homogeneity test for the cyt*b* and coI fragments in PAUP* (Swofford 2002) confirmed that these sequences contained a compatible signal and could justifiably be combined into a single data set (P = 1.00, 100 replicate branch-and-bound search). Phylogenetic relationships were estimated using maximum likelihood (ML) in PAUP* version 4.0b10 (Swofford 2002), and Bayesian inference (BI) in MrBayes version 3.0b4 (Huelsenbeck & Ronquist 2001). The best-fit model of sequence evolution was estimated for each data set using hierachical likelihood ratio tests as implemented in Modeltest version 3.06 (Posada & Crandàll 1998) for the maximum likelihood analysis, and MrModeltest version 2.2 (Nylander 2004) for the Bayesian analyses. The most appropriate model of sequence evolution for *A. kollari* was the Hasegawa-Kishino-Yano (1985) model with unequal base frequencies, gamma distributed rate heterogeneity (Γ)(Yang 1993) and invariant sites (I). The locus displayed a strong A-T bias, a high transition/transversion ratio (ti/tv = 12.4835), a gamma shape estimate of 0.7435, and a high proportion of invariant sites (0.7126). For *M. stigmatizans* the General Time-Reversible model (Lanave *et al.* 1984, Rodriguez *et al.* 1990) + Γ was selected for the cyt*b* and coI data sets when tested individually and in combination. The combined data set also had a strongly A-T biased base composition, and the gamma shape parameter estimate was 0.2257.

All amplified fragments were from protein-coding regions, and the data sets for both species revealed a high bias in relative site rates according to codon position (*A. kollari* pos1: 0.13, pos2: 0.04, pos3: 2.82; *M. stigmatizans* pos1: 0.51, pos2: 0.18, pos3: 2.31). Under a codon-based model of site-specific rate heterogeneity, implemented by partitioning according to codon position and allowing rates to vary across partitions, tree topology was the same but likelihoods (both ML and BI) were higher relative to models with just Γ (*M. stigmatizans*) or Γ+I (*A. kollari*). Also, Bayes factors indicated very strong evidence for the selection of site-specific models for both data sets according to the guidelines of Raftery (1996). Therefore, a codon-based model of site-specific rate heterogeneity was implemented for both data sets. Currently it is not possible to implement a codon-based model of site-specific rate heterogeneity when performing a bootstrap search in PAUP. Although bootstrap support values have been observed to be lower than corresponding measures of support derived from Bayesian posterior probabilities (Wilcox *et al.* 2002), this may account for the lower relative bootstrap support in comparison to the Bayesian support observed.

ML analyses were performed using a neighbour-joining (NJ) tree as a starting-point, followed by tree-bisection-reconnection (TBR) branch swapping (Swofford *et al.* 1996), specifying the parameters as determined by Modeltest. Bootstrap support was determined using 100 bootstrap replicates and TBR branch swapping. BI analyses were performed using the $(MC)^3$ algorithm, with four simultaneous Markov chains (three heated, one cold) as implemented by default in MrBayes 3, for 2 million generations. The best-fit model identified by MrModeltest was specified, with substitution model parameters (stationary base frequencies, substitution rates between states, rate variation across sites) estimated during the analyses. Analyses were started from random trees and trees were sampled every 100 generations, with the first 10% discarded as burn-in. Bayesian posterior probabilities were estimated for each clade from the 50% majority-rule consensus tree of the 18,000 sampled trees. Four independent runs were carried out to compare topology and support, in order to check that the chain had been run for long enough. There were no conspicuous differences between the results and chain length was judged to be sufficient. Acceptance rates were monitored to check that mixing was adequate, and a plot of log-likelihood scores against generation was examined in each case to check that stationarity had been achieved within the first 200,000 generations. Topology was identical in every case, with support varying only slightly between runs.

Molecular clock analysis

A molecular clock approach was adopted in order to date the time depth of the Iberian clade in both *A. kollari* and *M. stigmatizans*. A likelihood ratio test (LRT; Felsenstein 1981) was conducted for both species to assess whether the assumption of a constant rate of evolution across each tree could be upheld. The LRT statistic is defined as: LR = 2(lnL1-lnL2), where the null hypothesis is a homogeneous rate of sequence evolution. Trees were generated using ML in PAUP with and without a molecular clock enforced. The results supported a molecular clock model for both species (*A. kollari* χ^2 = 47.54, df=54, P=0.72; *M. stigmatizans* χ^2 = 36.48, df=39, P=0.59). A relative rate test was conducted using the program RRTree version 1.1 (Robinson-Rechavi & Huchon 2000), to test for a constant rate of evolution between species. Relative rate tests use a distance-based approach to compare the relative rate of evolution between the test sequences and a reference taxon (Sarich & Wilson 1973). The analysis was weighted according to topology, using a NJ phylogram constructed in PAUP for the *A. kollari* and *M. stigmatizans* cyt*b* sequences and a honeybee outgroup (*Apis melifera*). The phylogram revealed similar branch lengths between and within the *A. kollari* and *M. stigmatizans* clades, and Ka (the number of non-synonymous substitutions per non-synonymous site) was found to be highly non-significant (P=0.67), suggesting that the two lineages evolve at the same rate. However, it was not possible to calculate Ks (the number of synonymous substitutions per synonymous site), probably due to saturation associated with use of a distant outgroup. An additional analysis using an Ichneumonid wasp as a closer outgroup yielded non-significant values for B4 (the number of synonymous transversions per fourfold degenerate site) and Ka, but calculation of Ks was again not possible. Relationships within the apocrita are still poorly resolved, particularly within the clade that is believed to contain the Chalcidoidea and Cynipoidea. Choice of an outgroup that is both sufficiently close to the ingroups to avoid the problem of saturation, and equidistant from each, is currently difficult. In addition, no cyt*b* sequences for a possibly more suitable outgroup (such as from one of the superfamilies Evanioidea, Megalyroidea, Trigonaloidea or Stephanoidea), are currently available in Genbank. Shao *et al.* (2003) suggest that Ks and B4 are not a useful means of measuring the rate of nucleotide substitution for highly divergent taxa. Thus, even if the sequences were available, it is quite possible that the two superfamilies might prove to be too divergent to obtain meaningful results using this method. Thus, since no data suggested differently, the rate of evolution between the two lineages was assumed to be constant, and a pairwise sequence divergence rate of 2.3% per million years (1.15% per million years per lineage) was adopted, following the general molecular clock estimate for the evolution of arthropod mtDNA proposed by Brower (1994). Clades were dated using the program BEAST (Drummond & Rambaut 2003), which uses a Bayesian MCMC approach to date clades using molecular sequence data. MCMC chain length was 10 million, with parameters logged every

1000 steps. The model of sequence evolution specified was that found by MrModeltest. The data set was partitioned according to codon position, with rate heterogeneity unlinked across codon positions. All other settings were left as default, and no clades were enforced. Multiple independent runs were carried out, with similar results achieved in each. Effective sample size (ESS) values and frequency plots were examined to check that there was adequate mixing of the MCMC chain using Tracer (Rambaut & Drummond 2003). Burnin was also established and results viewed using this program.

To allow meaningful comparisons between *A. kollari* and *M. stigmatizans*, it is necessary to compare the range of possible colonisation dates for Iberian lineages (Fig.2). Both the most recent and most ancient extremes are considered. Consideration of the most recent common ancestor (MRCA) of the Iberian clade ignores demographic effects on genetic diversity that may have acted since colonisation of the Peninsula. Consideration of the most ancient common ancestor (MACA) assumes that colonisation was simultaneous with the appearance of the lineage. If the most recent for one species differes significantly from the most ancient for another, then it is reasonable to argue that the clades in each species originated at different times.

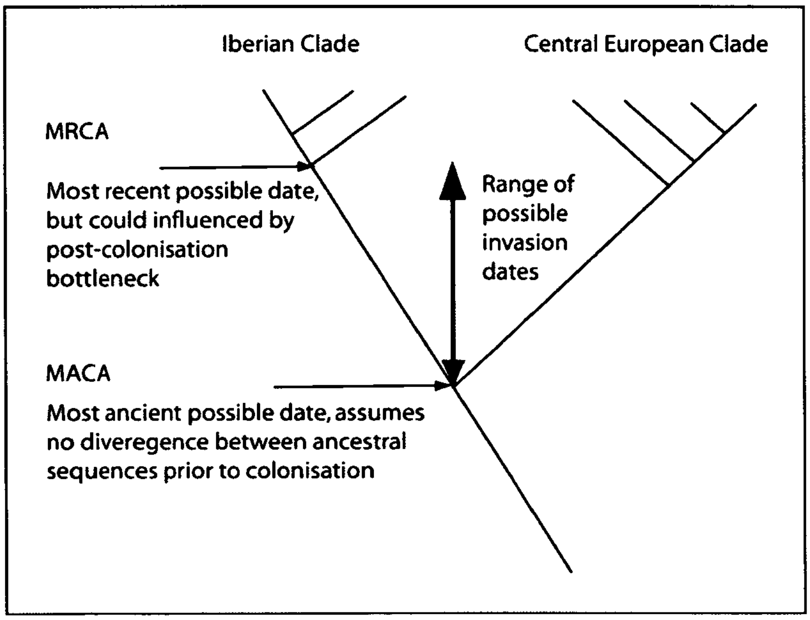


Figure 2. Diagramatic representation of the most recent and most ancient possible dates of common ancestry between a monophyletic Iberian clade and its outgroup. Consideration of the most recent common ancestor (MRCA) of the Iberian clade ignores demographic effects on genetic diversity that may have acted since colonisation of the Peninsular. Consideration of the most ancient common ancestor (MACA) assumes that colonisation was simultaneous with the appearance of the lineage.

3.3 Results

3.3.1 *Andricus kollari*

For *A. kollari*, 54 individuals yielded 36 haplotypes (Tables 1 & 2). Of the 433 sites examined 62 were variable, with 32 of these parsimony-informative. Both Bayesian and maximum likelihood trees for these sequences (Fig. 3) reveal a strongly supported monophyletic Iberian clade that is derived with respect to more eastern sequences. Independent Bayesian runs revealed a congruent overall topology, well supported by Bayesian posterior probabilities. Corrected pairwise distances (Table 4) show that the mean distance between haplotypes sampled within Iberia and those outside this region was 4.1%, almost double the mean pairwise distance for the clade as a whole (2.8%).

Table 2. *Andricus kollari* haplotype table for a 433bp fragment of the cytochrome b gene. The position of variable sites is given in base pairs at the top of the table, under which a consensus sequence is indicated. Deviations are listed as they occur for each haplotype. Question marks indicate missing data.

	7	10	16	32	43	46	49	55	69	76	79	82	86	106	121	124	130	181	193	196	202	205	214	238	241	244	262	271	292	307	308	321	322	323	325	329	337	340	355	376	382	397	421
	a	a	t	a	t	a	a	g	t	a	a	a	t	t	c	t	t	a	a	a	t	t	a	t	t	t	a	a	t	a	t	t	c	c	t	a	a	t	t	t	t	a	a
Haplotype 1	.	.	.	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Haplotype 2	.	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Haplotype 3	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	g	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Haplotype 4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	c	.	.	.	.	.	g	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Haplotype 5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	c	.	.	.	.	.
Haplotype 6	.	.	.	.	.	.	g	.	.	.	.	.	.	.	.	.	.	.	.	g	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	c	.	.	.	.	.
Haplotype 7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Haplotype 8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	c	.	g	.	.	.	.	.	.	.	.	.	.	.	.	?	.	.	.	.	.	.	.
Haplotype 9	.	g	.	.	.	.	.	a	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Haplotype 10	.	g	.	.	.	.	.	a	.	.	.	.	.	a	.	.	.	g	.	.	.	c	g	.	.	.	.	.	.	.	.	.	.	.	c	.	.	c	.	c	.	.	.
Haplotype 11	.	g	.	.	.	.	.	a	.	.	.	.	.	a	.	.	.	g	.	.	.	.	g	.	.	.	.	.	.	.	.	.	.	.	c	.	.	c	.	.	.	.	.
Haplotype 12	.	g	.	.	.	.	.	a	.	.	.	.	.	a	.	.	.	g	.	.	.	.	g	.	.	.	.	.	.	.	.	c	.	.	c	.	.	c	.	.	.	.	.
Haplotype 13	.	g	.	.	.	.	.	a	.	.	.	.	.	.	.	.	.	g	.	.	.	.	g	.	.	.	.	.	.	.	.	.	.	.	c	.	.	c	.	?	.	.	.
Haplotype 14	.	g	.	.	.	g	.	a	.	.	.	.	.	.	t	c	.	.	t	g	.	.	.	.	.	.	.	.	.	.	.	.	t	.	c	.	.	.	.	.	c	g	g
Haplotype 15	.	g	.	.	.	.	.	a	c	.	.	g	.	.	.	.	.	.	.	g	c	c	g	.	.	.	.	.	c	.	.	.	t	t	c	g	.	.	.	.	.	.	.
Haplotype 16	.	g	.	.	.	.	.	a	c	.	.	g	c	.	.	.	.	.	.	g	c	c	g	.	.	.	.	.	c	.	.	.	t	t	c	g	.	.	.	.	.	.	.
Haplotype 17	.	g	.	g	.	.	.	a	c	.	.	g	.	.	.	.	.	.	.	g	c	c	g	.	.	.	.	.	c	.	.	.	t	.	c	g	.	.	.	.	.	.	.
Haplotype 18	.	g	.	.	.	.	.	a	c	.	.	g	.	.	.	.	.	.	.	g	c	c	g	.	.	.	.	.	c	.	.	.	t	.	c	g	.	.	.	.	.	.	.
Haplotype 19	.	g	.	.	.	.	.	a	c	.	.	.	.	.	.	.	.	.	.	g	c	c	g	.	.	.	.	.	c	.	.	.	t	.	c	g	.	.	.	.	.	.	.
Haplotype 20	.	g	.	.	.	.	.	a	c	.	.	.	.	.	.	.	.	.	.	.	c	c	g	.	.	.	.	.	c	.	.	.	t	.	c	g	.	.	.	.	.	.	.
Haplotype 21	.	g	.	.	.	.	.	a	c	.	.	.	.	.	.	.	.	.	.	?	c	c	g	.	.	.	.	.	c	.	.	.	t	.	c	g	.	.	.	.	.	.	.
Haplotype 22	.	g	.	.	.	.	.	a	.	.	g	.	.	.	.	.	c	.	.	.	.	c	g	c	c	c	.	.	.	.	c	.	t	.	.	.	.	.	.	.	.	.	.
Haplotype 23	g	g	.	.	.	.	.	a	.	.	g	.	.	.	.	.	c	.	.	.	.	c	g	c	c	c	.	.	.	.	c	.	t	.	.	.	.	.	.	.	.	.	.
Haplotype 24	g	g	.	.	.	.	.	a	.	.	g	.	.	.	.	.	c	.	.	.	.	c	g	c	c	.	.	.	.	.	c	.	t	.	.	.	.	c	.	.	.	.	.
Haplotype 25	g	g	.	.	.	.	.	a	.	.	g	g	.	.	.	.	c	.	.	.	.	c	g	c	c	.	.	.	.	.	c	.	t	.	.	.	.	c	.	.	.	.	.
Haplotype 26	g	g	.	.	.	.	.	a	.	.	g	.	.	.	.	.	c	.	.	.	.	c	g	c	c	.	.	.	.	.	c	.	t	.	.	.	.	.	.	.	.	.	.
Haplotype 27	g	g	.	.	.	.	.	a	.	.	g	.	.	.	.	.	c	.	.	g	.	c	g	c	c	.	.	.	.	.	c	.	t	.	.	.	.	.	.	.	.	.	.
Haplotype 28	g	g	.	.	.	.	.	a	.	.	g	.	.	.	.	.	c	.	.	g	.	c	g	c	c	.	.	.	.	.	.	.	t	.	.	.	.	.	.	.	.	.	.
Haplotype 29	g	g	.	.	.	.	.	a	.	.	g	.	.	.	.	.	c	.	.	.	.	c	g	c	c	c	.	.	.	.	c	.	t	.	.	.	.	.	.	c	.	.	.
Haplotype 30	g	g	.	.	.	.	.	a	.	g	.	.	.	.	.	.	c	.	.	.	.	c	g	c	c	.	.	.	.	.	.	.	t	.	.	.	.	.	c	.	.	.	.
Haplotype 31	g	g	.	.	.	.	.	a	.	.	g	.	.	.	.	.	c	.	.	.	.	c	g	c	c	c	t	.	.	.	c	.	t	.	.	.	g	.	.	.	.	.	.
Haplotype 32	g	g	.	.	.	.	.	a	.	.	g	.	.	.	t	.	c	.	.	.	.	c	.	c	c	c	.	.	.	.	c	.	t	.	.	.	.	.	.	.	.	.	.
Haplotype 33	g	g	.	.	.	.	.	a	.	.	.	.	.	.	.	.	c	.	.	.	.	c	g	c	c	c	.	t	.	.	c	.	t	.	.	.	.	.	.	.	.	.	.
Haplotype 34	g	g	.	.	.	.	.	a	.	.	g	.	.	.	.	.	c	.	.	.	.	c	g	c	c	c	.	.	.	g	c	.	t	.	.	.	.	.	.	.	.	.	.
Haplotype 35	g	g	.	.	.	.	.	a	.	.	g	g	.	.	.	.	c	.	.	.	.	c	g	c	c	.	.	.	.	.	c	.	t	.	.	.	.	.	.	.	.	.	.
Haplotype 36	g	g	.	.	.	.	.	a	.	.	g	.	.	.	.	.	c	.	.	.	.	c	g	c	c	c	.	.	.	.	.	.	t	.	.	.	.	.	.	.	.	.	.

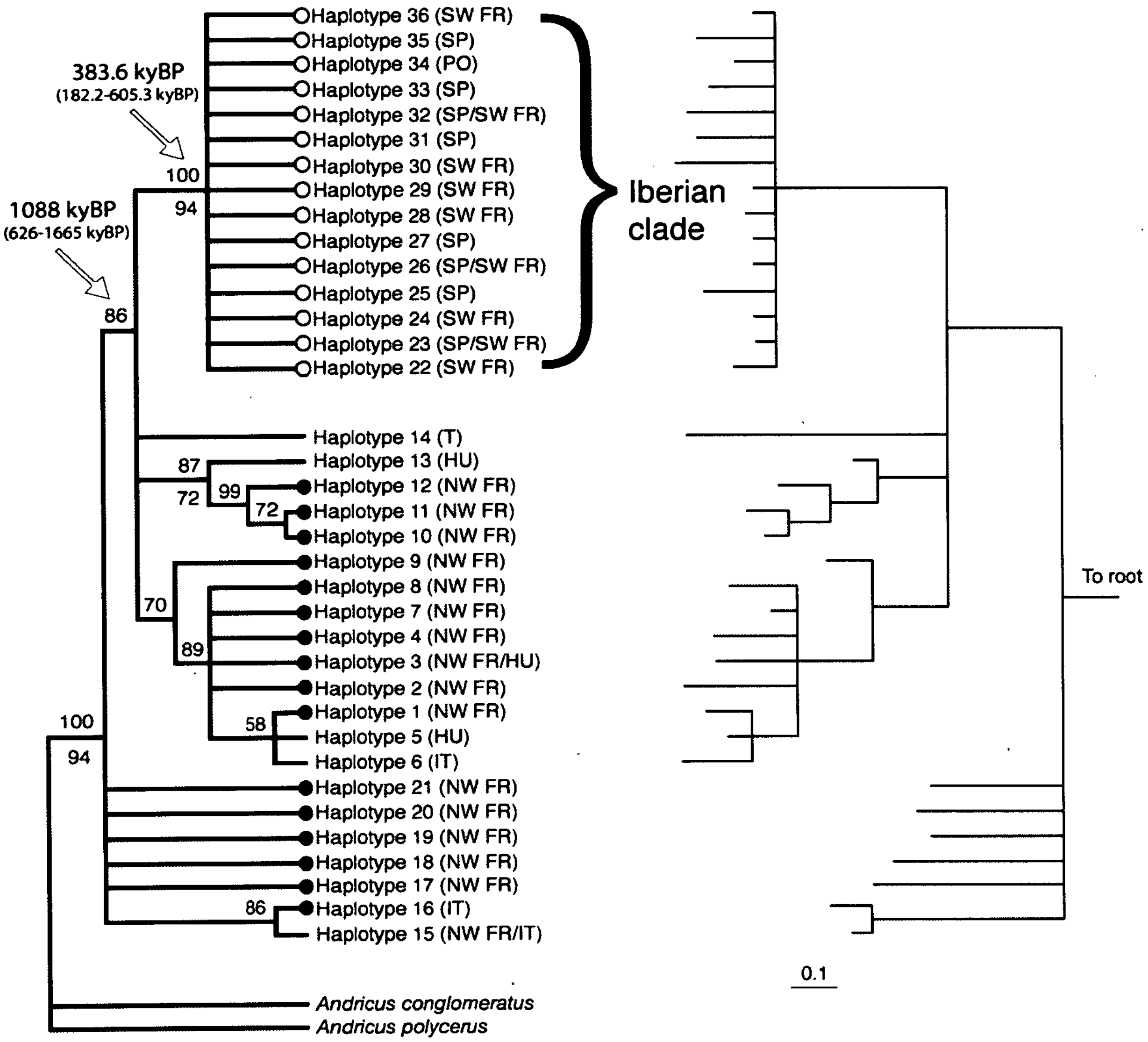


Figure 3. Relationships of *A. kollari* haplotypes based on a 433 bp fragment of the *cytb* gene. The left-hand cladogram displays the sample location of haplotypes collected in the invaded range or Iberia (Filled circle for invaded range, clear circle for Iberia inclusive of southwestern France). Support values are provided on the cladogram with Bayesian posterior probabilites displayed above nodes, and bootstrap proportions below. A phylogram version of the tree is provided on the right, with a scale bar indicating the percentage sequence divergence.

Table 3. *Megastigmus stigmatizans* haplotype table for a 643bp fragment of the cytochrome b gene and a 654bp fragment of the cytochrome oxidase gene. The position of variable sites is given in base pairs at the top of the table, under which a consensus sequence is indicated. Deviations are listed as they occur for each haplotype.

	Cytochrome oxidase I													Cytochrome b																			
	200	223	259	271	274	294	370	373	421	487	559	610	625	658	740	821	824	850	868	883	904	919	950	973	976	1006	1005	1117	1120	1123	1152	1153	1174
	t	t	t	t	a	c	a	t	a	g	a	t	g	c	t	a	g	a	a	a	a	t	g	a	a	c	g	a	c	t	c	a	t
Haplotype 1	.	c	.	.	.	.	g	c	.	a	g	.	.	.	.	.	.	.	.	.	.	a	.	.	.	.	.	t	t	.	.	.	.
Haplotype 2	.	.	.	.	.	.	.	.	g	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Haplotype 3	.	.	.	.	.	t	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Haplotype 4	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	a	.	.	.	.	.	.	.	.	.	.
Haplotype 5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	a	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Haplotype 6	.	.	.	.	.	.	.	c	.	.	.	.	.	t	.	g	.	.	.	.	.	a	.	.	.	.	.	.	.	c	.	g	.
Haplotype 7	.	.	c	.	g	.	.	.	.	.	.	c	a	t	.	.	.	.	g	.	.	a	.	.	.	.	.	.	.	.	.	g	.
Haplotype 8	.	.	c	.	g	.	.	.	.	.	.	.	a	t	g	.	.	.	.	g	g	a	.	g	.	.	.	.	.	.	.	g	.
Haplotype 9	c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Haplotype 10	.	c	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	a	.	.	g	.	.	.	.	.	t	.	c
Haplotype 11	.	c	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	a	.	.	g	.	a	.	.	.	t	.	c
Haplotype 12	.	c	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	a	.	.	.	.	.	.	.	.	t	.	c
Haplotype 13	.	c	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	a	.	.	g	t	.	.	.	.	t	.	c
Haplotype 14	.	c	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	g	.	.	.	a	.	.	g	.	.	.	.	.	t	.	c
Haplotype 15	.	c	.	c	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	g	a	.	.	g	.	.	.	.	.	t	.	c

3.3.2 *Megastigmus stigmatizans*

The 38 *M. stigmatizans* individuals yielded 15 haplotypes (Tables 1 & 3). Of the 1297 sites examined across both loci, only 33 were variable and 13 parsimony informative. For cyt*b*, there were 20 variable sites, of which 7 were parsimony informative, and for coI there were 13 variable sites, of which 6 were parsimony informative. BI and ML analyses revealed trees with a similar topology (Fig. 4), again showing a strongly supported monophyletic Iberian clade that is derived relative to eastern sequences. Corrected pairwise distances show *Megastigmus stigmatizans* to have far lower sequence diversity than it host (Table 4): *A. kollari* sequences have 5-6 times the diversity present in the parasitoid (sequence divergence for the entire *A. kollari* clade varies between 0 and 7%, whereas for *M. stigmatizans* the range is only 0-1.35%). The Iberian clade for *M. stigmatizans* contains a particularly low level of sequence diversity, with 6 haplotypes giving a mean pairwise distance of 0.03%. As for *A. kollari*, the mean pairwise distance between haplotypes sampled within Iberia and those outside this region (0.69%) was almost double the mean distance for the clade as a whole (0.38%).

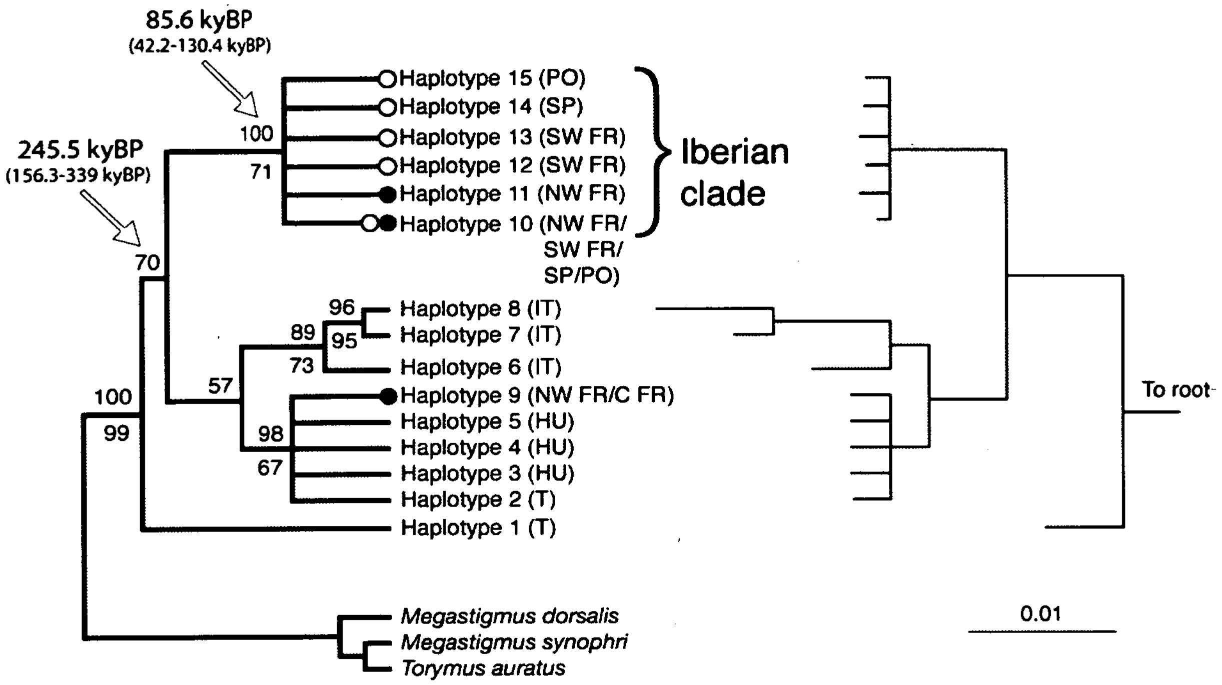


Figure 4. Relationships of *M. stigmatizans* haplotypes based on the combined coxI/cyt*b* dataset. Sample location and support values are provided as with *A. kollari*.

Table 4. Corrected distances (percentage)

		Whole clade	Iberian	Non-Iberian	Iberian versus non-Iberian
A. kollari	Range	0-7	0-1.86	0-5.25	2.5-7
	Mean	2.8	0.53	2.53	4.11
M. stigmatizans	Range	0-1.35	0-0.16	0-1.35	0.48-1.26
	Mean	0.38	0.03	0.45	0.69

Whole clade refers to all sequences of that species analysed. *Iberian* refers to sequences that fall in the Iberian clade. *Non-Iberian* referes to all sequences outside of the Iberian clade. *Iberian versus non-Iberian* is the average distance between Iberian and non-Iberian sequences.

3.3.3 Dating the divergence of the Iberian clades of *A. kollari* and *M. stigmatizans*

BEAST infers very different dates of most recent common ancestry (MRCA) for the gallwasp and its parasitoid. The MRCA for all *Andricus kollari* sequences is dated at 1.327 million years before present (with lower and upper highest posterior density estimates of 780-1906 kyBP), while the MRCA for Iberian *A. kollari* is estimated at 383.6 kyBP (182.2-605.3 kyBP). In contrast, the MRCA for all *M. stigmatizans* sequences is estimated at 262 kyBP (180.9-360.3 kyBP), with an MRCA for Iberian sequences of 85.6 kyBP (42.2-130.4 kyBP).

The most ancient possible common ancestor (MACA) of the Iberian *A. kollari* clade is dated at 1.088 million years before present (626-1665 kyBP), and at 245.5 kyBP (156.3-339 kyBP) for Iberian *M. stigmatizans*. If MRCA estimates accurately reflect relative dates of arrival of the two species in Iberia, *M. stigmatizans* thus followed its host into the Peninsula. Even if we take the earliest probable (MACA) date of arrival of *M. stigmatizans* and the latest (MRCA) for *A. kollari*, it is still unlikely that the two lineages arrived in Iberia at the same time.

3.3.4 Range expansion by Iberian and central European lineages of *A. kollari* and *M. stigmatizans*

No *Andricus kollari* collected north of the Gironde estuary yielded individuals with an Iberian haplotype, and no central European haplotypes were found in Iberia, confirming failure by gallwasp matrilines of either origin to cross the contact zone. In contrast, *M. stigmatizans* collected in the planted range of *Quercus cerris* in northern and central France include haplotypes of both Iberian and central European origin (Fig.4). For example, haplotype 9, which lies in the clade containing sequences from the native distribution of *Quercus cerris* in central Europe, was found in wasps collected in St. Malo, Vannes, Nantes and Perigueux (see Fig.1). This confirms the ability of central European populations of *M. stigmatizans* to pursue their natural gallwasp host. Similarly, haplotypes 10 and 11, which fall within the Iberian clade (Fig.4) were found in wasps collected in Rennes, St. Malo, and Nantes (See Fig.1). This confirms the ability of Iberian *M. stigmatizans* to shift onto a novel invading host, and so, in contrast to Iberian gallwasps, to escape their natural glacial refugium. In contrast, no haplotypes diagnostic of the central European clade were found in Iberia, implying unidirectional range expansion by *M. stigmatizans* across its host contact zone. Populations of *M. stigmatizans* derived from both eastern and western origins possess only a small subset of the genetic diversity present in their native range.

3.3.5 *Wolbachia* screening

A panel of 20 *M. stigmatizans* specimens from across the range were screened for the presence of *Wolbachia*. All specimens tested postive for infection.

3.4 Discussion

The low diversity observed in mtDNA for *M. stigmatizans* suggests that it may have undergone a symbiont induced selective sweep. Furthermore, a panel of *M. stigmatizans* individuals from Iberia to Turkey screened for the presence of *Wolbacchia* all tested positive for infection. The refugial organisation of haplotypes and level of associated mtDNA diversity suggest that this sweep must have occurred prior to longitudinal range expansion. Thus, the interpretation of patterns observed in this study are structured in a way that is unlikely to be compatible with symbiont-generated artifacts.

3.4.1 The Iberian clade, its time depth and genetic differentiation

Iberian haplotypes of both host and parasitoid form well-defined monophyletic clades that are derived with respect to more eastern populations. This implies that the Iberian matrilines of both species are derived from a single lineage with an origin further east, probably in central Europe. For both species, this underlines the rarity of successful past migration between southern glacial refugia, a pattern that is repeated in other oak gallwasps (Rokas *et al.* 2001, 2003a). It suggests, too, that eastern centres of oak gallwasp diversity (Rokas *et al.* 2003a,b, Stone *et al.* 2001, 2002) may also represent sources for range expansion by community members at the next trophic level.

However, the differences in sequence diversity in the Iberian clades of these species strongly suggest a more ancient MRCA for the gallwasp than its parasitoid. If we assume that diversity within these clades represents divergence following invasion of Iberia, without significant impacts of subsequent population bottlenecks, then the MRCA dates inferred by BEAST coincide with known interglacials – the Hoxnian for *A. kollari* and the Eemian for *M. stigmatizans*. If we assume instead that post-colonisation bottlenecks have affected genetic diversity only in the gallwasp, this represents a worst-case scenario for testing simultaneous colonisation of Iberia. While BEAST estimates still support a more anceint origin for Iberian *A. kollari*, the confidence limits for the MACA (Fig.2) of the Iberian *M. stigmatizans* clade (181-360 kyBP) do overlap with the MRCA of the Iberian *A. kollari* clade (200-600 kyBP). It is thus possible that both lineages colonised Iberia at the same time. However, aspects of the biology of the species involved probably predispose the parasitoid, rather than the host, to population bottlenecks. The gallwasp has 2 generations each year, one of which has high fecundity, and both generations can exist at high population densities (Stone *et al.* 2002). In contrast, the parasitoid has only one generation a year, and only attacks a proportion of one host generation in any location.

While the relative estimates of divergence between Iberia and the rest of Europe for these two species are probably correct, their absolute dating must be treated with care: they depend on an average rate calculated by Brower (1994) over seven arthropod taxa from 5 orders in which the rates varied between 1.7% and 4.2% per million years. Rates in some insects lie below this range. For example, Prüser & Mossakowski (1998) estimated a substitution rate of 0.39-0.98%/myr for mtDNA (NADH-dehydrogenase subunit 1 gene fragment) in a group of Mediterranean carabid beetles. No Hymenoptera were included in Brower's study, so it remains unclear whether this rate is appropriate to this group. This is particularly significant given Castro *et al.*'s (2002) finding that the parasitic Hymenoptera show an increased rate of genetic divergence. Our estimates could then represent overestimates of true divergence times. Furthermore, only two mitochondrial gene fragments were analysed for *M. stigmatizans*, and one for *A. kollari*, and dating estimates thus carry with them the usual caveats applied to analysis of mitochondrial sequence data (Hurst & Jiggins 2005). However, the putative most recent dates of range expansion into Iberia do coincide with known interglacials, and so periods in which oak forests (and hence, potentially, gall faunas) allowed dispersal between southern glacial refugia. Furthermore, independent of their absolute clock rate, there is no evidence that mutations are accumulating at different rates in host and parasitoid. In the absence of differential demographic impacts on mitochondrial sequence diversity in host and parasitoid, these data imply that this parasitoid pursued its host into Iberia. If the datings are correct, then *A. kollari* enjoyed one glaciation's worth of enemy free space from *M. stigmatizans*.

An alternative interpretation for the lower genetic diversity in *M. stigmatizans* is that this species colonised Iberia prior to the Eemian interglacial but that mitochondrial sequence diversity underwent a severe bottleneck at this time. Such an impact could result either from a population bottleneck (Avise 2000) or a selective sweep imposed, for example, by *Wolbachia* (Hurst & Jiggins 2005). For such effects to explain low diversity in Iberia alone, we expect the phylogenetic structure of Iberian haplotypes to differ from those sampled elsewhere. If the Iberian lineage is in fact ancient, but with low current diversity due to a bottleneck, Iberian haplotypes should show very little diversity and be joined to their outgroup by a relatively long branch, with substantially greater nucleotide diversity in more eastern sequences. This is not the case. Sequence diversity is low for the entire *M. stigmatizans* clade, suggesting that diversity in this species was reduced by an earlier, more basal demographic event in eastern populations, followed by westwards range expansion.

3.4.2 Recent range expansion

Our results show that *A. kollari* has been pursued by *M. stigmatizans* populations sharing the same native range, and has also been exploited by host-shifting Iberian populations of the same species. It is thus clear that *M. stigmatizans* can jump across gall wasp ecotypes, but gall wasps cannot jump across oaks. This suggests that the cues used by *M. stigmatizans* to locate its gallwasp hosts are not restricted to specific section Quercus oaks. Parasitoids of both origins show low genetic diversity, however, suggesting the establishment of small populations in both cases. More detailed sampling with more polymorphic markers is required to assess the extent of the demographic impacts involved in each case.

More generally, if other parasitoids respond in the same way, then the invasion of northern Europe by *A. kollari* and other host-alternating gallwasps may have major implications for the structure of communities centred on them. While the invading gallwasps are derived from a single eastern origin, their spread potentiates range expansion by two sets of parasitoids, by native pursuit and non native host-shift. Both routes will result in the recruitment of new natural enemies to northern gallwasp communities, with associated impacts on community dynamics. One major impact may be the indirect interactions between native and invading gallwasps mediated by shared natural enemies (Holt & Lawton 1993, Schönrogge & Crawley 2000). While *M. stigmatizans* is almost entirely restricted to the invading species in northern Europe, other parasitoids attacking the invaders have much broader host ranges. High populations of these parasitoids in invading gallwasps could have significant negative impacts on native gallwasp populations. To address these issues, we need to know which other parasitoids are expanding their ranges by both pursuit and host shift routes.

Chapter 4

Phylogeography of European oak gall wasp parasitoids and implications for oak gall wasp communities

4.1 Introduction

This is a complex chapter because it involves a number of interacting structuring processes. Nevertheless, my feeling is that they are best considered together. The chapter consists of analyses of population structure in three parasitoid species. Each species is dealt with in turn including discussion, with a general discussion to close the chapter.

A series of questions regarding the phylogeography and structure of European oak gall wasp parasitoid populations is examined here. In addition to providing knowledge about the parasitoid populations themselves, the questions have specifically been chosen to allow comparison with findings for the gall wasp hosts, as summarised in Chapter 1 of this thesis. Such comparisons are important since they allow us to develop an understanding of the way in which the communities evolve and function. Of specific interest are questions relating to the possibility that oak gall communities as a whole show a common phylogeographic origin, and whether the variability inherent within lower trophic levels exerts an effect upon the structure of higher trophic levels. The answers to these questions have relevance outside gall wasp research, and could have important implications for our understanding of biological communities and ecosystems, discussed in detail below.

4.2 Questions to be addressed

Here I summarise the main questions addressed in this chapter. Alternative hypotheses are then outlined in turn for each question in the next section.

1) What is the structure of oak gall wasp parasitoid populations across Europe?

(a) Is there evidence of population subdivision corresponding to putative refugia?

If so:

(i) how many and which refugia are implied?

(ii) does the pattern fit the oak or gall wasp model of refugia?

(b) What does the observed structure suggest about the ancient colonisation history of Europe by oak gall wasp communities?

(i) is there support for any refuge being basal to the others?

(ii) if so, is it Anatolia, thus supporting the 'out of Anatolia' hypothesis? (Chapter 1)

2) Is there evidence of host-mediated population subdivision?

As discussed in Chapter 1, it is probable that parasitoids have a hierarchical series of host location behaviours. Variability in the expression of these behaviours might act to drive population subdivision, resulting in the existence of host races. If so, which gall traits are important in driving this process:

(a) host oak section?

(b) gall location on the tree?

(c) gall phenotype, either at the level of key structural traits (such as surface spines, a sticky surface, or an internal airspace) or individual species?

3) By what origin have parasitoids come to attack invading gall wasps in the invaded range?

(a) Pursuit from the south?

(b) Host-shift from native gall wasp communities?

(c) Introduction from galls imported into the UK for industry during the 19th century?

4) Is there evidence of cryptic taxa within a morphologically conserved group of species?

Preliminary evidence from Acs *et al* (2002) suggests that the parasitoid *Eurytoma brunniventris* may be composed of a number of sibling species. Here, I will investigate this possibility further and ask:

(a) Is this finding upheld by mtDNA haplotype data?

(b) If so, do the data suggest subdivision at the level of species, or is variation compatible with that expected for host races?

(c) and are certain gall traits associated with the different groupings?

4.3 Significance & alternative hypotheses

The significance of the above questions and alternative hypotheses for each are now explored. Although ultimately generation of data for nuclear markers was not successful, discussion of predictions for each question are kept broad and are relevant to both nuclear and haplotypes markers in this section.

4.3.1 Question 1. Parasitoid population structure

(a) Refugial structure

Many hypothetical patterns of population subdivision according to putative refugia exist for parasitoid populations in Europe. Here I restrict my interpretation of the data to the following biologically meaningful hypotheses:

H_0 = no subdivision between putative refugial samples

H_1 = the "oak hypothesis" - three refugia: Iberia + Italy + the Balkans, these being the three major refugia for oaks in Europe (see Chapter 1).

H_2 = the "gall wasp hypothesis" - two refugia: Iberia + Italy/the Balkans, as suggested for the gall wasps *Biorhiza pallida* (Rokas *et al* 2001), *Andricus quercustozae* (Rokas *et al* 2003a) and *A. kollari* (Stone *et al* 2001).

The patterns of population subdivison expected under the alternative hypotheses described above are illustrated in Figure 4.1 below. Distinguishing between these possibilities is an important part of characterising the genetic variation found in different parasitoid populations, and will also be necessary in order to address the other questions outlined in section 4.2. In order to discriminate between the hypotheses, samples from each putative refugium need to be collected for each species. Sample selection for this and subsequent questions are discussed in detail in section 4.5. In addition to European samples, samples from Turkey will also be included in order to address question 1(b). A further consideration for Turkey is whether it represents a single refugium, or two refugia separated by the Anatolian diagonal. The Anatolian diagonal (Davis 1965, 1971) is an important floristic divide formed by a series of mountains (Binboga, Munzur, Kargasekmez Mountains, etc.) that cut across Anatolia laterally and separate the European continent from Western Asia.

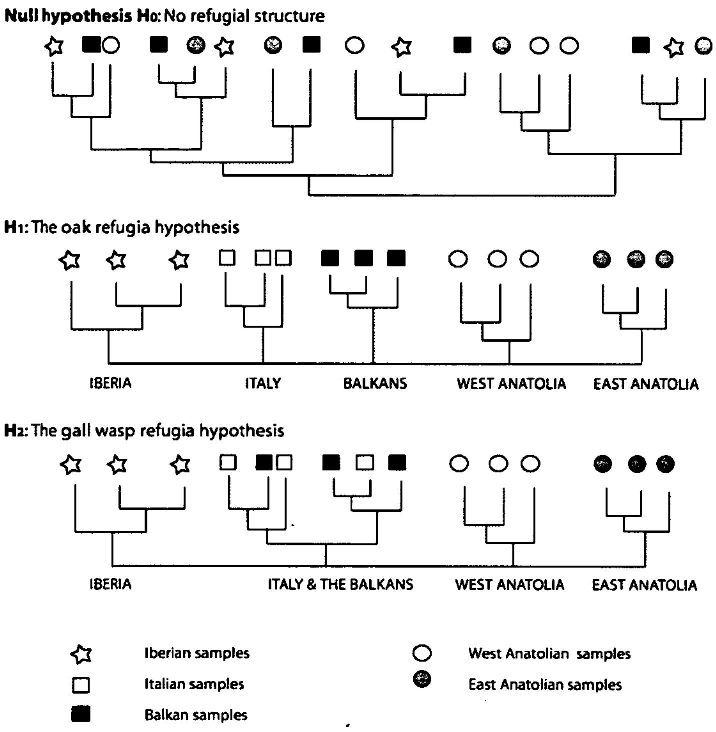


Figure 4.1 Alternative hypotheses for the possible pattern of refugial subdivision in European and Turkish parasitoid populations. Groupings could either represent clades of haplotype sequences or clusters of populations grouped by allele frequency.

(b) Colonisation history

If sufficient structure is discovered within parasitoid intraspecific phylogenies, it may be possible to examine whether a certain refugium is basal and so a possible origin for the species. The following are possible alternative hypotheses:

H_0 = refugial populations are connected by a polytomy within which no specific refuge is identified as basal to the others.

H_1 = Turkish populations basal, with more western populations most derived, *ie* the 'Out of Anatolia' hypothesis.

H_2 = Italian populations basal, with more eastern and western populations derived.

H_3 = Iberian populations basal, with more eastern populations most derived.

Evidence of **H_1**, a similar structure to that observed for the gall wasp hosts, would provide support for the 'Out of Anatolia' hypothesis (Rokas *et al* 2003a), suggesting a shared community origin with entry into Europe from the east (see Figure 4.2 below).

region not only contains high species richness, but may also represent a centre of genetic diversity for many taxa that are traditionally regarded as European. An aim of this Chapter is to assess the extent to which Anatolia represents a possible origin of genetic diversity across trophic levels in oak gall communities, and hence a 'cradle' (*sensu* Chown & Gaston 2000) for the entire trophic unit. If supported, such a pattern should stimulate the application of multispecies phylogeography to other communities.

One of the most intriguing questions in ecology is the degree to which species occur together in ecosystems as a result of specific assembly rules (Diamond 1975, Weiher & Keddy 1999). In order to test predictions regarding the existence of common rules in community assembly, ecologists typically pick what they assume to be replicate (*ie* statistically independent) communities for comparison. If specific patterns are replicated between communities, this is taken as evidence of the validity of the rule in question. However, the problem with this approach is that communities may show similar patterns not due to some underlying rule, but because they are linked by phylogeny (they are phylogenetically related) and therefore not truly independent. Thus, in order to investigate the existence of rules of community assembly through the comparative method, it is first necessary to establish the extent to which the communities under consideration are related by descent. The only means of doing this is through formal examination of community phylogeny.

4.3.2 Question 2. Host-mediated structure

Little is known about host-mediated structure in parasitoid populations. However, as outlined in the introduction to this thesis, oak gall wasp communities are characterised by an abundance of generalist parasitoids that occur in sympatry, and theory predicts partition of these apparent generalists (Thomson & Cunningham 2002). The possible existence of host races is highly important for the way in which we look at community ecology. Interpretation of community function on the basis of foodwebs depends critically on whether generalist parasitoids are true generalists or whether they actually consist of a series of specialist ecotypes. Generalists link many hosts together in a foodweb, and have the potential to mediate apparent competition among shared hosts within a single large foodweb compartment. Alternatively, the existence of host races would imply a set of smaller subcompartments within which only a subset of the hosts could interact.

Host gall recognition by parasitoids may occur by scent, vision, touch or all of these, and a number of possible factors exist by which apparently generalist oak gall wasp parasitoids may be structured into host races. Plant structures release complex volatile chemicals, some of which are known to have an attractant effect on parasitoids (Turlings *et al* 1995, Takabayashi & Dicke 1996). Volatile signature might vary between gall species, or alternatively be dependent upon the host oak or gall location. Different gall species certainly have different appearances, with shape, colour and the presence of surface characteristics varying widely. Given this diversity, parasitoid host races could be structured according to any of these traits. If partitioning of generalist species does occur, we would expect to see correlation between these traits and certain haplotypes or allele frequencies in intraspecific phylogenies (see section on methods 4.5).

In this study, the existence of host races will be examined at three levels: oak section (Cerris or Quercus, see Chapter 1), gall location, and gall morphology. Host oak section is a level that previous results have already suggested might be important (Acs *et al.* 2002), whereas the importance of gall location and morphology have yet to be demonstrated. The following alternative hypotheses are outlined for population subdivison according to key gall traits, conforming to a possible hierarchy of parasitoid search ques (see Chapter 1):

$\mathbf{H_0}$ = no population substructure
$\mathbf{H_1}$ = substructure by oak section
$\mathbf{H_2}$ = substructure by gall location
$\mathbf{H_3}$ = substructure by gall morphology
$\mathbf{H_4}$ = substructure by gall species generation

These hypotheses are not mutually exclusive and it is possible that population substructure could occur according to a hierarchical organisation of these traits. This possibility seems particularly likely if parasitoids respond to a set order of cues during host location.

Determining host races superposed on phylogeographic structure

Our ability to detect refugial structure in contemporary parasitoid populations will depend upon the dispersal ability of the parasitoids. If they are highly efficient dispersers, it is possible that any purely refugial structure once present may have been erased due to gene flow between populations. Host-mediated structure should not be affected by dispersal ability due to the continued action of structuring processes.

In light of these considerations, a series of predictions regarding the possible structure of parasitoid populations is made below:

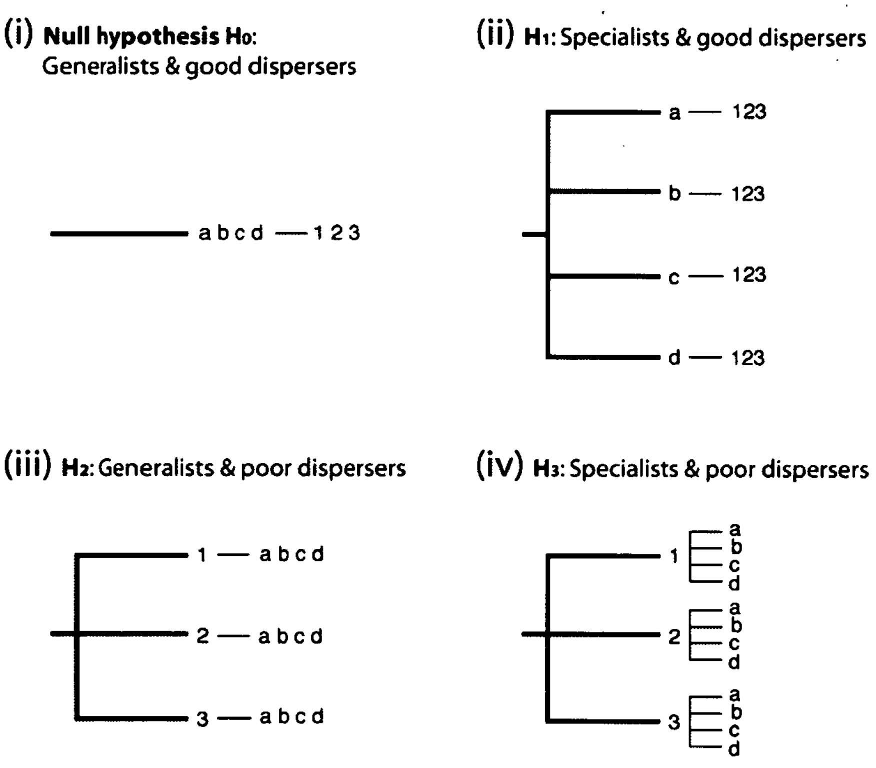


Figure 4.3 Hypotheses of population subdivision with respect to refugia, presence of host races, and dispersal ability. Bold lines indicate monophyletic lineages (not single haplotypes). The letters a,b,c,d represent lineages originating from galls differing according to a certain trait (oak tree host, gall location or morphology). The numbers 1,2,3 represent lineages from different refugia. Numbers or letters written in line indicate refugia or host associations that are not genetically differentiated.

Scenario (i) represents the expected pattern should parasitoids be generalists and good dispersers. In such a case, any refugial structure that once existed has been erased, and no structure is exerted by host gall traits because the parasitoids are true generalists. The resultant pattern will be no clear phylogenetic distribution of either refugial membership or host gall characteristics.

Scenario (ii) depicts the pattern we expect if parasitoids are specialists and good dispersers. In such a case, any refugial structure has been erased and population structure is imposed solely as a result of the important host traits. If this scenario were correct, we would expect to see the existence of genetic subdivision correlating with the number of host traits of importance.

Scenario (iii) corresponds to what we expect if parasitoids are generalists and poor dispersers. Under such a scenario, we expect a number of lineages (*eg* haplotypes) corresponding to the pattern of putative refugia, with no further structure imposed by host-mediated effects.

Scenario (iv) represents the most complex situation, with structure occurring as a result of putative refugia and host effects. This pattern is expected if parasitoids are specialists and poor dispersers.

These scenarios represent simplified versions of what we might expect to see given different combinations of structuring forces. It is quite possible that the resultant pattern in the actual data will not be as clear, and could lie somewhere between different scenarios (see methods section 4.5). However, in such a case the predictions will still act as a useful guide to aid interpretation of the data.

In addition to the above considerations, an important point is that predictions for the assortment of nuclear and haplotype markers differ depending on how host races are structured for the parasitoid in question. In the two most extreme scenarios (no host races, or long-lived host races), there will either be extensive gene flow between parasitoids associated with all hosts at a given site (no host races), or alternatively parasitoids will be divided into long-lived host specific lineages that attack subsets of the host range. Nonexistence of host races would be identified by the absence of significant host-associated genetic structure for any marker, and the existence of long-lived host races by nuclear and mtDNA data that reveal congruent signals of genetic-substructure consistent across different localities. However, it is also possible for substructure to exist due to short-lived local assortative mating among population subsets, or as a result of female host specificity without assortative mating. In the first of these cases, parasitoid subpopulations at a given locality would differ in microsatellite allele frequencies but would not necessarily constitute monophyletic haplotype clades, and patterns of subdivision may differ among localities. In the second case, substructure would be detected for haplotype sequences but not for nuclear markers.

4.3.3 Question 3. Recruitment of parasitoids in the invaded range

There are three possible mechanisms of recruitment of parasitoids attacking invading host-alternating gall wasps in the invaded range: pursuit, host-shift and (for the UK) introduction. These possibilities are not mutually exclusive since parasitoid populations in the invaded range could have arrived via a combination of routes. The significance of each of the hypotheses is discussed in turn below.

Pursuit

The pursuit hypothesis represents the possibility that parasitoid populations followed invading host-alternators from their origin in the south. Invading host-alternating gall wasps in northern continental Europe are thought to have originated from the Balkans and/or Italy. However, parasitoids may have pursued the gall wasps from one or more origins. For example, parasitoids may have escaped from refugia to which their natural hosts are restricted by 'hitchhiking' on invaders, as already demonstrated for *Megastigmus stigamtizans* in Chapter 3. The pursuit hypothesis thus comprises the following possibilities: (a) pursuit from the Balkans, (b) pursuit from Italy (c) pursuit via host shift from Iberia. Here I extend the analysis to another two parasitoid species, and examine whether it applies to the UK in addition to continental Europe.

If natural enemies can pursue introduced hosts into new ranges, it means that current limits to distributions are not restricted by environmental conditions but are a result of host distributions. In other words, their 'physiological envelope' is actually much bigger than is implied by their native range. Inbuilt physiological plasticity is of interest given global warming and the impact of invasions. Predictions regarding the impact of global warming on changes in animal distributions are commonly made by examining the maximum and minimum temperatures at species northern-most range limits, and interpolating where the corresponding isotherm will move to following a change in climate. The new location of the isotherm is then taken to be an estimate of the future range limit of the species (Pearson & Dawson 2003). These predictions are based on climate being the factor that limits species distributions. However, if pursuit of invasive plant hosts following introduction is common, this suggests that it is host distribution and not physiology that limits the spread of animal communities based on plant hosts.

Host shift

For parasitoids that are native to northern Europe and the UK, host-shifts to invaders from native gall wasp communities are a possible alternative form of recruitment. Oak gall parasitoid diversity in Europe decreases from north to south, and the set of parasitoids present in the north are a subset of those found in putative refugia. Therefore, in principle all parasitoid species found in the native range have the potential to pursue invading gall wasps. However, the host-shift hypothesis is only applicable to species native to northern Europe and the UK prior to the invasion of the host-alternators. If host-shifts are shown to have happened, this will demonstrate that parasitoids can recruit to novel host gall phenotypes over very short timescales.

Introduction

The last hypothesis of origin is accidental introduction during mass importation of *Andricus kollari* during the 19th Century. *A. kollari* was first recorded in the UK in 1834 following importation of galls for use in the dye or ink industries in the Exe Valley, Devon (Smith 1854, Parfitt 1856, Lindley & Moore 1870, Walker et al 2002). It appears to have established itself quickly after introduction (Walker et al 2002), and large numbers are documented to have occurred in Devon soon after the time of introduction (Smith 1861). Indeed, one common name by which it was known amongst British entomologists at the time is the 'Devonshire gall' (Walker 1874).

Haplotypes for *A. kollari* sampled in the UK include some not sampled anywhere else in Europe (Stone, unpublished data). This suggests that British *A. kollari* may originate from lineages that escaped from galls from a region outside Europe that has not yet been sampled. If *A. kollari* managed to escape from imported galls, it is highly likely that other inhabitants would also have escaped. For this scenario to be a possible explanation for the presence of parasitoids within invading gall species in the UK, the parasitoids in question would have had to be present in the putative origin of the imported *A. kollari* galls. The precise origin of *A. kollari* imported into the UK is unknown, but is almost certainly located close to the shipping point in the eastern Mediterranean. Sampling of parasitoid communities from Iran has shown that many of the species present in western Europe have distributions at least as far east as Iran. Therefore, it is reasonable to assume that most of the common species of parasitoid will also have been present in the imported galls.

If introduction is indicated, it will demonstrate that whole food web compartments can be transplanted and still function. If it is not supported, this will suggest either that not enough parasitoids were imported, or that they could not persist in the UK. This is of broader significance for the introduction of biocontrol agents from widely different physiological regimes. The presence of parasitoid lineages originating from imported galls will also suggest that current species distributions are of little use in predicting changes in distribution because of global warming.

Discriminating between the various hypotheses

The ability to identify the putative origins of parasitoids in invading gall wasps in the UK depends on the existence of region-specifc haplotypes. Presence in the UK of a haplotype characteristic of a certain European refugium that was traceable back across continental Europe would suggest pursuit from the corresponding refugium. Host shifts from native UK communities could potentially be identified by the presence of UK specific genetic variants in the invading gall wasps, and introduction would be suggested by the existence of a novel hapoltype not seen elsewhere in Europe for populations sampled from the UK. Alternatively, depending on the level of resolution present in the mtDNA haplotype marker, differences may only be visible in haplotype frequency as opposed to region specific haplotypes. This point is particularly pertinent for the host shift hypothesis, given that 'native' British parasitoid populations can only have existed in the UK for up to about 8000-10,000 years at most, the presence of UK specific haplotypes, for example although possible, is unlikely.

4.3.4 Question 4. Evidence of cryptic taxa

Results from the study of Acs *et al* (2002) suggest that the oak gall wasp parasitoid *Eurytoma brunniventris* might consist of multiple sibling species. The *Eurytoma rosae*-group is characterised by morphological uniformity (Claridge & Askew 1960) and no identifiable morphological differentiation has been found within *E. brunniventris*. In their study, Acs *et al.* (2001) examined 10 *E. brunniventris* specimens from five gall species and uncovered three species groups based on RAPD analysis (Random Amplification of Polymorphic DNA). A copy of their results is replicated below in Table 4.1. The species groups identified were split as follows: *Andricus lucidus, A. glutinosus, A. coriarius*: group I. *A. multiplicatus*: group II. *Aphenlonyx cerricola*: group III. These groups appear to be structured primarily according to host oak section, with further division by host gall species for the groups on section Cerris.

Table 4.1 Species groups identified by Acs *et al* (2001)

Host gall sp.	Host oak section	Group
Andricus lucidus	Quercus	I.
Andricus glutinosus	Quercus	
Andricus coriarius	Quercus	
Andricus multiplicatus	Cerris	II.
Aphelonyx cerricola	Cerris	III.

As mentioned above, these are preliminary results based on an extremely small sample size. Here I investigate the possible existence of cryptic species in *E. brunniventris* in more detail, and ask the following questions:

(a) is this finding upheld with respect to mtDNA haplotype data?
(b) if so, do the data suggest subdivision into cryptic species, or is genetic divergence more representative of that expected for host races?
(c) and, are certain gall traits associated with different groupings?

The possibility of uncovering cryptic species in the other parasitoids also exists, particularly for *M. dorsalis*, which in common with *E. brunniventris* has a very large host range. It is important to investigate the possible existence of cryptic species, so that it is clear that similar biological entities are being examined when species are compared.

4.5 Methods

4.5.1 Selection of parasitoids

In the initial stages of research for this thesis, a range of parasitoid species were targeted for study. These included: *Ormyrus nitidulus* and *O. pomaceus* (Ormyridae), *Sycophila biguttata* and *Eurytoma biguttata* (Eurytomidae), and *Torymus auratus*, *T. cyaneus* *Megastigmus stigmatizans* and *M. dorsalis* (Torymidae). These species display a range of biologies and were selected to allow comparisons within and between different genera and families. For example, *Sycophila bigutatta* is one of the few endoparasitoids in oak gall wasp communities and because of the tighter host association endoparasitism implies, it could have been an interesting species in which to investigate the presence of host races. However, in the end it was necessary to focus on a subset of these for various limiting factors (time, money, and the availability of samples from a range of locations and gall hosts), which given the available diversity of oak gall wasp parasitoids only represent a small proportion of available taxonomic and life history diversity. Therefore, the questions posed above are examined with reference to the following three oak gall wasp parasitoids: *Megastigmus stigmatizans, M. dorsalis* and *Eurytoma brunniventris*. These consist of two widespread generalists (*M. dorsalis* and *E. brunniventris*) and one southern specialist (*M. stigmatizans*). *M. dorsalis* and *E. brunniventris* are very common and widespread in oak gall wasp communities, and are therefore important parasitoids for consideration. *M. stigmatizans* is more rare and represents a different life history strategy, thus providing a contrast. The biologies of the species and further reasons for including them in this study are now discussed in more detail:

<u>Megastigmus stigmatizans</u>

M. stigmatizans is a large idiobiont ectoparasitoid (see Chapter 1) of the gall former, recorded from the galls of 14 oak gall wasp species (Askew et al. 2005). The majority of its hosts are large host-alternating members of the genus Andricus, however, it has been reared very occasionally from 2 *Cynips* species. *M. stigmatizans* has only one generation a year and always attacks its hosts during their asexual generation on section Quercus hosts. Since all galls attacked are on section Quercus hosts, the possible impact of oak host on population structure will not be examined for this species. Two of the invading gall wasps that have reached the UK, *Andricus kollari* and *A. quercuscalicis*, are common hosts of *M. stigmatizans*.

Figure 4.4 A female M. stigmatizans specimen

M. stigmatizans native range in Europe is restricted to areas corresponding to putative ice age refugia in the south. However, it has recently expanded its range to northern Europe (see Chapter 3) and has been recorded in the UK from the 19th Century (Walker 1874). *M. stigmatizans* lineages sampled from invading gall wasp hosts in the UK could thus have originated by either pursuit or introduction. Recruitment is not a possibility since *M. stigmatizans* is not a native parasitioid in northern Europe. Introduction is a possible source of *M. stigmatizans* since it was probably introduced to the UK alongside its host *A. kollari* during the 19th Century (see above). All specimens of *M. stigmatizans* sampled from the UK in this study originated from Holcombe down in Devon. In fact, this is the only place known in which *M. stigmatizans* can be found in the UK, and the location is tied in with what is known about the importation history of *A. kollari*. Records of large numbers of *M. stigmatizans* having been reared from galls collected in southern England appear in the latter part of the 19th Century and it is reported to have been common in certain localities at this time (Walker 1874, Fitch 1880). However, it is now extremely rare in the UK, and in a study during which 38,901 *A. quercuscalicis* galls were collected from locations around the country and the inhabitants reared, only a single specimen of *M. stigmatizans* was recovered (Schönrogge et al 1996), which originated from a gall collected on Holcombe Down.

As discussed in Chapter 1, mature gall traits are not relevant to all parasitoids. However, *M. stigmatizans* is one parasitoid where this does apply, since it attacks late in gall development, so final gall traits for those that change (gall morphology & resource size) are relevant to considering host choice. Therefore, it makes particular sense to look at the correlation between haplotypes and gall traits for *M. stigmatizans*. This is not true for *M. dorsalis* or *E. brunniventris*.

Figure 4.5 A female *M. dorsalis* specimen

<u>*Megastigmus dorsalis*</u>

M. dorsalis is closely related to *M. stigmatizans*, but is smaller and has a much wider host range, having been reared from 42 different species of oak gall, including many non host-alternating species (Schönrogge & Askew, in press). *M. dorsalis* is also an idiobiont and parasitises both gall inducers and inquilines. It has two generations a year, and attacks both spring and summer generation galls. Because it is native to northern Europe, individuals reared from invading gall inducers found in this range could be traced to either: (a) lineages that pursued the gall inducers, (b) lineages that were recruited from native populations, or (c) lineages that are descended from introduced galls, or a combination of these. *M. dorsalis* thus allows testing of all three hypothesis of recruitment to invading gall wasp hosts. It was also specifically selected to allow testing of the host race hypothesis between congeneric specialists and generalists, with the expectation of less subdivision in *M. stigmatizans*.

Figure 4.5 A female *E. brunniventris* specimen

<u>*Eurytoma brunniventris*</u>

E. brunniventris is of special interest in this study for two reasons. Firstly it is even more generalist than *M. dorsalis*, with 77 recorded host species of oak gall wasp (Schönrogge & Askew, in press) and also has a native range including northern Europe. Therefore, it was selected to allow testing of all three alternative hypotheses for enemy recruitment in northern Europe as a replicate with *M. dorsalis*. Secondly, it was selected because it is part of a possible cryptic species set which molecular tools could help to dissect (see also Chapter 6). *E. brunniventris* differs from the other parasitoids considered here in that it is able to mature by feeding solely on plant tissue. It can also act as a facultative hyperparasitoid and cannibal (Askew 1999). Furthermore, it is reputed to feed mainly on inquilines and seldomly on cynipid gall inducers (Claridge & Askew 1960). This is of relevance, since if *E. brunniventris* only attacks inquilines close to the surface of a gall, it might be less sensitive to host gall traits. Therefore, in contrast to *M. stigmatizans* the relevance of mature gall phenotypes is something that needs to be considered for *E. brunniventris* and *M. dorsalis*.

4.5.2 Analysis

Lab techniques and phylogenetic analyses were carried out as outlined in Chapter 2 of this thesis. Outgroups were selected from available material to be closely related to the ingroup. For *M. stigmatizans* these were *M. dorsalis, M. synophri* and *Torymus auratus*. For *M. dorsalis* outgroups were: *M. dumicola, M. synophri* and *Torymus auratus*. For *E. brunniventris* outgroups were: *M. dorsalis, Torymus auratus* and *Eurytoma pistacina*.

As described in Chapter 2, the hypotheses discussed above were initially to have been tested using a combination of nuclear and mtDNA markers. However, for reasons discussed in Chapter 2, only haplotype data were available and specific points refer to this data type.

The hypotheses discussed above predict ideal distributions of haplotypes across characters, and very strong patterns may be so obvious that they do not need formal statistical testing. Where character distributions are less obvious, hypothesis testing can be approached in two ways: (a) we can ask whether haplotype partitioning across characters is perfect by investigating whether a single haplotype is associated with multiple character states (*eg* two refugia or two gall locations etc). If this is true it means either that perfect population subdivision does not exist for this marker or that sorting of ancestral polymorphism conceals such structure. (b) We can ask if the correlation between genetic distance and specimen character states is significantly non-random. I became aware of the existence of this group of methods too late in my thesis to apply them, but I discuss them here because of their importance. Two similar approaches to answering this question are Mantel tests and analysis of similarities (ANOSIM), both of which are forms of matrix correlation test.

The Mantel test (Mantel 1967, Mantel & Valand 1970) is used to test the correlation between two distance matrices and is commonly implemented in the population genetics computer program Arlequin (Schneider *et al* 2000). As an example, to test if genetic distance were correlated with geographical distance, the matrices required would be genetic distance between haplotypes and geographical distance between the sample locations of haplotypes. A correlation coefficient is calculated for these matrices, the columns and rows within one of the matrices are permutated, and the correlation coefficient is recalculated. Significance is measured as the number of times the permutated coefficient exceeds the original correlation coefficient. The logical basis of the test is that if the null hypothesis (no significant correlation) is correct, then permutation of the matrix should be equally likely to result in a larger or smaller correlation coefficient.

ANOSIM (Clark 1993) uses rank similarities among replicates contained within a certain grouping and rank similarities from all pairs of replicates between different groupings to calculate a test statistic. This test statistic is then compared to a recomputed statistic calculated for a permutated data set. Under the null hypothesis (no significant difference between groups), permutation is expected to result in little effect (on average) to the value of the test statistic. In this way the test indicates the degree to which points within the same groupings have closer biological similarity to each other than average levels of similarity occurring across the wider dataset. ANOSIM is usually implemented using the computer package PRIMER (Clarke & Gorley, 2001).

The Mantel test and ANOSIM work in a similar way, but test slightly different things. In a Mantel test the correlation between two matrices is measured. An ANOSIM tests how well an imposed grouping explains the variation in the real data, using the average distance between data points within each imposed class, and the average distances between classes. An important limitation associated with the use of the Mantel test is that it can only cope with distance data. Binary gall traits such as presence or absence of surface spines are compatible with representation as distance, however, traits such as gall location are not amenable to such an approach. Therefore, the ANOSIM would be the test of preference for the data applicable in this study.

An alternative approach to conducting character-by-character analyses, would be to conduct a multivariate analysis. This could be applied in order to examine if groups of gall traits have a combined effect that it is not possible to pick up when examining gall traits individually. A commonly implemented example being Principal Components Analysis (PCA), often conducted using the program Statistical Package for Social Sciences (SPSS)(SPSS inc. 1999). PCA looks for a pattern in multivariate data by attempting to reduce the dimensionality inherent within the data, as summarised by a series of principal components.

Character by character analyses are easy to interpret, but risk a type II error since they may miss complex patterns. However, even when detected, these patterns can be very hard to interpret. Furthermore, such analyses might prove unnecessary if more simple analyses reveal strong patterns.

4.5.3 Sampling design

Specimen selection requires sampling of host galls (a) distributed across appropriate putative refugia and (b) showing diversity in the gall traits whose impact on parasitoid population structure I wish to assess. In this study hosts were selected from a catalogue by Dr K Schönrogge & Dr RR Askew (Schönrogge & Askew, in press), and scored by the following gall trait categories:

(a) putative refugia: Iberia, Italy, the Balkans, eastern Turkey, Western Turkey
(b) oak section: sections Quercus, Cerris
(c) gall location: bud, acorn, leaf, catkin, shoot, root
(d) gall morphology: presence or absence of surface spines
(e) gall morphology: presence of absence of surface glue
(f) gall morphology: presence or absence of an internal airspace

The following discussion illustrates sampling design for *M. stigmatizans*. The intention is to show *a priori* host gall selection for the hypotheses to be tested and the inevitable compromise associated with sampling from the real world. A similar approach, though not summarised, was applied for the other two parasitoid species.

Host gall species are identified by the following abbreviations:

ac = *Andricus coriarius*
acm = *A. caputmedusae*
an = *A. coronatus*
ad = *A. dentimitratus*
ag = *A. glossulariae*
ah = *A. hungaricus*
ai = *A. infectorius*
ak = *A. kollari*
al = *A. lucidus*
ap = *A. pictus*
aqc = *A. quercuscalicis*
aqt = *A. quercustozae*
clv = *Cynips longiventris*
cqf = *C. quercusfolii*

The following tables show known recorded hosts by refuge and character states. The aim was to balance sampling across refugia, and within each refugium to sample adequate numbers of parasitoids from hosts showing the full diversity of gall location and morphology character states. Numbers in brackets indicate the actual numbers of samples obtained from each host gall and treatment (refugium, gall location, etc). *M. stigmatizans* only attacks hosts on section Quercus oaks, therefore, oak section is not considered here.

Plate. 4.1 Recorded hosts of *M. stigmatizans*. All are asexual generation galls. From left to right, by rows down the figure, the galls are *Andricus caputmedusae*, *A. coriarius*, *A. coronatus*, *A. dentimitratus*, *A. hungaricus*, *A. kollari*, *A. lucidus*, *A. pictus*, *A. quercuscalicis*, *A. quercustozae*, *Cynips longiventris* and *C. quercusfolii*.

Table 4.2 Known hosts for *M. stigmatizans*, separated by putative refugia. Numbers in brackets show actual numbers of specimens sampled

Iran (1)	Iberia (18)	Italy (13)	the Balkans (10)	Turkey (8)	N. Europe/UK (33)
aqt (1)	ac	acm (2)	acm (5)	acm (2)	ac
	ad	ak (3)	ac	ad (2)	acm
	ak (14)	aqc	an	ak (2)	ad
	ap	aqt (5)	ak (3)	aqt (2)	ak (27)
	aqt (4)	ai (1)	ah		aqc (6)
		ag (1)	al		aqt
		ad (1)	aqc (2)		
			aqt		
			clv		
			cqf		

Table 4.3 Known hosts for *M. stigmatizans*, separated by gall location. Numbers in brackets show actual numbers of specimens sampled

Bud (63)	Acorn (20)	Leaf (0)
ac	acm (9)	Clv
an	ad (3)	Cqf
ak (49)	al	
ah	ap	
al	aqc (8)	
aqt (12)		
ai (1)		
ag (1)		

Table 4.4 Known hosts for *M. stigmatizans*, separated by presence or absence of surface spines. Numbers in brackets show actual numbers of specimens sampled.

Spines

With (10)	Without (73)
acm (9)	an
ac	ad (3)
al	ak (49)
ag (1)	ah
	ap
	aqc (8)
	aqt (12)
	ai (1)
	clv
	cqf

Table 4.5 Known hosts for *M. stigmatizans*, separated by presence or absence of a sticky surface. Numbers in brackets show actual numbers of specimens sampled.

Sticky surface

With (33)	Without (50)
an	ac
ad (3)	ak (49)
acm (9)	ai (1)
al	clv
ap	cqf
aqc (8)	
aqt (12)	
ag (1)	

Table 4.6 Known hosts for *M. stigmatizans*, separated by presence or absence of an internal airspace. Numbers in brackets show actual numbers of specimens sampled.

Internal airspace	
With (23)	Without (60)
an	acm (9)
ad (3)	ac
ah	ak (49)
ap	al
aqc (8)	ai (1)
aqt (12)	ag (1)
	clv
	cqf

Table 4.7 Diversity of host gall species sampled for *M. stigmatizans*. Numbers in brackets show actual numbers of specimens sampled.

Species
Gall species (83)
ac (0)
acm (9)
an (0)
ad (3)
ag (1)
ah (0)
ai (1)
ak (49)
al (0)
ap (0)
aqc (8)
aqt (12)
clv (0)
cqf (0)

Consequences of incomplete sampling

Samples from as many hosts and refugia as possible were reared from my own and other collections. Full details of all sampling locations and hosts are provided in Table 4.10 below. However, Tables 4.2-4.7 show that whilst sampling across refugia is relatively balanced, sampling across hosts and host characters within refugia is weak. This means that if refugial structure is strong, detection of host races within refugia (see hypothesis 4, Figure 4.1) may be difficult or impossible. For gall location the only contrast available from sampled hosts is between bud and acorn. *M. stigmatizans* is extremely rare in leaf galls (*Cynips quercusfolii*, *C. longiventris*) and none were collected. The bud sampling is heavily influenced by samples from one host (*A. kollari* 49/62) so gall location may be confounded with other traits of this specific host gall. If refugial structure exists then no gall location analysis is possible for Iberia since all host galls are on buds. For gall morphology: (i) Spines - a contrast is possible but sampling is heavily biased towards parasitoids reared from host galls without spines (73/83) (ii) sticky surface - sampling is relatively balanced between galls with and without a sticky surface, but 49/50 parasitoids reared from non-sticky galls are from a single host (*A. kollari*). (c) Internal air space - sampling allows contrast between these two states, but sampling is heavily biased in galls without an internal air space by samples reared from *A. kollari.* Having noted these limitations in available data, improvements in taxon sampling remain an important issue for this parasitoid.

Summaries of host gall traits sampled for *M. dorsalis* and *E. brunniventris* are now provided, and full details of all sampling locations are given in Tables 4.11 and 4.12 respectively.

Table 4.8 Summary of gall traits sampled for *M. dorsalis.* Numbers in brackets show actual numbers of specimens sampled.

Host gall species		Oak section	Location on tree	Spines	Sticky surface	Internal airspace
Andricus coriarius	(16)	Quercus	bud	-	-	-
Andricus grossulariae	(6)	Quercus	multiple	+	+	-
Andricus hungaricus	(1)	Quercus	bud	+	-	+
Andricus kollari	(13)	Quercus	bud	-	-	-
Andricus lignicola	(2)	Quercus	bud	-	-	-
Andricus lucidus	(4)	Quercus	bud	+	+	-
Synophrus politus	(2)	Cerris	multiple	-	-	-

Table 4.9 Summary of gall traits sampled for *E. brunniventris.* Numbers in brackets show actual numbers of specimens sampled.

Host gall species		Location on tree	Oak section	Spines	Sticky surface	Internal airspace
Andricus conglomeratus	(3)	bud	Quercus	-	-	-
Andricus coriarius	(2)	bud	Quercus	+	-	-
Andricus coronatus	(4)	bud	Quercus	-	+	+
Andricus infectorius	(5)	bud	Quercus	-	-	-
Andricus kollari	(8)	bud	Quercus	-	-	-
Andricus polycerus	(1)	bud	Quercus	-	-	-
Andricus quercustozae	(7)	bud	Quercus	-	+	+
Aphelonyx persica	(2)	shoot	Cerris	-	-	-
Cynips quercus	(7)	leaf	Quercus	-	-	-
Cynips quercusfolii	(2)	leaf	Quercus	-	-	-
Trigonaspis synaspis	(2)	leaf	Quercus	-	-	+

As for *M. stigmatizans,* for consideration of the above questions it was necessary to include specimens for each species from a wide range of refugial and northern European locations. It was also necessary to include as many specimens as possible from each putative refugium, balanced across available hosts to allow analysis of the impact of the host traits of interest. Constraints on specimen selection meant that it was not always possible to include all the specimens preferable in each case. Precise specimen locations for the parasitoids used in this study are provided in Tables 4.10, 4.11 and 4.12 below.

Table 4.10 Sample site details for *Megastigmus stigmatizans* specimens. Haplotype refers to the haplotype number designated below (see table 4.13).

Megastigmus stigmatizans

Host gall Sp.	Country	Site name	Latitude	Longitude	Haplotype
Andricus caputmedusae	Hungary	Godollo	47°36'0N	19°19'48E	H1
Andricus caputmedusae	Hungary	Balaton	46° 57' 0N	17° 52' 60E	H2
Andricus caputmedusae	Hungary	Eger	47°53'60N	20°22'60E	H6
Andricus caputmedusae	Italy	Gargnano	45° 40' 60N	10° 40' 0E	H9
Andricus caputmedusae	Italy	Gargnano	45° 40' 60N	10° 40' 0E	H10
Andricus caputmedusae	Turkey	Egirdir	41°1'0N	35°54'0E	H18
Andricus caputmedusae	Turkey	Suluova	40° 50' 11N	35° 38' 34E	H29
Andricus caputmedusae	Austria	Vienna	41°1'0N	35°54'0E	H30
Andricus caputmedusae	Hungary	Eger	47°53'60N	20°22'60E	H32
Andricus caputmedusae	Turkey	Tokat	39° 12' 14N	29° 8' 21E	H33
Andricus caputmedusae	Turkey	Suluova	40° 50' 11N	35° 38' 34E	H34
Andricus dentimitratus	Turkey	Beybesli	41°1'0N	35°54'0E	H2
Andricus dentimitratus	Italy	Gargnano	45° 40' 60N	10° 40' 0E	H15
Andricus dentimitratus	Turkey	Beybesli	41°1'0N	35°54'0E	H29
Andricus grossulariae	Italy	Greve	43°34'48N	11°19'12E	H16
Andricus infectorius	Italy	Cilento	39° 58' 60N	15° 43' 0E	H14
Andricus kollari	France	Vannes	47°40'0N	2°45'0W	H2
Andricus kollari	Hungary	Godollo	47°36'0N	19°19'48E	H2
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H2
Andricus kollari	France	St Malo	48°38'60N	2°1'0W	H2
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H2
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H2
Andricus kollari	France	Nantes	47°13'48N	1°34'48W	H2
Andricus kollari	France	Nantes	47°13'48N	1°34'48W	H2
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H2
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H2
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H2
Andricus kollari	France	Perigueux	45°10'60N	0°43'0E	H2
Andricus kollari	Turkey	Refahiye	39°54'0N	38°45'0E	H2
Andricus kollari	Hungary	Sopron	47°40'12N	16°34'48E	H3
Andricus kollari	Hungary	Godollo	47°36'0N	19°19'48E	H4
Andricus kollari	Italy	Pruno	43° 58' 60N	10° 13' 0E	H11
Andricus kollari	Italy	Chiusi	43°1'0N	11°57'0E	H12
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H15
Andricus kollari	France	Vannes	47°40'0N	2°45'0W	H21
Andricus kollari	France	Rennes	48°6'0N	1°40'12W	H21
Andricus kollari	France	Pau	43°17'60N	0°22'0W	H21
Andricus kollari	France	Pau	43°17'60N	0°22'0W	H21
Andricus kollari	France	Pau	43°17'60N	0°22'0W	H21
Andricus kollari	France	Rennes	48°6'0N	1°40'12W	H21
Andricus kollari	Spain	Olot	42°10'60N	2°28'60E	H21
Andricus kollari	France	Ceret	42°28'60N	2°45'0E	H21
Andricus kollari	Spain	El Escorial	40°34'60N	4°9'0W	H21
Andricus kollari	Spain	El Escorial	40°34'60N	4°9'0W	H21
Andricus kollari	Portugal	Lagarelhos	41°2'60N	8°4'0W	H21
Andricus kollari	France	Bayonne	43°28'60N	1°28'60W	H21
Andricus kollari	France	St Malo	48°38'60N	2°1'0W	H21
Andricus kollari	France	St Malo	48°38'60N	2°1'0W	H21
Andricus kollari	France	Nantes	47°13'48N	1°34'48W	H21
Andricus kollari	France	Nantes	47°13'48N	1°34'48W	H21
Andricus kollari	France	Rennes	48°6'0N	1°40'12W	H21
Andricus kollari	France	Rennes	48°6'0N	1°40'12W	H21
Andricus kollari	Spain	Sant Celoni	41°41'07N	2°29'57E	H21
Andricus kollari	France	Rennes	48°6'0N	1°40'12W	H21
Andricus kollari	France	Rennes	48°6'0N	1°40'12W	H21
Andricus kollari	France	Rennes	48°6'0N	1°40'12W	H22
Andricus kollari	France	Amou	43°36'0N	0°45'0W	H23
Andricus kollari	Portugal	Cabeca Gorda	37°55'0N	7°47'60W	H24
Andricus kollari	France	Ceret	42°28'60N	2°45'0E	H25
Andricus kollari	Spain	El Escorial	40°34'60N	4°9'0W	H26
Andricus kollari	Turkey	Beybesli	41°1'0N	35°54'0E	H27
Andricus kollari	Italy	Pruno	43° 58' 60N	10° 13' 0E	H28
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H31
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H31
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H31
Andricus quercuscalicis	Germany	Ludwigsburg	50°31'60N	12°25'0E	H2
Andricus quercuscalicis	Hungary	Sopron	47°40'12N	16°34'48E	H5
Andricus quercuscalicis	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H17
Andricus quercuscalicis	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H31
Andricus quercuscalicis	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H31
Andricus quercuscalicis	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H31
Andricus quercuscalicis	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H31
Andricus quercustozae	Italy	Gildone	41° 31'0N	14° 43'60E	H7
Andricus quercustozae	Italy	Greve	43°34'48N	11°19'12E	H8
Andricus quercustozae	Italy	Massa Marittima	43° 2' 60N	10° 52' 60E	H8
Andricus quercustozae	Italy	Chiusi	43°1'0N	11°57'0E	H13
Andricus quercustozae	Italy	Greve	43°34'48N	11°19'12E	H15
Andricus quercustozae	Turkey	Küllüce	38° 12' 0N	34° 36' 0E	H18
Andricus quercustozae	Turkey	Küllüce	38° 12' 0N	34° 36' 0E	H19
Andricus quercustozae	Morocco	Azrou	34° 31' 60N	4° 20' 60W	H20
Andricus quercustozae	Portugal	Monchique	37° 19' 0N	8° 33' 0W	H20
Andricus quercustozae	Morocco	Azrou	34° 31' 60N	4° 20' 60W	H20
Andricus quercustozae	Spain	Prado del rey	36° 47' 60N	5° 32' 60W	H21
Andricus quercustozae	Iran	Kordestan	/	/	H29

Table 4.11 Sample site details for *Megastigmus dorsalis* specimens. Haplotype refers to the haplotype number designated below (see table 4.15).

Megastigmus dorsalis

Gall species	Country	Site name	Latitude	Longitude	Haplotype
Andricus coriarius	Greece	Pisodherion	40°46'60N	21°15'0E	H1
Andricus coriarius	Greece	Pisodherion	40°46'60N	21°15'0E	H1
Andricus coriarius	Greece	Pisodherion	40°46'60N	21°15'0E	H1
Andricus coriarius	Spain	Soto del Real	40° 45' 0N	3° 46' 48W	H1
Andricus coriarius	Turkey	Madenli	40° 53' 60N	33° 15' 0E	H2
Andricus coriarius	Italy	San Venanzo	42° 52' 0N	12° 16' 0E	H2
Andricus coriarius	Italy	San Venanzo	42° 52' 0N	12° 16' 0E	H2
Andricus coriarius	Spain	Soto del Real	40° 45' 0N	3° 46' 48W	H2
Andricus coriarius	Hungary	Tatabanya	47° 34' 0N	18° 25' 0E	H3
Andricus coriarius	Hungary	Tatabanya	47° 34' 0N	18° 25' 0E	H3
Andricus coriarius	Turkey	Egirdir	37° 52' 12N	30° 51' 0E	H5
Andricus coriarius	Italy	Massa Marittima	43° 2' 60N	10° 52' 60E	H8
Andricus coriarius	Italy	Massa Marittima	43° 2' 60N	10° 52' 60E	H8
Andricus coriarius	Italy	Lame	45° 40' 60N	13° 4' 0E	H9
Andricus coriarius	Italy	San Venanzo	42° 52' 0N	12° 16' 0E	H13
Andricus coriarius	Spain	Soto del Real	40° 45' 0N	3° 46' 48W	H13
Andricus grossulariae	Hungary	Godollo	47°36'0N	19°19'48E	H1
Andricus grossulariae	Hungary	Godollo	47°36'0N	19°19'48E	H1
Andricus grossulariae	Italy	Poppi	43° 43' 0N	11° 46' 0E	H1
Andricus grossulariae	Turkey	Niksar	40° 35' 24N	36° 56' 60E	H11
Andricus grossulariae	Turkey	Egirdir	37° 52' 12N	30° 51' 0E	H12
Andricus grossulariae	Iran	Pyran Shar	36°41'43N	45°8'21E	H15
Andricus hungaricus	Hungary	Godollo	47°36'0N	19°19'48E	H1
Andricus kollari	France	Ceret	42°28'60N	2°45'0E	H1
Andricus kollari	France	Ceret	42°28'60N	2°45'0E	H1
Andricus kollari	France	Rennes	48°6'0N	1°40'12W	H1
Andricus kollari	France	Rennes	48°6'0N	1°40'12W	H1
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H1
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H1
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H1
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H1
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H1
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H1
Andricus kollari	Britain	Teignmouth	50° 32' 60N	3° 28' 0W	H1
Andricus kollari	France	Rennes	48°6'0N	1°40'12W	H4
Andricus kollari	UK	Teignmouth	50° 32' 60N	3° 28' 0W	H7
Andricus lignicola	Germany	Ludwigsburg	50°31'60N	12°25'0E	H1
Andricus lignicola	Hungary	Godollo	47°36'0N	19°19'48E	H1
Andricus lucidus	Italy	Castel del Alpi	44°12'37N	11°15'16E	H6
Andricus lucidus	Turkey	Niksar	40° 35' 24N	36° 56' 60E	H10
Andricus lucidus	Iran	Shena	33°32'76N	48°04'34E	H14
Andricus lucidus	Iran	Shena	33°32'76N	48°04'34E	H16
Synophrus politus	Greece	Pisodherion	40°46'60N	21°15'0E	H1
Synophrus politus	Greece	Pisodherion	40°46'60N	21°15'0E	H1

Table 4.12 Sample site details for *Eurytoma brunniventris* specimens. Haplotype refers to the haplotype number designated below (see table 4.17).

Eurytoma brunniventris

Host gall sp.	Country	Site name	Latitude	Longitude	Haplotype
Andricus conglomeratus	Greece	Chalkidiki	40° 22' 28N	23° 26' 18E	H9
Andricus conglomeratus	Greece	Chalkidiki	40° 22' 28N	23° 26' 18E	H12
Andricus conglomeratus	Greece	Chalkidiki	40° 22' 28N	23° 26' 18E	H26
Andricus coriarius	Hungary	Matrafured	47° 49' 60N	19° 58' 0E	H5
Andricus coriarius	Hungary	Godollo	47°36'0N	19°19'48E	H7
Andricus coronatus	Spain	Lerida	41° 37' 0N	0° 37' 0E	H1
Andricus coronatus	Spain	Lerida	41° 37' 0N	0° 37' 0E	H4
Andricus coronatus	Turkey	N. Anatolia	/	/	H11
Andricus coronatus	Turkey	/	/	/	H17
Andricus infectorius	Italy	Gargnano	45° 40' 60N	10° 40' 0E	H2
Andricus infectorius	Italy	Gargnano	45° 40' 60N	10° 40' 0E	H2
Andricus infectorius	Italy	Gargnano	45° 40' 60N	10° 40' 0E	H2
Andricus infectorius	Italy	Cupi	42° 58' 60N	13° 5' 60E	H2
Andricus infectorius	Italy	Cupi	42° 58' 60N	13° 5' 60E	H2
Andricus kollari	Spain	Montseny	41° 46' 0N	2° 23' 60E	H2
Andricus kollari	Spain	Montseny	41° 46' 0N	2° 23' 60E	H2
Andricus kollari	Italy	Gargnano	45° 40' 60N	10° 40' 0E	H2
Andricus kollari	Spain	Montseny	41° 46' 0N	2° 23' 60E	H3
Andricus kollari	France	Nantes	47°13'48N	1°34'48W	H3
Andricus kollari	France	Nantes	47°13'48N	1°34'48W	H3
Andricus kollari	France	Nantes	47°13'48N	1°34'48W	H3
Andricus kollari	Hungary	Godollo	47°36'0N	19°19'48E	H10
Andricus polycerus	Greece	Komnina	37° 53' 60N	23° 45' 0E	H8
Andricus quercustozae	Italy	Campora	40° 19' 0N	15° 16' 60E	H19
Andricus quercustozae	Portugal	Vila de Ala	41° 36' 0N	7° 1' 0W	H22
Andricus quercustozae	Portugal	Vila de Ala	41° 36' 0N	7° 1' 0W	H23
Andricus quercustozae	Portugal	Vila de Ala	41° 36' 0N	7° 1' 0W	H23
Andricus quercustozae	Portugal	Vila de Ala	41° 36' 0N	7° 1' 0W	H23
Andricus quercustozae	Portugal	Vila de Ala	41° 36' 0N	7° 1' 0W	H24
Andricus quercustozae	Portugal	Caneiro	40° 13' 60N	8° 19' 0W	H28
Aphelonyx persica	Iran	Tatineh	33°48'60N	47°56'51N	H25
Aphelonyx persica	Iran	Tatineh	33°48'60N	47°56'51N	H27
Cynips quercus	Hungary	Szentkut	47° 58' 60N	19° 46' 60E	H5
Cynips quercus	Hungary	Sopron	47° 40' 60N	16° 36' 0E	H6
Cynips quercus	Turkey	Suluova	40° 50' 11N	35° 38' 34E	H13
Cynips quercus	Turkey	Tokat	39° 12' 14N	29° 8' 21E	H14
Cynips quercus	Turkey	Tokat	39° 12' 14N	29° 8' 21E	H15
Cynips quercus	Turkey	Cekerek	40° 4' 23N	35° 29' 41E	H16
Cynips quercus	Turkey	Cekerek	40° 4' 23N	35° 29' 41E	H16
Cynips quercusfolii	Hungary	Matrafured	47° 49' 60N	19° 58' 0E	H5
Cynips quercusfolii	Hungary	Matrafured	47° 49' 60N	19° 58' 0E	H18
Trigonaspis synaspis	Spain	Valgallego	40° 49' 60N	3° 31' 60W	H20
Trigonaspis synaspis	Spain	Valgallego	40° 49' 60N	3° 31' 60W	H21

In addition, no specimens of any species were obtained for east of the Anatolian divide in Turkey. Therefore, the existence of multiple refugia within Turkey cannot be investigated here. However, several specimens were obtained from western Iran, which lies east of the Anatolian divide and possibly belongs to the same putative refugium.

Figure 4.6 Location of samples for specimens of all species obtained from Turkey

4.6 Results and discussion by species

4.6.1 *Megastigmus stigmatizans*

83 individuals revealed a total of 34 haplotypes for the 643 bp cyt*b* fragment. Of the 643 nucleotide sites examined, 35 were variable and 19 were parsimony informative. Corrected pairwise distances (Table 4.13) show that the distance between haplotypes is very low (mean=0.64%, range=0-1.98%).

Table 4.13 Haplotypes recovered for *M. stigmatizans*. Numbers at the top of the table indicate nucleotide position from the 3' end of the fragment to a variable site. The column labelled *f* indicates the frequency with which each haplotype was sampled. Letters in the column labelled 'geographical location' refer to the country from which the specimen was sampled: T=Turkey, P=Iran, H=Hungary, A=Austria, F=France (excluding SW France), D=Germany, U=Britain, I=Italy, B=Iberia.

Megastigmus stigmatizans Cytochrome b haplotype table

Haplotype number				1	1	1	1	2	2	2	2	2	2	3	3	3	3	4	4	4	4	4	4	4	4	4	5	5	5	5	5	5	5	5	6	*f*	Geographical location
		8	8	6	7	7	9	1	2	5	6	8	9	1	2	5	7	0	0	6	6	6	6	7	9	9	2	3	3	5	6	8	9	9	3		
	4	5	6	7	0	6	6	4	9	0	5	0	6	9	2	2	2	1	6	0	3	6	9	2	8	9	0	2	4	4	2	0	2	5	7		
	C	T	T	A	G	A	A	A	A	A	T	T	G	A	A	C	T	G	C	A	A	C	T	C	C	A	T	T	G	T	T	T	T	T	A		
Haplotype 1	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	H
Haplotype 2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	16	T, H, F, D, U
Haplotype 3	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	H
Haplotype 4	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	H
Haplotype 5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	1	H
Haplotype 6	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	H
Haplotype 7	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.	.	.	C	.	.	1	I
Haplotype 8	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	2	I
Haplotype 9	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	.	1	I
Haplotype 10	T	.	.	G	.	.	.	.	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	1	I
Haplotype 11	T	.	.	.	.	.	.	G	.	.	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	1	I
Haplotype 12	T	.	G	.	.	.	.	.	G	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	1	I
Haplotype 13	T	.	.	.	.	.	.	G	.	G	A	.	.	G	.	.	A	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	1	I
Haplotype 14	T	.	.	.	.	.	.	G	G	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	1	I
Haplotype 15	T	.	.	.	.	.	.	G	G	G	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	3	I, U
Haplotype 16	T	.	.	.	.	.	.	G	G	G	A	.	.	G	.	.	.	.	.	G	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	1	I
Haplotype 17	T	.	.	.	.	.	.	G	G	G	A	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	U
Haplotype 18	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	T
Haplotype 19	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	1	T
Haplotype 20	.	.	.	.	.	.	.	.	.	.	A	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	3	B
Haplotype 21	.	.	.	.	.	.	.	.	.	.	A	.	.	.	G	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	.	.	21	B, F
Haplotype 22	.	.	.	.	.	.	.	.	.	.	A	.	.	.	G	.	.	A	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	.	.	1	B, F
Haplotype 23	.	.	.	.	.	.	.	.	.	.	A	.	.	.	G	T	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	.	.	1	B
Haplotype 24	.	.	.	.	.	.	.	.	.	G	A	.	.	.	G	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	.	.	1	B
Haplotype 25	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	.	.	1	B
Haplotype 26	.	.	.	.	.	.	G	.	.	.	A	.	.	.	G	.	.	.	.	.	.	.	.	.	T	.	C	.	.	.	.	.	.	.	.	1	B
Haplotype 27	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	T	T	.	.	.	.	.	.	.	.	.	.	.	.	.	1	T
Haplotype 28	T	.	.	G	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	C	.	.	G	.	.	.	.	.	.	.	.	.	1	I
Haplotype 29	.	.	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	3	T, P
Haplotype 30	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	1	A
Haplotype 31	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	7	U
Haplotype 32	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	1	H
Haplotype 33	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	T
Haplotype 34	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	1	T

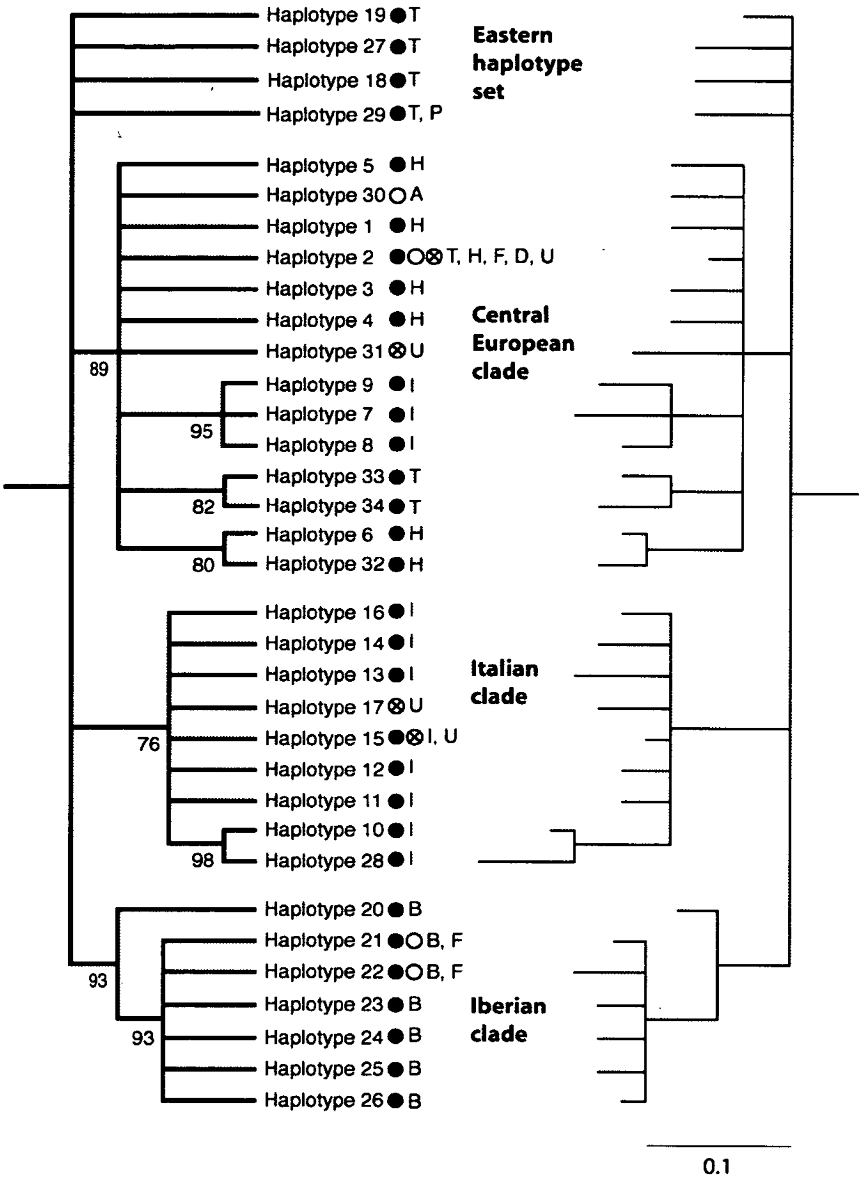


Figure 4.7 *Megastigmus stigmatizans* haplotype tree for a 643 bp fragment of the cyt*b* gene (left-hand side), and a phylogramatic respresentation of the same tree (right-hand side) in which branch lengths indicate the divergence between sequences. Filled circles indicate specimens sampled from a putative refugial area. Clear circles indicate specimens sampled from the invaded range. Circles with a cross indicate specimens sampled from Britain. Letters at the end of branches refer to the following locations: T=Turkey, P=Iran, H=Hungary, A=Austria, F=France (excluding SW France), D=Germany, U=Britain, I=Italy, B=Iberia. The term 'Central European clade' is used because the haplotypes present in this clade are represented predominantly by samples from central Europe, however, it is important to note the presence of Turkish sequences in this clade, as well as sequences derived from samples within the invaded range. Numbers adjacent to clades indicate Bayesian posterior probabilities. The scale bar indicates percentage sequence divergence for the phylogram.

Table 4.14 Corrected distances (in percent) with haplotype and specimen frequency per clade for the *M. stigmatizans* haplotype data

	Whole clade	Eastern set	Central European clade	Italian clade	Iberian clade
Range	0-1.98	0-0.48	0-0.67	0-1.38	0-0.32
Mean	0.64	0.31	0.2	0.69	0.09
No. haplotypes	34	4	14	9	7
No. specimens	84	7	36	11	30

Haplotypes are divided into four groups labelled as follows: the Iberian clade (which contains haplotypes 20-26), the Italian clade (which contains haplotypes 10-17 & 28), the Central European clade (which contains haplotypes 1-9 & 30-34), and the Eastern Haplotype Set (which contains haplotypes 18, 19, 27 & 29).

The Iberian clade contains samples collected from the Iberian putative refuge, and northern France in the invaded range. The Italian clade represents the Italian putative refuge and the UK in the invaded range. The Central European clade represents the Balkan, Italian, and Turkish putative refugia, and northern France, Germany, and the UK in the invaded range. The eastern haplotype set represents Turkish and Iranian putative refugia. The Italian clade contains the highest sequence differences among haplotypes with a mean between of 0.69% (range of 0-1.38%) uncorrected sequence divergence between sequences.

The following important results are apparent:

(a) Within species genetic diversity is low.
(b) Haplotype frequency is skewed towards a small number of haplotypes (H2 & H21)
(c) Only a small number of haplotypes contribute the majority of those found in the invaded range (H2 & H21, discussed below)

Haplotype 2 (Central European clade) and haplotype 21 (Iberian clade) are very frequent (Table 4.13), accounting for 45% of all individuals sampled between them. These haplotypes represent most of the individuals sampled from within the non-native range (24/35 haplotypes).

1. Structure of populations

(a) Evidence of refugial subdivision

The haplotype tree illustrated in Figure 4.7 provides strong evidence for the refugial structure of European populations of *M. stigmatizans*. Thus, in terms of the predictions made in Figure 4.1 $\mathbf{H_0}$ can be rejected. Three clades containing samples from putative refugia are supported, a 'central European clade, an 'Italian clade', and an 'Iberian clade', with several eastern haplotypes arising from a basal polytomy. Of the clades inferred, the Central European and Iberian clade are well supported, but the Italian clade is less well supported with a Bayesian posterior probability of only 76%.

Italian haplotypes do not fall into just one clade, but are distributed between the Italian (haplotypes 10-17 & 28) and Central European clades (haplotypes 7-9). The separation of Italian haplotypes could be due to lineage sorting or more recent gene flow. Given the apparent refugial subdivision in the data and the dispersal ability of *M. stigmatizans*, the data are taken as being most compatible with $\mathbf{H_1}$: the oak hypothesis, with recent geneflow partially disrupting the pattern, rather than $\mathbf{H_2}$: the gall wasp hypothesis.

The reason for the split in Italian haplotypes is not immediately apparent. As noted in Chapter 1, it is possible that Italy was divided into more than one refugium during the ice ages. Thus, one possible cause for the observed pattern could be that northern Italian populations were more closely related to populations in the Balkan refuge than southern Italian populations. However, there is no obvious geographical separation for the specimens sampled, since division between Italian haplotypes in the different clades is not correlated with a north-south or even an east-west division (Table 4.10). Another possible explanation is that post-glacial geneflow between Italian and Balkan refugial populations either during, or subsequent to ice ages has disrupted the refugial signature in both groups of wasp. Italy and the Balkans are adjacent refugia, and it is easy to imagine a situation whereby taxa from either putative refugium could be displaced from one to the next, depending on their dispersal ability. Chalcids are generally small, weak-flying insects that are easily distributed by wind currents. Furthermore, there is evidence that chalcid wasps might actually deliberately use the wind as a dispersal strategy (Compton *et al* 2000). In comparison, the dispersal ability of oaks is limited by the production of large seeds that are dispersed either by gravity or over relatively short-distances by animals. Although oaks are wind pollinated, this has only been demonstrated over relatively short distances. The dispersal ability of gall wasps is probably much greater than that of *M. stigmatizans*. *Andricus* species have a sexual generation that are effectively aerial plankton (Stone & Sunnucks 1993), being both smaller and typically far more numerous than *M. stigmatizans*. The observed pattern is compatible with dispersal across the Adriatic via easterly winds carrying wasps from the Balkans. Thus, perhaps *M. stigmatizans*, which most probably has a dispersal ability that is less than its gall wasp hosts but significantly greater than oaks, shows a population structure that is effectively between that of the two for this reason.

1(b) *Colonisation history*

All clades in the phylogram illustrated in Figure 4.7 arise from a polytomy. Thus, **Ho**, the hypothesis that refugial populations are connected by a polytomy with no obvious basal group, cannot be refuted.

The long-branch length separating haplotypes 10 and 28 from the other haplotypes in the Italian clade is a mystery. Also interesting is haplotype 20, which falls within the Iberian clade. This haplotype originates from specimens obtained from Moroccan and Portuguese *A. quercustozae* and appears to be appreciably distant to the other haplotypes present in Iberia. However, with reference to Table 4.13 it is apparent that this haplotype is distinct only due to the absence of a single thymine at position 498. This illustrates the limited resolution provided by the locus used for this question. It appears that resolution at a finer scale will be necessary before strong evidence regarding the colonisation of Europe by *M. stigmatizans* can be gained.

2. *Evidence of host-mediated subdivision*

(i) *Host oak section*

As discussed above, there is no variation in oak section for *M. stigmatizans* so this trait cannot be discussed.

(ii) *Gall location*

Gall location is mapped onto the haplotype phylogeny for *M. stigmatizans* in Figure 4.8 below. The Iberian clade shows 100% bud galling. This is probably an artifact of sampling since no acorn gall hosts in Iberia (*Andricus pictus, A. grossulariae asex, A. dentimitratus*) were sampled. Several haplotypes were reared from galls that originated from different locations (haplotypes 2, 15, 31). This suggests that the existence of long-lived host race lineages structured by gall location is unlikely. However, the possibility that gall location might exert an effect upon structuring *M. stigmatizans* populations cannot be ruled out. Host races may occur as a result of short-lived local assortative mating among population subsets, which would not necessarily result in mtDNA haplotype differentiation. In addition, it is possible that ancestral polymorphism is masking any underlying structure. These possibilities can only be tested using a higher resolution nuclear marker.

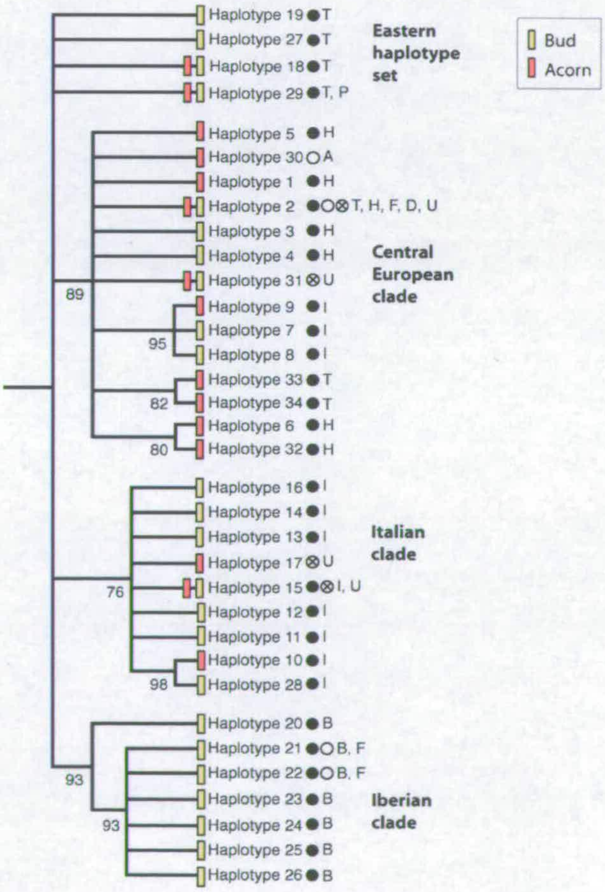


Figure 4.8 *Megastigmus stigmatizans* cyt*b* haplotype tree with gall location mapped onto corresponding branches

(iii) *Gall phenotype*

Spines

The pattern for the distribution of surface spines over haplotypes (Figure 4.9) is similar to that of gall location. The Iberian clade contains only spine-free galls, but as above this pattern is an artifact of sampling. Although spiny galls are present in Iberia (*A. grossulariae*, *A. coriarius*) no parasitoids from these galls were available for sampling. Again, some haplotypes (haplotypes 2, 18 and 29) were obtained from gall species that displayed both character states, suggesting that the existence of long-lived host races partitioned according to this trait is unlikely. However, the possible effect of spines on population subdivision cannot be ruled out.

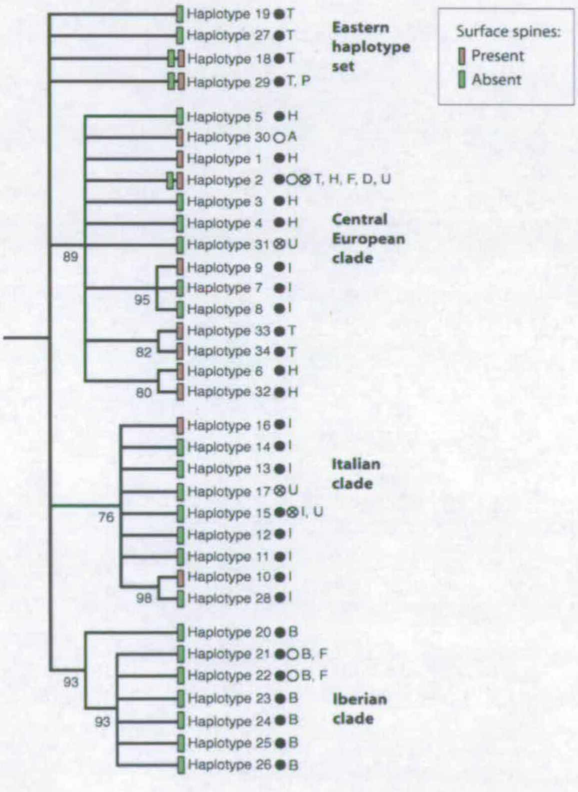


Figure 4.9 *Megastigmus stigmatizans* cyt*b* haplotype tree with presence or absence of surface spines mapped onto corresponding branches

Sticky surface

Presence or absence of a sticky surface is mapped onto haplotype distributions in Figure 4.10 below. Again some haplotypes (haplotypes 2, 15, 21 and 31) were recovered from galls with and without a sticky surface. However, the same conclusions as reached above are applicable for this trait.

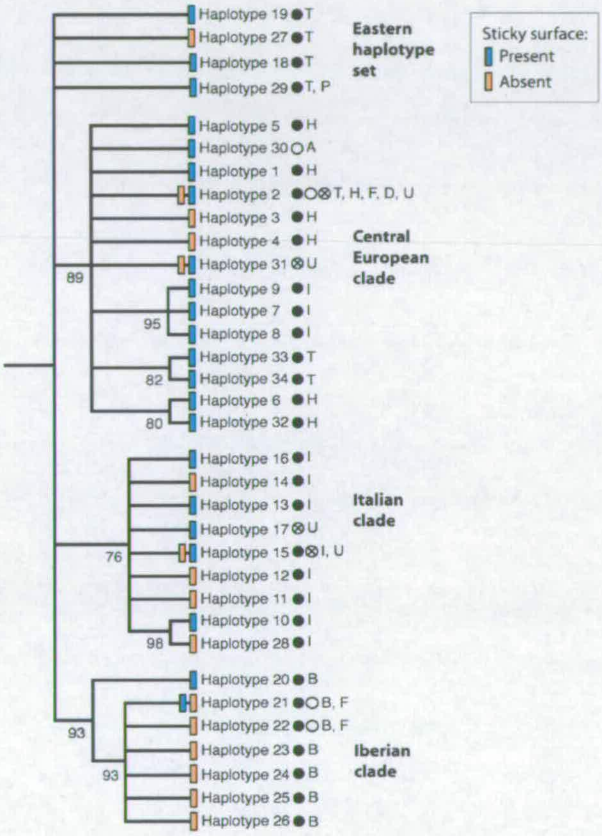


Figure 4.10 *Megastigmus stigmatizans* cyt*b* haplotype tree with presence or absence of sticky surface mapped onto corresponding branches

Internal airspace

Presence or absence of an internal airspace is mapped onto haplotype distributions in Figure 4.11 below. Once more there is no strong pattern, and multiple haplotypes were recovered from both galls with and without internal airspaces.

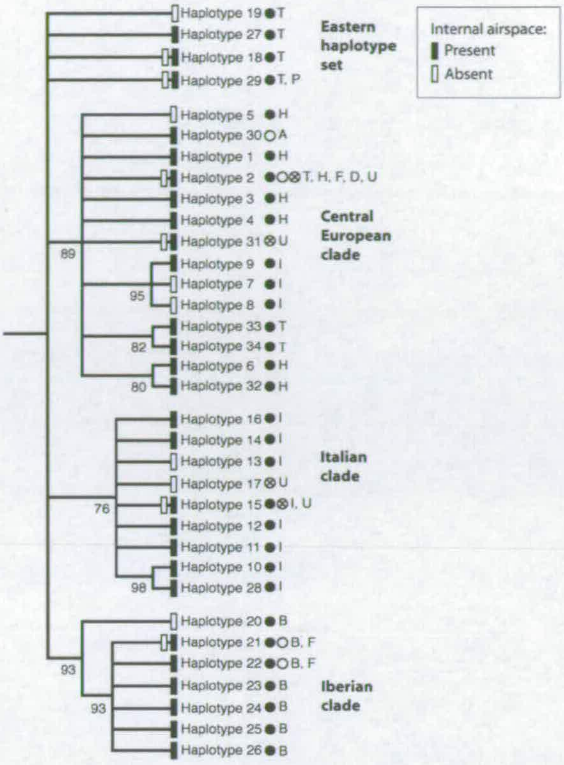


Figure 4.11 *Megastigmus stigmatizans* cyt*b* haplotype tree with presence or absence of an internal airspace mapped onto corresponding branches

Host gall species

In Figure 4.12 the gall species from which the specimen emerged is mapped onto the haplotype distribution. No clear pattern is apparent, and further sampling is required from Iberia, since a low diversity of hosts were sampled from this refugium (samples many consist of *A. kollari*).

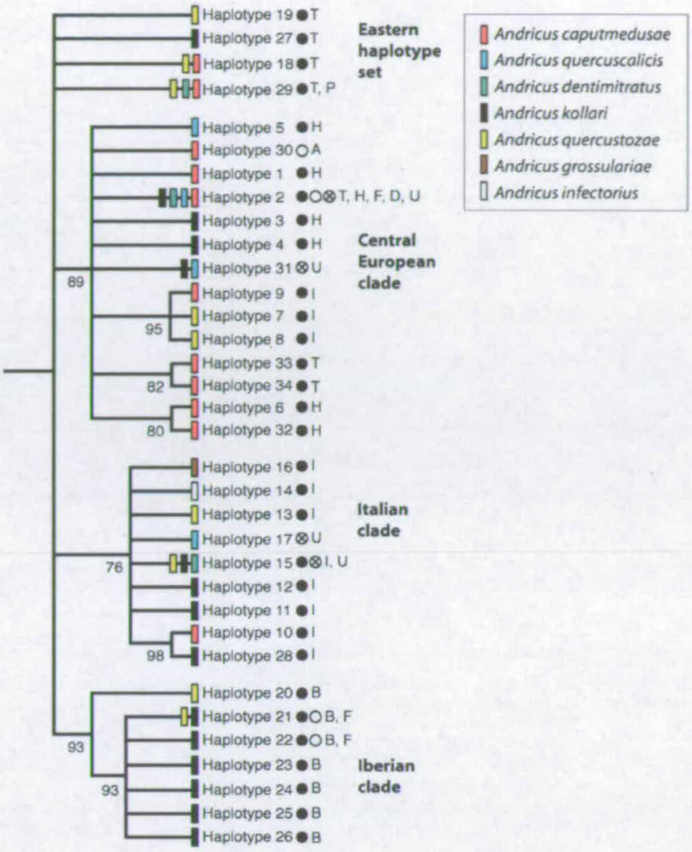


Figure 4.12 *Megastigmus stigmatizans* cyt*b* haplotype tree with host gall species mapped onto corresponding branches

Overall, there is little support for the existence of host-races structured according to host gall traits. Haplotypes arising from galls possessing different traits are scattered throughout clades with no clear pattern visible. Furthermore, the existence of several haplotypes originating from specimens sampled from galls possessing multiple traits argues against substructure due to the presence of long-lived host races based on host gall traits for this species. If prediction (iv) was correct, that of refugial structure and specialism, we would expect to see some level of structure according to gall species or traits occurring within clades. However, the existence of host races cannot be ruled out. In addition to the points made above regarding the possible existence of short-lived host races as a result of local assortative mating among population subsets, and potential clouding of genetic signal due to the sorting of ancestral polymorphism, other possibilities might also prevent the ability to detect host races. As raised in the analysis section above (section 4.5.2), the presence of host races might be missed due to a failure to sort host haplotypes by an appropriate combination of gall traits. Such patterns might only be picked up in multivariate analyses of gall traits. Furthermore, analyses of haplotype frequency might be required to pick up weaker effects. Thus, in summary population structure as inferred based on mtDNA haplotype data corresponds with prediction (iii) made in the introduction to this chapter, the case expected when parasitoids are generalist and dispersal between refugia is rare and lineage sorting complete. Clear evidence of refugial subdivision exists, but further structure within refugia due to host gall traits is absent. Clarification of the possible existence of host races requires further work, utilising nuclear markers. This point is revisited at the end of this Chapter.

3. *Contemporary range expansion*

The haplotypes of wasps sampled in the UK fall within two clades, the Italian clade (1 specimen each of haplotypes 15 & 17) and the Central European clade (6 specimens of haplotype 2 and 7 specimens of haplotype 31). Host shifts from native hosts are not a possibility for *M. stigmatizans* in invading gall wasp hosts in the UK, since the parasitoid is not native in the north of Europe. However, wasps with these haplotypes in the UK could originate either through the pursuit or introduction hypotheses. These possibilities will be discussed in turn.

Pursuit

Evidence is compatible directly only with pursuit from the Balkans and pursuit from Italy. However, Italian clade haplotypes are absent from northern Europe. Were UK sequences in fact to be derived from Italy, such a gap in their distribution could only be explained by incomplete sampling. In contrast, Central European clade sequences have a continuous distribution from the Balkans to the UK, making this scenario more likely. However, though evidence is compatible with pursuit from the Balkans, it cannot exclude introduction for haplotype 2 (see below).

Although haplotype 31 is unique to specimens sampled from the UK, haplotype 2 can be traced back through France and Germany to refugial populations in Hungary and Turkey. Its presence in specimens sampled in the UK is compatible with pursuit of invading host-alternators from their Balkan refuge. Since haplotype 2 is also present in Turkey, its presence in the UK is also compatible with introduction. As mentioned previously, some

Andricus kollari sampled within the UK display a haplotype unlike any other sampled within Europe, suggesting that they are descended from individuals that were imported and that invaders from Europe have been unable to become established by crossing the channel. Under the pursuit hypothesis, *M. stigmatizans* individuals sampled from *A. kollari* in the UK displaying haplotype 2 therefore presumably pursued either *A. kollari* or *A. quercuscalicis* to the north of France and were then blown across the channel. It is possible that haplotype 31 is also derived from individuals that pursued their hosts across northern Europe, and that it was not found in hosts sampled from northern Europe due to incomplete sampling.

As noted in Chapter 3, Iberian populations of *M. stigmatizans* have crossed the oak divide that limits the distribution of their gall wasp hosts, and now attack host-alternators in northern Europe. However, although they have managed to reach the very north of France as demonstrated by specimens sampled from locations such as St Malo, Iberia-specific haplotypes have yet to be detected in the UK. This could be due to an inability either to cross the channel, or to survive once in the UK, or due to incomplete sampling.

Haplotypes 17 and 15 were not sampled elsewhere in northern Europe, but it is possible that this is due to incomplete sampling and specimens possessing this haplotype in the UK could derive from lineages that pursued gall wasps originating from Italy.

Introduction

Haplotype 2 was sampled at a high frequency in the UK and northern Europe. Presence of a widespread, common haplotype in northern Europe and the UK is uninformative for distinguishing between the pursuit and introduction hypotheses since the widespread haplotype could also have been present the origin of the introduced galls.

Haplotype 31 could either be an unsampled Balkan–specific haplotype, comparable with pursuit, or a more eastern haplotype that clusters in the same clade, as for haplotype 33 and 34 from Turkey. Thus, this haplotype in the UK was either:

(a) introduced from the east, but has not been sampled form there, or

(b) originates as a result of pursuit form Hungary/C.Europe but has not been sampled there.

Given that haplotype 31 was not found in any northern European or Balkan samples, the evidence suggests that introduction might be the more likely explanation.

Although haplotypes 15 and 17 fall into the Italian clade, in addition to a potential Italian origin, they may be derived from a centre of origin somewhere in the east that has yet to be sampled. The evidence for a centre of origin for *M. stigmatizans* in the east is inconclusive, but if this were the case the haplotypes in the Italian clade could represent a subset of diversity from this centre of origin. Under the introduction hypothesis, the UK sampled haplotypes in the Italian clade would represent sorting of ancestral polymorphism in two independent introduction events, 1 to Italy, and 1 to the UK. Alternatively, it is also possible that *A. kollari* was imported into the UK from Italy. Sharing of genotypes between Italy and Turkey/Greece has been demonstrated for one gall wasp (*Andricus quercustozae*), and a number of eastern Mediterranean gall wasps absent from most of the Balkans and northern Italy are found in southern Italy (the *insana* form of *A. quercustozae*, *A. megalucidus, A. curtisii*). The same is true of some oaks including *Q. trojana* (section Cerris) and *Q. infectorius* (section Quercus). While these distribution patterns are compatible with ancient links between Italy and Turkey, further sampling is required, perhaps with more polymorphic markers, before this hypothesis can be explicitly tested.

Therefore, no strong conclusions can be made to separate pursuit and introduction based on the available data.

4.6.2 *Megastigmus dorsalis*

44 individuals revealed a total of 16 haplotypes for the 433 bp cyt*b* fragment. Of the 433 sites examined 21 were variable and 14 were parsimony informative. The level of diversity between haplotypes was low, with corrected pairwise distances with a range of 0-3.92% and mean of 0.98%.

Table 4.15 Haplotypes recovered for *M. dorsalis*. Numbers at the top of the table indicate nucleotide position from the 3' end of the fragment to a variable site. The column labelled *f* indicates the frequency with which each haplotype was sampled. Letters in the column labelled 'geographical location' refer to the country from which the specimen was sampled: T=Turkey, P=Iran, H=Hungary, F=France (excluding SW France), G=Greece, U=Britain, I=Italy, B=Iberia.

Megastigmus dorsalis Cytochrome b haplotype table

Haplotype number	4	10	14	79	89	92	95	118	147	157	199	211	244	253	265	270	287	310	325	340	355	*f*	Geographical location
	T	G	A	T	G	G	T	T	C	C	A	A	C	T	A	C	C	T	C	T	C		
Haplotype 1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	23	H, B, I, G, F, U
Haplotype 2	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	4	T, I
Haplotype 3	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2	H
Haplotype 4	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.	.	1	F
Haplotype 5	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.	.	.	.	1	T
Haplotype 6	.	.	.	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	1	I
Haplotype 7	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G	.	.	A	.	.	.	1	U
Haplotype 8	.	.	G	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	2	I
Haplotype 9	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	1	I
Haplotype 10	.	A	.	.	A	A	A	A	T	T	G	.	T	.	G	.	.	.	.	C	T	1	T
Haplotype 11	C	A	.	.	A	A	A	C	T	.	G	.	.	.	.	T	T	.	.	C	T	1	T
Haplotype 12	.	A	.	.	A	A	A	A	T	T	G	.	.	.	G	.	.	.	.	C	.	1	T
Haplotype 13	.	A	.	.	A	A	A	.	T	.	.	.	.	.	G	.	.	.	.	C	.	2	B, I
Haplotype 14	.	A	.	.	A	A	A	A	T	T	.	.	T	.	.	.	.	.	.	C	.	1	P
Haplotype 15	.	A	.	.	A	A	A	A	T	T	.	.	.	.	G	.	.	.	.	.	.	1	P
Haplotype 16	.	A	.	.	A	.	A	A	T	T	G	.	.	.	G	.	.	.	.	C	.	1	P

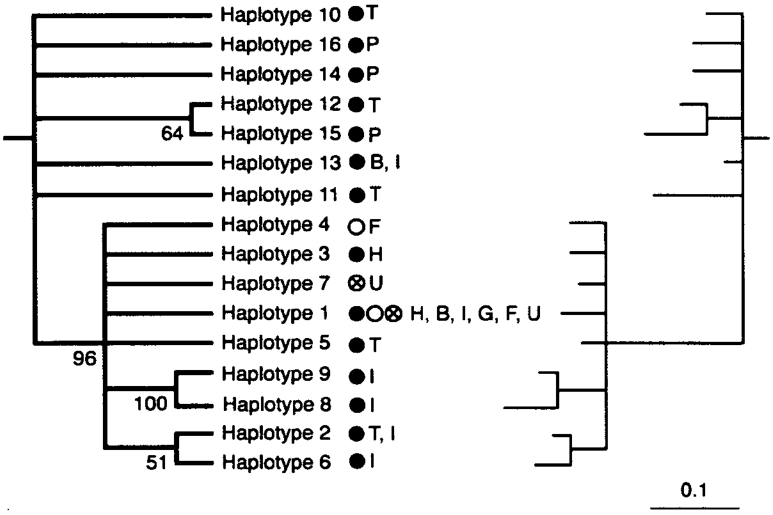


Figure 4.13 *Megastigmus dorsalis* haplotype tree for a 433 bp fragment of the cyt*b* gene (left-hand side), and a phylogramatic respresentation of the same tree (right-hand side). Filled circles indicate specimens sampled from a putative refugial area. Clear circles indicate specimens sampled from the invaded range. Circles with a cross indicate specimens sampled from Britain. Letters at the end of branches refer to the following locations: T=Turkey, P=Iran, H=Hungary, F=France (excluding SW France), G=Greece, U=Britain, I= Italy, B=Iberia. Numbers adjacent to clades indicate Bayesian posterior probabilities. The scale bar indicates percentage sequence divergence for the phylogram.

Examination of the results reveals the following general points:

(a) Haplotype frequency is skewed towards a small number of haplotypes (Table 4.15).

(b) Haplotype 1 contributes all but one of the haplotypes found in the invaded range.

Eastern sequences are mostly basal, with central European sampled haplotypes generally derived.

1. Structure of populations

(a) Evidence of refugial subdivision

Figure 4.13 illustrates that there is one well-supported clade consisting of haplotypes 1-9, which were mainly sampled from central Europe. Basal to this clade is a polytomy containing haplotypes sampled predominantly from Turkey and Iran. Unlike *M. stigmatizans,* haplotypes from specimens collected in Iberia do not form a derived clade. This pattern is not consistent with any of the specific refuge hypotheses made at the start of the Chapter, and suggests that further sampling is required.

(b) *Colonisation history*

A group of predominantly eastern sequences lie basally as sister groups to a derived clade consisting mainly of haplotypes sampled from central Europe. Apart from the presence of an Iberian and an Italian haplotype lying outside the defined clade, this pattern fits the 'out of Anatolia' hypothesis well. However, due to the limited sampling and resolution it is not possible to be conclusive.

2. *Evidence of host-mediated subdivision*

Table 4.16 summarises the range of host-associated traits that were sampled in this study. As discussed above, this range represents a compromise between which traits were targeted, and which traits were successfully sampled.

Table 4.16 Table showing haplotype distribution and gall traits in *M. dorsalis*. All hosts are asexual generation except for *S. politus*

Host gall species	Oak section	Location on tree	Spines	Sticky surface	Internal airspace	Haplotypes associated
Andricus coriarius	Quercus	bud	-	-	-	H1, H2, H3, H5, H8, H9, H13
Andricus grossulariae	Quercus	multiple	+	+	-	H1, H11, H12, H15
Andricus hungaricus	Quercus	bud	+	-	+	H1
Andricus kollari	Quercus	bud	-	-	-	H1, H4, H7
Andricus lignicola	Quercus	bud	-	-	-	H1
Andricus lucidus	Quercus	bud	+	+	-	H6, H10, H14, H16
Synophrus politus	Cerris	multiple	-	-	-	H1

(i) *Host oak section*

All specimens originated from galls on section Quercus oaks, except for two specimens from *Synophrus politus*, an inquiline species that attacks small host galls on section Cerris oaks. However, *M. stigmatizans* reared from this species share a haplotype with specimens derived from section Quercus oaks. This argues against discrete clades for haplotypes associated with specific oak sections. Thus, if host oak section is important for population structure, the mtDNA marker used for this study does not offer enough resolution to pick up the effect.

(ii) *Gall location*

Figure 4.14 illustrates the pattern observed when gall location is mapped onto the *M. dorsalis* haplotype tree.

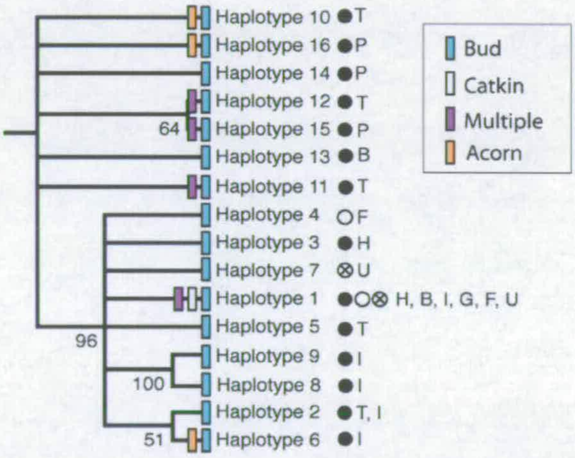


Figure 4.14 *Megastigmus dorsalis* cyt*b* haplotype tree with galling location mapped onto corresponding branches. The term 'multiple' refers to *Synophrus politus,* which can gall stems and the leaf petiole.

No clear pattern is present for population subdivision according to gall location. The majority of samples originate from bud galls, so more complete sampling is necessary to investigate the issue further. However, haplotype 1 contains specimens sampled from catkins and *Synophrus politus,* which is able to gall multiple locations, thus suggesting that the existence of long-lived host races according to galling location might be unlikely.

(iii) *Gall phenotype*

Spines

No clear pattern is apparent for the distribution of surface spines over haplotypes illustrated in Figure 4.15. It is evident that much wider sampling is required to include more specimens reared from galls with surface spines.

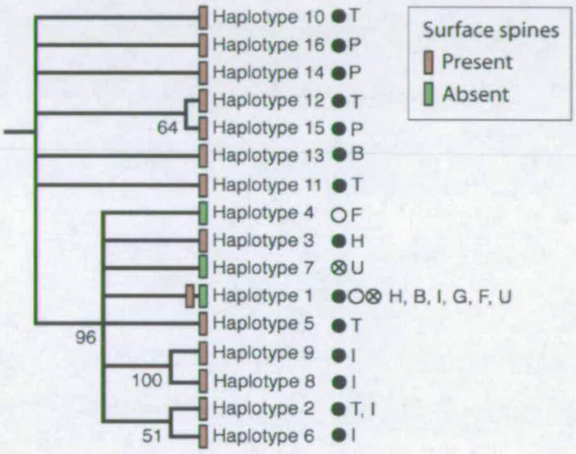


Figure 4.15 *Megastigmus dorsalis* cyt*b* haplotype tree with presence or absence of surface spines mapped onto corresponding branches

Sticky surface

Figure 4.16 illustrates the presence or absence of a sticky surface. The pattern is very similar to that observed for surface spines. Both of the trees show that haplotype 1 was reared from gall in which both trait states were present. However, more data is required before generalisations can be drawn.

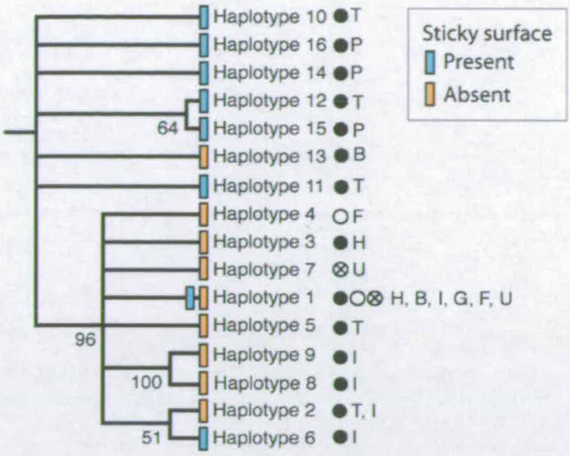


Figure 4.16 *Megastigmus dorsalis* cyt*b* haplotype tree with presence or absence of a sticky surface mapped onto corresponding branches

Internal airspace

A. hungaricus was the only species with an internal airspace for which a specimen was successfully reared. The specimen shares haplotype 1 with specimens reared from galls without an internal airspace. However, as with the previous traits more much wider sampling to include a more balanced set of trait distributions is required before generalisations can be drawn.

Host gall species

Samples from seven different host species were included, six of which were from the genus *Andricus* and one of which was from the genus *Synophrus*. Figure 4.17 illustrates the pattern observed when gall species is mapped across haplotype distributions.

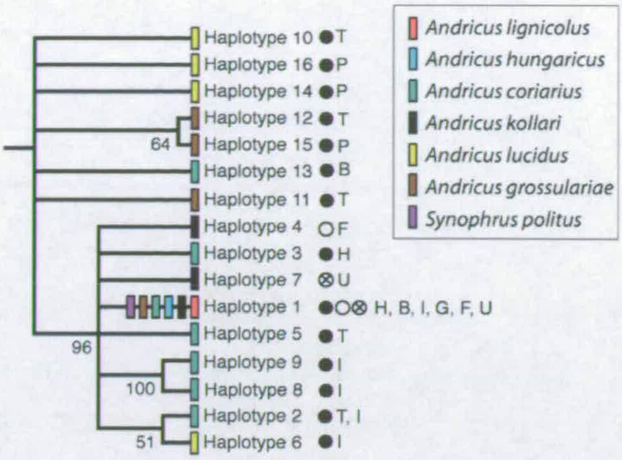


Figure 4.17 *Megastigmus dorsalis* cyt*b* haplotype tree with host gall species mapped onto corresponding branches

Haplotype 1 was obtained from specimens that emerged from all of the host species considered here, except for *Andricus lucidus*. Therefore, there is no strong support for host species association with specific haplotypes. This means either that there are no host races according to species or that host races are present but are obscured by persistent sorting of ancestral polymorphism.

In general, further sampling is required for many of the traits examined here. Traits that require further sampling are listed in Table 4.17 below.

Table 4.17 Gall traits that should be targeted in order to obtain better samples. Ticks indicate that the parasitoid attacks the corresponding gall species.

		Present/absent	
Gall trait	Gall species	*M. dosalis*	*E. brunniventris*
Section Cerris	*Aphelonyx*	✓	✓
	Pseudoneuroterus macropterus	✓	✓
	Neuroterus languginosus	✗	✓
	Synophrus politus	✓	✗
Acorn	*Andricus caputmedusae*	✓	✓
	Andricus dentimitratus	✓	✓
	Andricus quercuscalicis	✓	✓
Leaf galls	*Cynips quercus*	✓	✓
	Cynips quercusfolii	✓	✓
	Cynips longiventris	✓	✓
Surface spines	*Andricus lucidus*	✓	✓
	Andricus grossulariae	✓	✓
	Andricus caputmedusae	✓	✓
Sticky surface	*Andricus quercuscalicis*	✓	✓
	Andricus dentimitratus	✓	✓
	Andricus coronatus	✓	✓
Internal airspace	*Andricus quercustozae*	✓	✓
	Andricus quercuscalicis	✓	✓
	Andricus coronatus	✓	✓
	Andricus glutinosus	✓	✓

3. *Contemporary range expansion*

Pursuit

One haplotype (haplotype 1) was very common, and contributed most of the haplotypes sampled from specimens in northern Europe including Britain. Haplotypes from specimens collected in the UK are not distinct from other sequences within the central European clade. Seven specimens were found to possess haplotype 1, whilst one specimen was found to possess haplotype 7. Haplotype 1 is found along a transect leading back to the Balkans from the UK, a pattern that is compatible with the pursuit hypothesis for *M. dorsalis* found in the UK. However, haplotype 1 also is found in Iberia, so this refuge also cannot be excluded as a potential origin. Generally, common and geographically widespread haplotypes are largely uninformative regarding possible origins.

Host shifts

Since there is very little refugial structure, it is impossible to conclude whether *M. dorsalis* native to the UK have been recruited to the invading gall wasps. A marker with a greater degree of resolution is necessary to answer this question.

Introduction

The position of UK haplotypes in the *M. dorsalis* tree is compatible with the introduction hypothesis from the eastern Mediterranean. Turkish samples cluster within the main clade, and haplotype 1, though not actually sampled in Turkey, is widespread and common enough to suggest that further sampling could reveal it there. The sampling location for UK specimens, Teignmouth in Devon (Table 4.11) is near the site of import of *A. kollari* during the 19th Century, and so is compatible with introduction. A marker with greater polymorphism, and with region-specific alleles capable of discriminating between Balkan and Turkish origins, is required to resolve the issue for *M. dorsalis.*

4.6.3 *Eurytoma brunniventris*

43 individuals revealed a total of 27 haplotypes for the 433 bp cyt*b* fragment. Of the 433 sites examined, 136 were variable, and 105 were parsimony informative. Table 4.20 shows corrected pairwise distances between clades in Figure 4.18 The average distance between sequences for the entire clade is 16%. Such a figure is substantially more than currently held beliefs of what constitutes the usual limits of intraspecific divergences for mtDNA (although see Chapter 6), which are rarely greater than 2% and usually less than 1% (Avise 2000). Three well-supported clades are apparent from the haplotype tree with an additional separate haplotype, and the haplotype Table 4.18 shows that the sites defining these divisions are unambiguous.

Major clade divisions have been indicated with the use of shaded boxes on Figure 4.18 and in Table 4.18. Type 1 consists of haplotypes 1-17, Type 2 consists of haplotypes 18-24, type 3 consists of haplotypes 25-27, and Type 4 consists of only haplotype 28. Type 1 can be further broken down by the presence of a haplotypes 1,2 and 3, which form a small distinct grouping (Subtype 1a).

When haplotypes are split according to type as specified above, the variable site data and corrected distances (Table 4.13) are as follows:

Type 1 (whole clade, n=17 haplotypes): 31 variable sites, of which 20 are parsimony informative

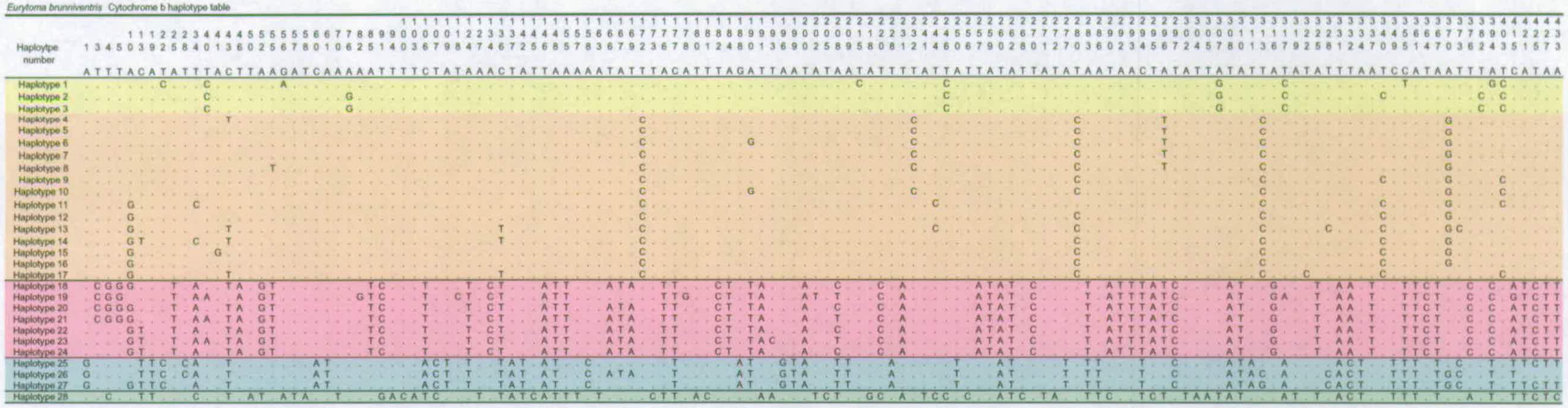
Type 1a (subclade 1a, n=3 haplotypes): 8 variable sites, of which 2 are parsimony informative

Type 1-1a (subclade, n=14 haplotypes): 18 variable sites, of which 18 are parsimony informative

Type 2 (whole clade, n=7 haplotypes): 18 variable sites, of which 12 are parsimony informative

Type 3 (whole clade, n=3 haplotypes): 12 variable sites, of which 8 are parsimony informative

Table 4.18 *Eurytoma brunniventris* cytochrome b haplotype table. Colours indicate putative species groups. Yellow = Type 1a, Orange = Type 1, Pink = Type 2, Blue = Type 3, Green = Type 4



Eurytoma brunniventris Cytochrome b haplotype table

Table 4.19 Haploltype frequencies and distributions for *E. brunniventris*

Haplotype No.	*f*	Geographical location
Haplotype 1	1	B
Haplotype 2	8	B, I
Haplotype 3	4	B, F
Haplotype 4	1	B
Haplotype 5	3	H
Haplotype 6	1	H
Haplotype 7	1	H
Haplotype 8	1	G
Haplotype 9	1	T
Haplotype 10	1	H
Haplotype 11	1	T
Haplotype 12	1	G
Haplotype 13	1	T
Haplotype 14	1	T
Haplotype 15	1	T
Haplotype 16	2	T
Haplotype 17	1	T
Haplotype 18	1	H
Haplotype 19	1	I
Haplotype 20	1	B
Haplotype 21	1	B
Haplotype 22	1	B
Haplotype 23	3	B
Haplotype 24	1	B
Haplotype 25	1	P
Haplotype 26	1	G
Haplotype 27	1	P
Haplotype 28	1	B

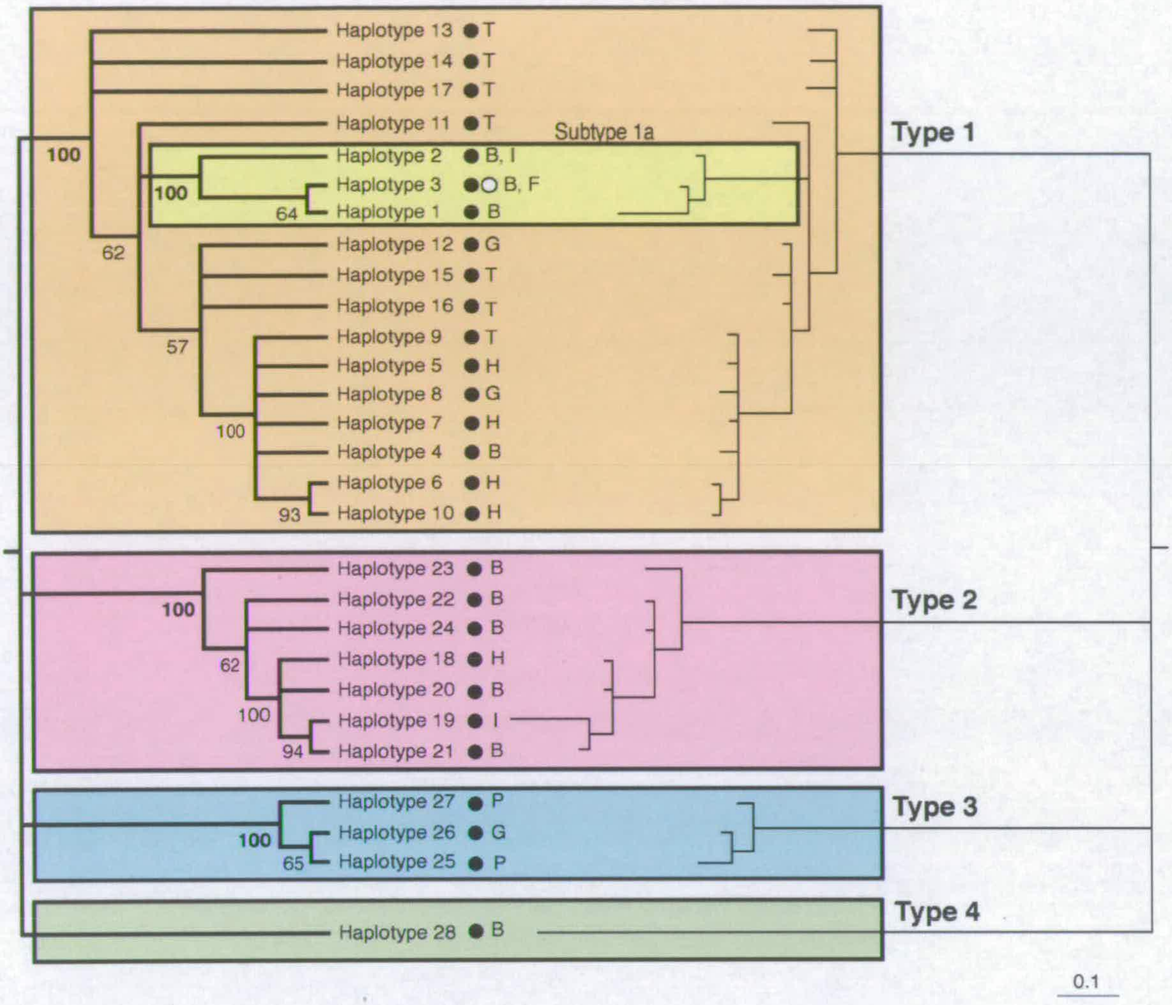


Figure 4.18 *Eurytoma brunniventris* haplotype tree for a 433 bp fragment of the cyt*b* gene (left-hand side), and a phylogramatic respresentation of the same tree (right-hand side). Filled circles indicate samples from a refugial area. Clear circles indicate samples collected from the invaded range. T=Turkey, P=Iran, H=Hungary, F=France (excluding SW France), G=Greece, I= Italy, B=Iberia. Coloured boxes indicate putative divisions between species. Colours correspond to those used in Table 4.18 and indicate distinct clades. The scale bar indicates percentage sequence divergence for the phylogram.

Table 4.20 Corrected distances (in percentage) for the *E. brunniventris* haplotype data

	Whole clade	Type 1	Subtype 1a only	Type 1 excluding subtype 1a	Type 2	Type 3
Range	0-65	0-6.1	0-2.02	0-2.9	0-3.58	0.47-1.22
Mean	15.97	2.4	0.39	1.24	1.39	0.8

1. Structure of populations

(a) Evidence of refugial subdivision

The Type 1 clade includes an Iberian subclade (1a) and a 'Central European' clade of the kind described for *M. stigmatizans*. However, the very limited sampling from Italy precludes discussion of which hypothesis of refugial structure (oak or gall wasp) best describes this type. Type 2 is composed mainly of haplotypes from Iberia, whereas Type 3 is composed of two eastern haplotypes from Iran and a Greek haplotype. In addition to substructure within Type 1, the types themselves show apparent geographic structure. However, additional sequences in Types 2-4 are required for each putative species in order to further investigate phylogeography *ie* patterns both across and within types.

1(b) *Colonisation history*

Type 1 is the only clade for which colonisation history can currently be discussed. In this clade, Turkish sequences are basal, and the pattern is clearly consistent with the 'Out of Anatolia' hypothesis.

2 Evidence of host-mediated subdivision

Given the findings of Acs *et al* (2001) I tried to rear as many specimens from different section oak hosts as possible. In the end the specimens I obtained were mainly derived from section Quercus hosts, which limited which hypotheses could be tested.

Table 4.21 Distribution of haplotypes over host gall species and character states

Host gall species	Location on tree	Oak section	Spines	Sticky surface	Airspace	Haplotypes derived
Andricus conglomeratus	bud	Quercus	-	-	-	H9, H12, H26
Andricus coriarius	bud	Quercus	+	-	-	H5, H7
Andricus coronatus	bud	Quercus	-	+	+	H1, H4, H11, H17
Andricus gallaetinctorae	bud	Quercus	-	-	-	H2
Andricus kollari	bud	Quercus	-	-	-	H2, H3
Andricus polycerus	bud	Quercus	-	-	-	H8
Andricus quercustozae	bud	Quercus	-	+	+	H19, H22, H23, H24, H28
Aphelonyx persica	shoot	Cerris	-	-	-	H25, H27
Cynips quercus	leaf	Quercus	-	-	-	H5, H6, H13, H14, H15, H16
Cynips quercusfolii	leaf	Quercus	-	-	-	H5, H18
Trigonaspis synaspis	leaf	Quercus	-	-	+	H20, H21

(i) *Host oak section*

Only two specimens were successfully reared from section Cerris hosts: haplotypes H25 and H27, and both were reared from *Aphelonyx persica*. Together with haplotype 26 these individuals comprise Type 3. Haplotype 26 was obtained from a gall of *A. conglomeratus* on a section Quercus oak. The presence of this haplotype in the clade containing the haplotypes obtained from the *Aphelonyx* galls argues against the possibility that Type 3 might be a Cerris-specific lineage. More haplotypes are needed from a greater diversity of hosts for all types, before the impact of host oak section can be assessed meaningfully.

(ii) *Gall location*

Figure 4.19 illustrates gall location mapped across haplotype distributions for *E. brunniventris*. On the basis of the available evidence, no conclusive pattern is apparent for the effect of gall location on population subdivision.

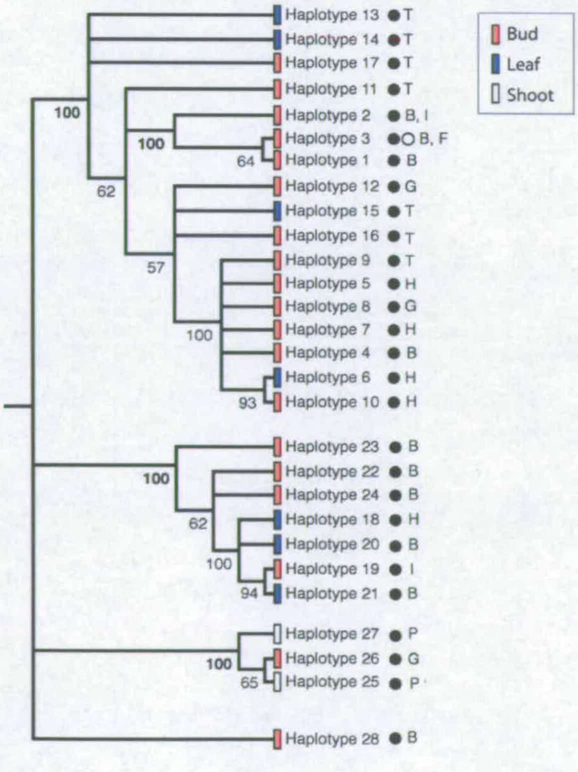


Figure 4.19 *Eurytoma brunniventris* cyt*b* haplotype tree with galling

location mapped onto corresponding branches

(iii) *Gall phenotype*

Spines

Out of the galls sampled, only *A. coriarius* has surface spines, therefore sampling is not adequate to investigate the effect of this gall trait. *E. brunniventris* is known from *A. lucidus A. grossulariae* and *A. caputmedusae* which all possess spines; therefore this issue could be addressed through further sampling of these hosts in future.

Sticky surface

Figure 4.20 illustrates presence or absence of a sticky surface mapped across haplotype distribution for *E. brunniventris*. Again, no clear pattern is visible, but sampling is inadequate for a conclusive assessment. The pattern in the clade comprising haplotypes 25-27 is a sampling artefact, since sticky galls are present in the areas in which these galls were collected, but were not successfully sampled for *E. brunniventris*.

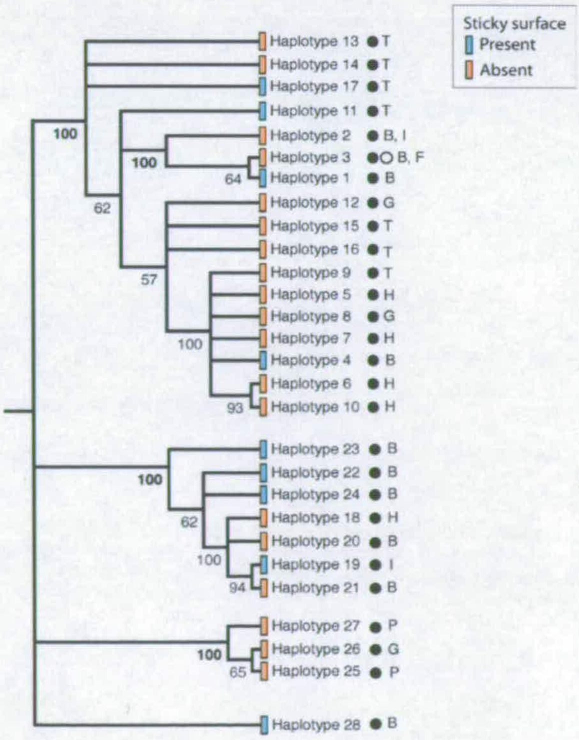


Figure 4.20 *Eurytoma brunniventris* cyt*b* haplotype tree with presence or absence of a sticky surface mapped onto corresponding branches

Internal airspace

Figure 4.21 illustrates the presence or absence of an internal airspace mapped across haplotype distribution. No clear pattern is visible, but sampling is inadequate for a conclusive assessment. A further consideration is that when closely related haplotypes possess different host character states (such as for haplotypes 1 and 3 below), this argues against the existence of long-lived lineage-specific patterns.

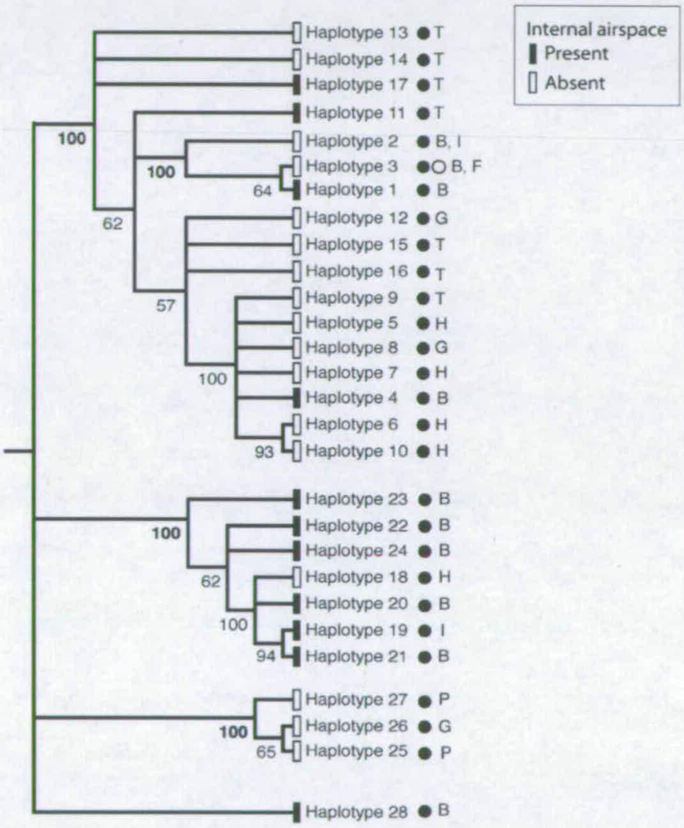


Figure 4.21 *Eurytoma brunniventris* cyt*b* haplotype tree with presence or absence of a sticky surface mapped onto corresponding branches

Host gall species

No structure according to host gall species is immediately apparent. However, more intensive sampling is required for the clades of Type 2, 3 and 4. Particularly since no specimens from *A. lucidus, A. glutinosus or A. multiplicatus* were included in this study (corresponding to Acs *et al* Groups I & II).

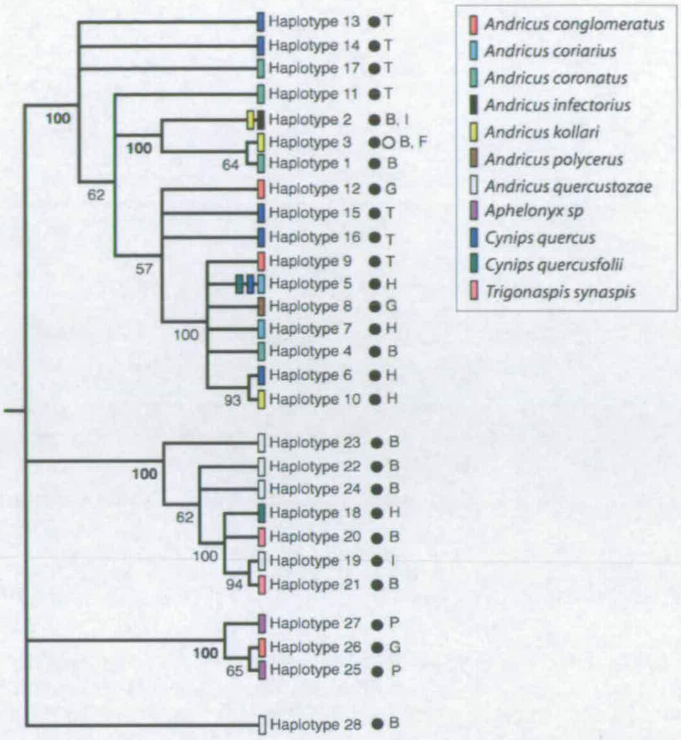


Figure 4.22 *Eurytoma brunniventris* cyt*b* haplotype tree with host gall species mapped onto corresponding branches

Cryptic taxa in E. brunniventris

It is clear that there is strong substructure within the species currently named as *Eurytoma brunniventris*. The level of genetic differentiation within the clades uncovered here strongly supports the existence of multiple sibling species. However, before more detailed discussion of possible structure according to host gall species or specific gall traits can be analysed, more extensive sampling of each subdivision will be required.

4.7 Discussion

This chapter has demonstrated the following patterns in the target species:

(a) At least one of the species, *M. stigmatizans*, shows subdivision into haplotype clades corresponding to putative refugia, in a pattern broadly compatible with the Oak refugia hypothesis. Some haplotypes sampled in Italy were found to lie in a more eastern Balkan clade, suggesting either imperfect refugial separation in the past or recent post-glacial gene flow between these regions. Neither *M. dorsalis* nor *E. brunniventris* show an equivalent strong pattern for available samples.

(b) Two species (*M. dorsalis* and *E. brunniventris*) show patterns consistent with the 'Out of Anatolia' hypothesis, while *M. stigmatizans* shows a basal polytomy.

(c) Colonisation of northern Europe is associated with loss of haplotype diversity, consistent with sequential bottlenecks during range expansion. Current data, which lack haplotypes specific to the eastern Mediterranean, do not allow absolute discrimination between the pursuit and host shift hypotheses.

(d) No haplotype-specific host races were revealed in any of the three target species for the host characters examined.

(e) *Eurytoma brunniventris* showed strong evidence of subdivision into at least four cryptic taxa, whose mutual divergence is probably compatible with designation as separate species.

4.7.1 Refugial subdivision

Only one of the three species tested shows a strong signature of refugial subdivision. As it stands, this suggests that *M. dorsalis* and *E. brunniventris* either never established separate refugial lineages, or that the geographic patterning of such lineages has been broken down by subsequent post-glacial dispersal – in essence, a more extreme version of the explanation for the presence of haplotypes from Italian *M. stigmatizans* in the Central European haplotype clade for that species. However, I believe that further sampling is essential before conclusions can be made for either *E. brunniventris* or *M. dorsalis*. For *E. brunniventris*, such sampling is complicated by the fact that specimens should be evenly distributed (where possible) across Types within this species, but the Types are not distinguishable morphologically!

In the case of *M. stigmatizans*, refugial subdivision is clear. Such subdivision in principle allows each refugial population to follow a separate evolutionary trajectory over the period for which gene flow with other refugia is prevented. Two potentially important variables between refugia are the oak cynipid hosts, and the oaks on which these develop. If these aspects of the host spectrum and possible cues used in host location (see Chapter 1) vary substantially between refugia, and separation is long enough, subpopulations could diverge in host preferences and host location behaviour. *M. stigmatizans* has a relatively narrow host range of large asexual generation host galls, predominantly of

Andricus species (parasitoids reared from the two remaining hosts, both asexual generation galls of *Cynips* species, are very rare). Of the *Andricus* hosts, 3 are found through the whole range of this parasitoid (*A. coriarius, A. kollari* and *A. quercustozae*). One species is restricted to the Iberian Peninsular and northwest Africa (*A. pictus*), while the rest (*A. caputmedusae, A. coronatus, A. dentimitratus, A. hungaricus, A. lucidus* and *A. quercuscalicis*) are found from Italy eastwards, but are absent from Iberia. The primary contrast in available hosts is thus between Iberia and regions further east. The same regions differ in their oak taxa, with the Iberian peninsular showing a regional flora distinct from a more broadly distributed set of species extending from Italy eastwards to Turkey (Jalas and Suominen 1987). One way in which divergence in host preference could be demonstrated would be if haplotypes escaping from the Iberian refugium (as demonstrated in Chapter 3) were to remain restricted to familiar hosts found in northern Europe (primarily *A. kollari* and *A. quercustozae*), and prove unable to colonise unfamiliar hosts, such as the vastly abundant knopper gall, *Andricus quercuscalicis.* To date, however, too few specimens have been reared from knopper galls in France to test this hypothesis.

4.7.2 Longitudinal range expansion

Two parasitoids support the 'Out of Anatolia' hypothesis, suggesting that, as for oaks and gall wasps, this region has acted as a 'cradle' for the diversification of at least 2 parasitoid lineages. It is too early to say whether this pattern is general in parasitoids, and it remains possible that other species will support the 'Out of Iberia' hypothesis. If Anatolia is supported as the cradle for all oak cynipid lineages, then the first parasitoids recruited to Iberian cynipid communities may well have shifted from non-cynipid hosts.

Establishment of a community phylogeography for oak cynipids and their parasitoids requires not only demonstration of suitable phylogenetic relationships between regions, but also an estimation of the time depth of westward expansion in each species. I have yet to date divergence between the relevant clades in each species, but this is planned using the methods applied in Chapter 3. Concordance in both topology and timing of cladogenesis would support the hypothesis that westwards range expansion by gall wasps was followed very closely by range expansion by their natural enemies. If true, this would suggest that the gall wasp hosts did not experience much 'enemy-free space' (Jeffries and Lawton 1984) during their escape from refugia. We know from work on recent gall wasp invasions of northern Europe that even rapidly invading gall wasps recruit communities of natural enemies over periods of decades (Schönrogge *et al.* 1995, 1996a,b, 1998), at least in part by range expansion of parasitoids from glacial refugia (Chapter 3). Westwards range expansion by gall wasps would probably have been far slower than these recent invasions: in contrast to rapid human dispersal of Turkey oak, natural range expansion by oaks proceeds at rates of 500-2000m per year (Bennet 1986; Huntley and Birks 1983). Slower rates of range expansion should facilitate parasitoid pursuit of their hosts.

4.7.3 Parasitoid colonisation of northern Europe

Megastigmus stigmatizans must have become established in southern Britain through either pursuit or introduction, with the latter mechanism strongly supported for some haplotypes. Host shift followed by pursuit is also supported for this species during escape from Iberia (Chapter 3). Both results suggest that the principal factor limiting the distribution of this species is the distribution of suitable hosts, rather than any physiological limitation. Human dispersal of Turkey oak has set in motion a 'trophic cascade' of range expansion, involving first oak, then gall wasps, and then natural enemies. This leads to the prediction that any parasitoids associated with the gall wasps currently invading northern Europe, and absent from existing northern communities, may well also expand their distributions northwards. Examination of the host parasitoid catalogue compiled by Schönrogge & Askew (in press), and the on-line British chalcidoid list (http://www.mapmate.co.uk/checklist/chalcidoidea.htm) shows that most of the parasitoids attacking the invading gall wasps are already recorded from Britain. However, were the remainder to track their hosts from the Balkans (in the case of the invaders *Andricus aries, A. corruptrix, A. grossulariae, A. kollari, A. lignicola, A. lucidus, A. quercuscalicis* and *Aphelonyx cerricola*) or from Spain (only possible for *A. grossulariae* and *A. kollari*), the British list would be enriched by up to 15 species, including 1 eurytomid (*Sycophila iracemae*), 2 torymids (*Megastigmus dumicola* and *Adontomerus crassipes*), 2 pteromalids (*Cyrtoptyx robustus* and *Mesopolobus lichtensteini*), 2 eupelmids (*Eupelmus cerris* and *E. rostratus*) and 8 eulophids (*Aprostocetus cerricola, A. domenichinii, A. viridinitens, Aulogymnus obscuripes, A. testaceoviridis, Baryscapus anasillus, B. berhidanus* and *B. pallidae*). More generally, the range expansions triggered by human dispersal of Turkey oak mean that predictions of changes in species distributions base on climatic envelope modeling (see discussions in Araujo *et al.* 2005, Whittaker *et al.* 2005) may sometimes be based on simplistic assumptions of what limits species distributions.

M. stigmatizans may have reached its distribution limit in the U.K. As discussed earlier in this chapter, it was common in southern England shortly after the import of *Andricus kollari*. Its distribution is now very local, and concentrated in southwestern England. This limit could have several mutually compatible causes, including physiological limitation and low diversity or abundance of available gall wasp hosts (only *A. kollari* and *A. quercuscalicis*).

4.7.4 Host races

No host races were detected using the samples and host traits used here. The small sample size allowed only the detection of the strictest and longest-established host races, associated with subdivision of a species into mutually monophyletic lineages associated with different host subsets. It is possible that less completely separated or more recently formed host races exist in all of the study species, and could be detected using molecular markers with higher resolution. Larger sample sizes would allow the use of haplotype frequency data, while the greatest resolution would probably be achievable with multiallelic nuclear markers, such as microsatellites. As discussed earlier, the intention was to combine analysis of haplotype sequence with microsatellite data, but difficulties in marker isolation and development made this combined approach impossible within my Ph.D. Nevertheless, my results do suggest that strong, long-lived intraspecific partitioning of lineages has not taken place in response to the gall characters examined.

A second possible reason for inability to detect patterns that are in fact there is inappropriateness of the potential structuring traits examined. So little is known about how parasitoids locate and identify gall targets that the selection of traits to test may be inaccurate or oversimplistic. It is possible, for example, that parasitoids partition ‘host character space’ in a complex way, involving interaction of host oak, gall location and gall morphology. One approach for the future would be to examine correlations between genetic structure and combinations of traits, perhaps combined into a single metric using cluster analysis. However, such an approach still needs to overcome the problems associated with generating distances from characters (such as gall location) whose alternative states are not easily converted into meaningful numerical values.

It may simply be true that many of the parasitoids attacking oak cynipid galls really are true generalists. A feature of gall wasp populations is that they fluctuate significantly from year to year (Stone *et al.* 2002), and specialisation on a specific host subset may not be favoured because it carries a risk of local extinction due to failure of host populations. True generalists do have the potential to mediate apparent competition between multiple gall wasp hosts (Holt and Lawton 1993), and the wide host ranges of species such as *Megastigmus dorsalis* mean that the full interaction web in the richest oak cynipid communities would be extremely complex.

4.7.5 Cryptic taxa

Eurytoma brunniventris shows clear subdivision into four taxa. The four types show partial geographic separation, but samples from a single refugial region, such as Iberia, are found in more than one type (Types 1a, 2 and 4). This pattern thus does not conform to the predictions of refugial subdivision. There is also no evidence of strong host-associated subdivision on the basis of the limited sampling available here. Further extensive sampling is necessary to establish the distribution limits and host associations of the four types. One obvious question is the extent to which the types represent separate biological species. Use of mitochondrial markers cannot eliminate the possibility of gene flow through non-assortative mating by males, and future work will use microsatellites to examine the extent to which nuclear markers corroborate the picture obtained using haplotype sequence. The molecular barcodes discussed in Chapter 6 use haplotype sequence for mtDNA, for which sequence divergence between species is commonly above 2% (Avise 2000). Unfortunately, although *E. brunniventris* was included in the barcoding work, there was insufficient time after detection of the four types to include samples from each in the cytochrome oxidase sequencing work. This is an obvious aim of future work.

Demonstration of the cryptic taxa within *Eurytoma brunniventris* is potentially of major significance for gall wasp foodwebs. If the host ranges of the four types differ qualitatively or quantitatively, links currently attributed to a single species will have to be subdivided between types. Then, as discussed earlier for host races, the potential for apparent competition between hosts currently attributed to the single *E. brunniventris* taxon will be reduced. This example illustrates the need to examine population structure in other highly generalist parasitoids including, for example, *Eupelmus urozonus*.

Chapter 5

Life history analysis of the Torymidae

5.1 Introduction

5.1.1 Trophic diversity in the Parasitica

The parasitic wasps (Hymenoptera; Parasitica) are predominantly carnivores, and as larvae feed primarily on the immature stages of insects. However, despite this generalisation the Parasitica show considerable trophic diversity, including species that are partially phytophagous while feeding primarily as parasitoids, others that feed on plant material, particularly seeds, and yet others that induce plant galls. This diversity is found particularly in the superfamilies Chalcidoidea (which includes the parasitoids studied in this thesis) and the Cynipoidea (which includes the gall wasp hosts). In both of these taxa, trophic diversity is found both between families (such as the comparison between the wholly phytophagous agaonid fig wasps and their parasitoid sister group), and within particular families. Examples of chalcid families known to contain a range of phytophagous and carnivorous life histories include the Torymidae, Eurytomidae, Eulophidae, Pteromalidae and Tanaostigmatidae. Of these, the Torymidae, Eurytomidae and Eulophidae include each of parasitoids, seed feeders and gall inducers (La Salle 2005). In the Cynipoidea, the Cynipidae includes both phytophagous gall inducers and inquilines, and their sister group the Figitidae includes both phytophagous inquilines and entomophagous parasitoids (Csóka *et al.* 2005). This diversity raises the issue of the direction and frequency of shifts between alternative trophic strategies. There is general consensus that a parasitoid lifestyle is basal in the Parasitica, but evolutionary transitions between strategies remain little studied. The Cynipoidea is the only group for which switches between trophic strategies have been examined in detail. In the Cynipoidea, switches are extremely rare, from entomophagy to phytophagy, and in the Cynipidae the evolution of gall induction has been associated with dramatic adaptive radiation relative to the Figitidae parasitoid sister group (Ronquist 1994). La Salle (2005) states that gall induction must have evolved at least five times in the Eulophidae and Eurytomidae, once in each of the Pteromalidae and Tanaostigmatidae and twice in the Torymidae. These are minimal estimates based on phylogenetic relationships assuming that recognised taxa are natural groups, and that morphology is an adequate guide to phylogeny. In the absence of detailed phylogenetic reconstructions for these groups, the frequency and directions of trophic shifts remain unresolved. More generally, reconstructing the order and frequency of life history switches is of value in identifying whether taxa that share the same niche are part of a radiation within that niche, or the result of multiple transitions. The first pattern suggests that transitions are difficult, but once achieved give rise to an adaptive radiation. The second suggests that transitions between strategies are 'easy' in evolutionary terms, but that they do not potentiate further speciation.

5.1.2 Hypotheses for the evolution of phytophagy in Chalcidoidea

Given an ancestral parasitoid lifestyle, the existence of two forms of phytophagy (gall induction and seed feeding) raises the additional question of whether shifts can occur directly from entomophagy to either phytophagous state, or whether one is an intermediate step in the evolution of the other. There are four possibilities for shifts from entomophagy: (a) direct shifts to gall induction (b) direct shifts to seed feeding (c) a transition to gall induction and then to seed feeding (d) a transition from seed feeding and then to gall induction.

The possibility of direct transitions is raised by patterns in the Cynipoidea, where there appears to be a direct transition from entomophagous parasitoid to gall inducer in the Cynipidae (Ronquist 1994), and from parasitoid to phytophagous inquiline in the Figitidae (Csóka *et al.* 2005). This pattern cannot exclude the possibility of extinct intermediate lineages showing alternative trophic strategies, however. Direct transitions are also suggested by a number of parasitoids that attack gall inducing cynipids but which also feed on plant material (*eg Eurytoma* species in oak and rose cynipid galls, *Torymus cyaneus* in oak galls, *Dichatomas acerinus* in *Pediaspis* galls *on* Acer, and *Glyphomerus stigma* in rose galls; Csóka *et al.* 2005). Such mixed trophic strategies could in principal be an intermediate on the route to either seed feeding or gall induction. In the latter case, gall induction would provide both additional food resources and potentially provide a protected environment for the developing wasps (Stone and Schönrogge 2003). An additional intermediate step here is perhaps illustrated by the cynipid inquiline *Synophrus politus*, a cynipid inquiline that causes the death of its oak cynipid host early in gall development. Although apparently unable to induce gall development *de novo*, *S. politus* is able to continue gall induction once started by the host. Similar gall modification is shown by *Periclistus* inquilines in rose cynipid galls, and also by some parasitoids in oak galls.

It is not immediately clear which of the sequential transitions from entomophagy to phytophagy is more likely *a priori*. Seeds offer a ready-made highly nutritious food source and seed feeding may have arisen as a result of oviposition mistakes when searching for concealed insect larvae in seeds. Given that the seed is available, one could argue that evolutionary progression to gall induction would be redundant. However, one could argue with equal conviction that evolution of the ability to induce highly nutritious additonal gall tissues would increase fitness.

Distinction between these two 'transition' hypotheses should be possible using phylogenetic relationships between living taxa. If seed eating is an intermediate stage in the evolution of galling from entomophagy, gallers should represent a crown group within a paraphyletic seed-eating assemblage. If galling is an intermediate stage in the evolution of seed eating, the opposite should be true.

5.1.3 Selection of study taxa

A choice of taxa is available in which to investigate the issues outlined above. The taxa that show the greatest diversity are probably the Eurytomidae, the Eulophidae and the Torymidae. Of these, I hoped to work on the European Eurytomidae and Torymidae. Within the timescale of this thesis, I was only able to obtain a reasonable number of specimens for the Torymidae, and thus I only provide results for this family. Addition of the Eurytomidae would have been beneficial in order to examine parallels in patterns in life history evolution in independent lineages. Although not possible here, specimens of a diverse set of Eurytomidae have now been obtained through the generous help of Dr. R.R. Askew, and future analyses will include this family (see below). I now survey the trophic diversity present in the Torymidae and then focus on specific questions for this family.

The Torymidae is a large family including about 875 valid species, 54 genera and 2 subfamilies: the Megastigminae and the Toryminae (Grissell 1995). It is known to be biologically diverse, despite the fact that host records exist for only about one-third of the species (Grissell 1995). A diverse range of feeding types are present in the family, including

(a) the (presumed) basal state of entomophagous parasitoid in non-galling hosts, (b) parasitoids of gall formers, (c) parasitoids of gall formers that also consume gall tissue, (d) seed feeders and (e) gall formers (Grissell 1995, 1999). The dominant feeding strategy is entomophagy as parasitoids on gall inducing insects.

The Torymid bauplan is of adult females with a long, strongly exerted ovipositor (see Figure 5.1 and Plate 5.1 at the end of this chapter). This reflects the fact that the egg is usually deposited through some kind of substrate, usually ether a gall, fruit or seedpod, a task that is accomplished by drilling through the barrier with the long ovipositor. The biologies of the 2 subfamilies are now summarised in turn.

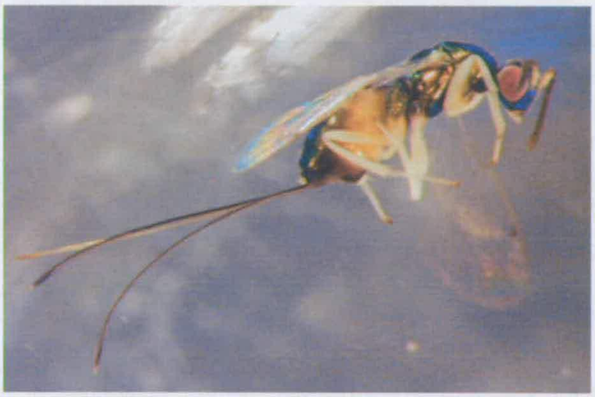


Figure 5.1 A female *Torymus geranii* illustrating the typical Torymid bauplan.

(a) Toryminae: According to Grissell (1995) the Toryminae are primarily entomophagous and are associated with 51 families of insect in 8 orders, with the main hosts being gall-forming Hymenoptera and Diptera. The largest genus in the subfamily by far is *Torymus* with 317 currently recognised species (Grissell 1995). Two species of the genus *Torymus* (*T. druparum* and *T. varians*) are strictly phytophagous and are seed-feeders in apple, pear (Rosaceae) and maple (Aceraceae) (Winter 1978, 1979). At least one species of *Torymus* (*T. philippii*) is a gall inducer on Euphorbiaceae, and another is an inquiline in galls of a Eurytomid on Bamboo (Takashi & Mizuta 1971). Graham and Gijswijt (1998) defined a number of species groups for European *Torymus*, predominantly based on female morphology. The groups including species attacking cynipid hosts for which specimens were available for this study (given after the group name below) are as follows:

Cyaneus group: *Torymus cyaneus, T. notatus.*

Erucarum group: *T. erucarum, T. nobilis.*

Bedugaris group: *T. geranii, T. auratus,* with *T. bedeguaris* in rose galls.

Torymus flavipes is recognised as belonging to a single species group of its own. The status of these species groupings remains untested using molecular markers. Two parasitoids known to have a mixed trophic strategy, involving parasitoid development with consumption of gall plant tissue, are known in the Toryminae: *Torymus cyaneus* attacks oak galls, while *Glyphomerus stigma* develops in rose galls (Csóka *et al.* 2005)

(b) Megastigminae. The Megastigminae are described as primarily phytophagous by Grissell (1995). However, the host associations of many species are unknown. Of a total of 125 described species, 59 (47%) are known to be seed feeders (Roques & Skrzypczynska 2003). The phytophagous species have been reared from at least seven plant families, and most are associated with the seeds of coniferous trees and rosaceous plants. Three species of the genus *Bortesia* are known gall inducers on Proteaceae (La Salle 1995). It is not clear whether additional gall inducing species exist in the genus *Megastigmus*: this has been proposed by Boucek (1988), but I was unable to locate any literature on the biology of specific species. As with the Toryminae, entomophagous members are primarily associated with Hymenopteran and Dipteran gall formers.

Given the trophic diversity described above, I would have liked to include the following taxa for a formal analysis of life-history switches within the Torymidae: *Torymus* seed feeders, *Torymus* gall inducers, *Torymus* inquilines, Toryminae parasitoids of gall inducing Diptera and Hymenoptera, *Bortesia* gall inducers, *Megastigmus* seed feeders and *Megastigmus* parasitoids of gall inducing Diptera and Hymenoptera. In order to bring together specimens for such an analysis, I contacted the following specialists around the world:

- Dr John Noyes, Research Entomologist & chalcid specialist, the Natural History Museum, London, UK.
- Dr Mark Shaw, Curator of Zoology (now retired), the National Museum of Scotland, UK.
- Dr George Melika, taxonomist at the Systematic Parasitoid Laboratory, Köszeg, Hungary.
- Dr Edward Grissell, chalcid systematist and current world authority on the Torymidae, US Department of Agriculture, USA.
- Dr Ebrahim Sadeghi, collaborator on oak gall wasp research, Research Institute of Forests & Rangelands, Teheran, Iran.
- Dr Jocelyn Berry, biocontrol researcher, New Zealand arthropod collection, New Zealand.
- Dr Chris Winks, biocontrol researcher, Landcare Research Institute, New Zealand.

- Dr Gary Taylor, researcher working on gall flies, Adelaide University, Australia.
- Dr Lars Vilhelmsen, Curator, Zoological Museum, University of Copenhagen.
- Dr James Amrine, Division of Plant and Soil Sciences, West Virginia University, USA.
- Dr Chris Burwell, Curator of Entomology, Queensland Museum, Australia.

Ultimately, it was not possible to gather the full range of species I had hoped for and the specimens that were available are listed in Table 5.1 below.

Table 5.1 Torymidae species included in this study. The host plant and/or insect genus from which the specimens was reared is included, in addition to the host family of either the insect (for entomophagous species) or plant (for phytophagous species). Cyt*b*, 28S and coI refer to the gene fragments used in this study, and symbols indicate the successful amplification of the respective fragment. An asterisk after the species name indicates that sequence data was obtained from Genbank for that species. The oak-associated species are shown in Plate 5.1 at the end of this chapter.

Species	Feeding-type	Host insect/plant genus	Host Family	Distribution	cytb	28S	coI
Glyphomerus stigma	Entomophagous	*Diplolepsis* on *Rosa*	Cynipidae	Palaearctic	✓	✓	✓
Megastigmus dumicola	Entomophagous	*Plagiotrochus* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Megastigmus almusiensis	Entomophagous	*Andricus* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Megastigmus dorsalis	Entomophagous	*Andricus* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Megastigmus dorsalis	Entomophagous	*Andricus* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Megastigmus politus	Entomophagous	*Synophrus* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Megastigmus stigmatizans	Entomophagous	*Andricus* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Megastigmus stigmatizans	Entomophagous	*Andricus* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Torymus auratus	Entomophagous	*Cynips* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Torymus beduguaris	Entomophagous	*Diplolepsis* on *Rosa*	Cynipidae	Palaearctic	✓	✓	✓
Torymus cyaneus	Entomophagous	*Cynips* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Torymus erucarum	Entomophagous	*Cynips* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Torymus erucarum	Entomophagous	*Andricus* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Torymus flavipes	Entomophagous	*Neuroterus* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Torymus flavipes	Entomophagous	*Andricus* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Torymus geranii	Entomophagous	*Cynips* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
Torymus nobilis	Entomophagous	*Biorhiza* on *Quercus*	Cynipidae	Palaearctic	✓	✓	✓
T. bedugaris	Entomophagous	*Diplolepsis* on *Rosa*	Cynipidae	Palaearctic	✓	✓	✓
Glyphomerus stigma	Entomophagous	*Diplolepsis* on *Rosa*	Cynipidae	Palaearctic	✓	✓	✓
*T. azureus**	Entomophagous	*Plemeliella* on *Picea*	Cecidomyiidae	Palaearctic	✓		
M. brevivalvus	Entomophagous	*Eurytoma* on *Microcitrus*	Eurytomidae	Australasian		✓	
M. sp 1	Entomophagous	*Procecidochares utilis* on *Aster*	Tephritidae	Australasian		✓	
M. sp 2	Entomophagous	*Fergusonina* on *Eucalyptus*	Fergusonidae	Australasian		✓	
M. sp 3	Entomophagous	? on *Ficus*	?	Australasian		✓	
M. sp 4	Phytophagous	*Syzygium*	Myrtaceae	Afrotropic		✓	
M. transvaalensis	Phytophagous	*Schinus*	Anacardiaceae	Afrotropic		✓	
M. pistacina	Phytophagous	*Pistacia*	Anacardiaceae	Palaearctic		✓	
M. aculeatus var. nigroflavus	Phytophagous	*Rosa*	Rosaceae	Palaearctic		✓	
*M. pingii**	Phytophagous	*Juniperus*	Cupressaceae	East-Asian	✓		
*M. thyoides**	Phytophagous	*Chamaecyparis*	Cupressaceae	Nearctic	✓		
*M. bipunctatus**	Phytophagous	*Juniperus*	Cupressaceae	Palaearctic	✓		
*M. atlanticus**	Phytophagous	*Cupressus*	Cupressaceae	Palaearctic	✓		
*M. wachtli**	Phytophagous	*Cupressus*	Cupressaceae	Palaearctic	✓		
*M. likiangensis **	Phytophagous	*Picea*	Pinaceae	East-Asian	✓		
*Megastigmus specularis**	Phytophagous	*Abies*	Pinaceae	Nearctic	✓		
*Megastigmus tsugae**	Phytophagous	*Tsuga*	Pinaceae	Nearctic	✓		
*Megastigmus rafni**	Phytophagous	*Abies*	Pinaceae	Nearctic	✓		
*Megastigmus pinus**	Phytophagous	*Abies*	Pinaceae	Nearctic	✓		
*M. milleri**	Phytophagous	*Abies*	Pinaceae	Nearctic	✓		
*Megastigmus lasiocarpae**	Phytophagous	*Abies*	Pinaceae	Nearctic	✓		
*Megastigmus hoffmeyeri**	Phytophagous	*Tsuga*	Pinaceae	Nearctic	✓		
*Megastigmus atedius**	Phytophagous	*Picea*	Pinaceae	Nearctic	✓		
*Megastigmus schimitscheki**	Phytophagous	*Cedrus*	Pinaceae	Palaearctic	✓		
*Megastigmus suspectus**	Phytophagous	*Abies*	Pinaceae	Palaearctic	✓		
*Megastigmus pinsapinis**	Phytophagous	*Cedrus*	Pinaceae	Palaearctic	✓		
*Megastigmus pictus**	Phytophagous	*Larix*	Pinaceae	Palaearctic	✓		
*Megastigmus borriesi**	Phytophagous	*Abies*	Pinaceae	Palaearctic	✓		
*Megastigmus rosae**	Phytophagous	*Rosa*	Rosaceae	Palaearctic	✓		
*Megastigmus cryptomeriae**	Phytophagous	*Cryptomeria*	Taxodiaceae	East-Asian	✓		

*Sequence obtained from Genbank

The reduced suite of specimens listed in Table 5.1 prevented me from examining the full range of questions described above. However, these specimens did allow the following questions to be addressed:

1. What is the relationship between seed feeding and entomophagy in *Megastigmus* (the only two feeding types currently known in this genus). Have these feeding types evolved repeatedly within the genus, or does one (and if so, which) represent the basal trophic strategy of the genus?

2. Has a mixed trophic strategy of entomophagy with a phytophagous component evolved repeatedly in the Toryminae? As mentioned above, *Glyphomerus stigma* and *Torymus cyaneus* are known to feed on both insect and plant material. Such a feeding habit might represent a transition between entomophagy and phytophagy. Do these two taxa represent independent evolutions of a mixed feeding strategy within the Toryminae?

3. What is the relationship between Toryminae attacking cynipid hosts on rose and oaks? Have there been transitions between the host gall wasp tribes Cynipini (on oaks) and Diplolepidini (on roses), or are parasitoids attacking one host group basal to the other?

In addition to these life history questions, I will also address the following taxonomic issues:

4. What is the relationship between *Megastigmus* and *Torymus*? Do both genera represent natural (monophyletic) groupings, and which, if either, is ancestral to the other?

5. What is the status of the *Torymus* species groups suggested by Graham and Gijswijt (1998)? Are these groups monophyletic, natural groups, or are they artificial paraphyletic or polyphyletic assemblages?

5.2 Methods

Laboratory techniques

A total of 17 species were sequenced for fragments of the cytochrome *b* (cyt*b*, 433 bp fragment), cytochrome oxidase I (coI, 654 bp fragment), and 28S (658 bp fragment) genes.

There were no missing data and all sequences for a given fragment were the same length, except for 28S where there was a slight variation in sequence length due to the presence of indels (insertion-deletions). Two specimens of each species were sequenced using the molecular techniques detailed in Chapter 2.

Phylogenetic analysis

The three genes listed above were selected for the following reasons: cyt*b* provides resolution in rapidly evolving lineages, whereas 28S resolves deeper nodes and avoids problems of sequence saturation associated with rapidly evolving genes (see Rokas *et al.* 2002 for a comparison of all 3 genes). An additional mtDNA fragment (COI), also relatively rapidly evolving, was also chosen to provide additional resolution among closely related species, since the cyt*b* fragment used here is fairly small. Cytochrome oxidase was also chosen to contribute Torymid sequences to the molecular barcoding study described in Chapter 6. The aim was to amplify all three gene fragments for every species. However, amplification of cyt*b* was not successful in every case, probably as a result of mutations in the primer binding site. Phylogenies were reconstructed for each gene fragment separately and in combination. The combined dataset has the potential advantage of resolving both terminal nodes (cyt*b* and COI) and deeper nodes in the tree, and signals from the three fragments are expected to be concordant. MrBayes allows phylogenetic inference under mixed models to accommodate data heterogeneity (Ronquist & Huelsenbeck 2003), and combined analyses of mtDNA and 28S have already proved useful for resolving relationships in parasitic wasps (*eg* Dowton & Austin 2001, Chen *et al* 2004). Thus, relationships were inferred using Bayesian methods as outlined in the methods chapter.

Models of sequence evolution were estimated using MrModeltest. The conditions used were as follows: For the separate analyses, mtDNA fragments were partitioned by codon position, with the rate heterogeneity model and substitution model unlinked across positions. The GTR model with gamma (Γ) distributed rate variation was specified for the CO1 data set, and GTR + Γ + invariant sites (I) were also specified for cyt*b*. The 28S fragment was also analysed using a GTR + Γ + I model of sequence evolution. However, it is not a protein coding fragment and was not partitioned according to codon position. 28S data can be broken down into 'stem' and 'loop' regions, according to the secondary structure of the RNA molecule for which it encodes. Loop regions are under less constaint due to a lack of nucleotide pairing and generally show higher rates of evolution. However, divison of the 28S gene into such regions requires specialist attention, and was not attempted in this study. The combined analysis was carried out using a GTR + Γ + I model with the following four partitions: mtDNA by codon position, and 28S. Outgroups selected from the Chalcidoidea were *Caenacis lauta* and *Mesopolobus xanthocerus* (Pteromalidae), *Sycophila variegatta* (Eurytomidae) and *Ormyrus nitidulus* (Ormyridae). The relationships between families within the Chalcidoidea are a current focus of phylogenetic research but remain unclear. The set selected includes a group long thought to lie within the Torymidae, and so closely related to them (Ormyridae), with two other families within the Chalcidoidea.

Character mapping

For a formal analysis of life history strategy, it would be necessary to reconstruct ancestral states across trees, instead of simply noting where species arose within a tree, since it is not always possible to identify the correct origin of the trait in quesiton. Ancestral states for each node in the torymid phylogeny could be reconstructed using a computer program such as Bayes-Multistate (Pagel *et al* 2004), and is planned for a later augmentation of this study.

5.3 Results

The fullest availability of samples is for the cyt*b* tree (see Figure 5.2 below), which shows good resolution even among closely related species. The same relationships were supported by the data for COI, and so are not shown separately here. The 28S tree (Figure 5.3) provides moderate support for deep nodes but provides little or no resolution within the clades of closely-related species resolved by cyt*b* sequence. The topology is compatible across both trees, supporting the use of sequences for both genes in a combined analysis (Figure 5.4). One aspect of the combined evidence tree is the apparent inclusion of *Sycophila biguttata* and *Ormyrus nitidulus* within the Torymidae. This is an artifact of taxon sampling and long-branch attraction (Felsenstein 1978) and genuine relationships for these taxa are probably better represented by the analysis in Chapter 6.

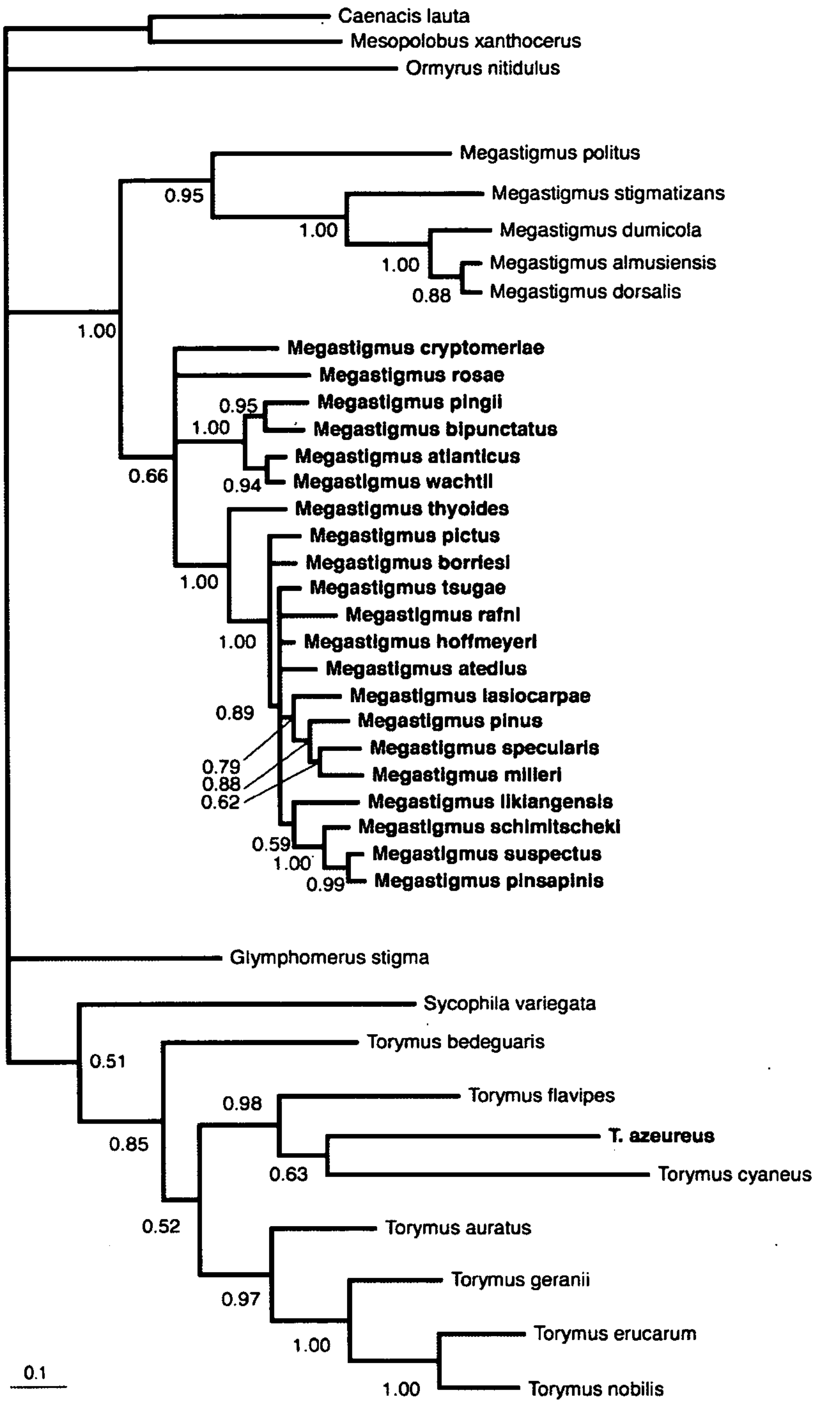


Figure 5.2 Phylogeny for the European oak gall attacking Torymidae for cyt*b* only, including sequences from Genbank (in bold text). Figures beside clades represent Bayesian posterior probabilities.

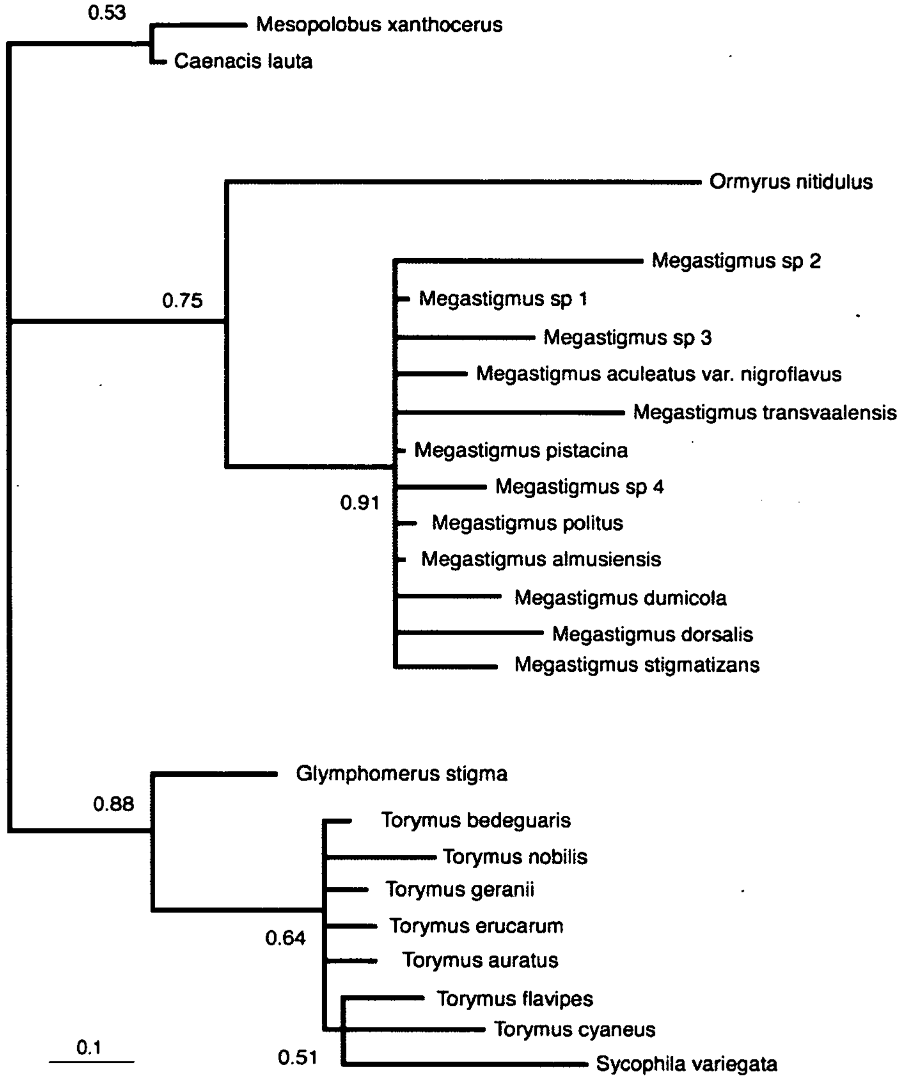


Figure 5.3 Phylogeny for the European oak gall attacking Torymidae for 28S only, including sequences for 7 additional *Megastigmus* species showing a diversity of individual trophic strategies. Figures beside clades represent Bayesian posterior probabilities.

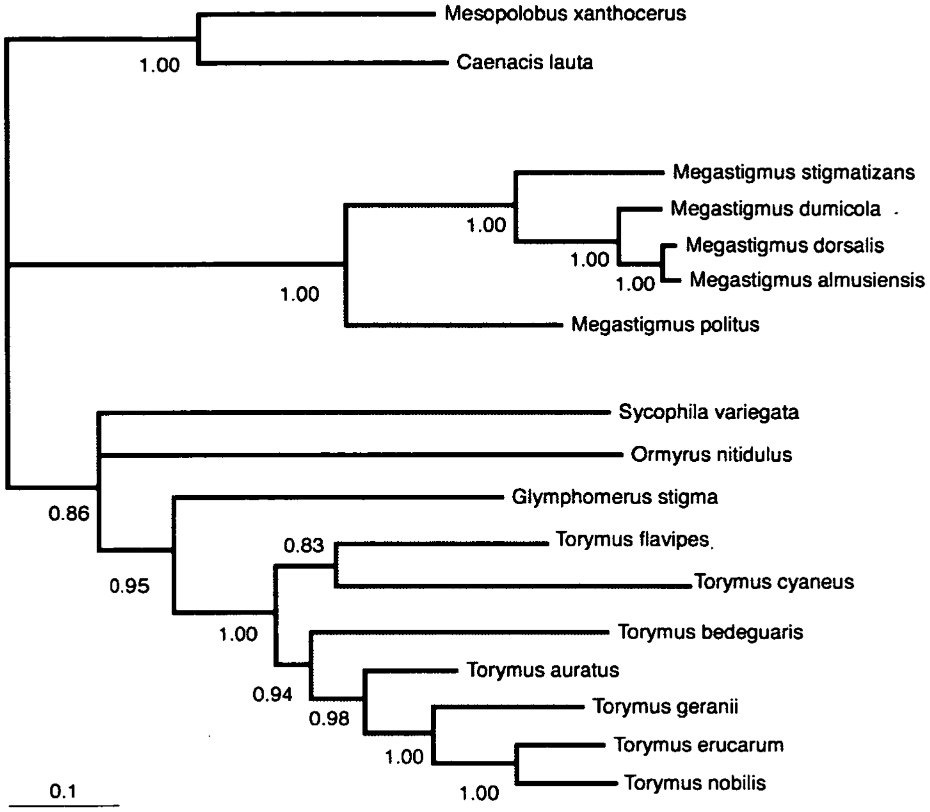


Figure 5.4 Phylogram for the European oak gall attacking Torymid species included here, based on a combined analysis consisting of 28S, cyt*b* and coI gene fragments. Figures beside clades represent Bayesian posterior probabilities.

The implications of the phylogenetic hypotheses in Figures 5.2-5.4 for the questions raised earlier are now discussed in turn below.

1. The relationship between seed feeding and entomophagy in *Megastigmus*

The cyt*b* data strongly support the separation of seed feeding and entomophagous *Megastigmus* into separate sister group clades. Each trophic strategy is thus monophyletic within the genus. While data for *Megastigmus* alone cannot resolve which of these life histories is ancestral, the outgroups within the Chalcidoidea are parasitoids, making this the probable ancestral state. However, such a conclusion requires confirmation by more extensive sampling within the Torymidae. All the European entomophagous *Megastigmus* are included in the tree, while only a subset of the phytophagous species are. The short branches in the phytophagous clade (Fig. 5.2) suggest rapid speciation in this group relative to the entomophagous clade, and this pattern will probably be strengthened by further sampling of phytophagous species.

2. Evidence of repeated evolution of entomophagy with a phytophagous component in the Toryminae

The separation between the two species showing mixed trophic strategies, *Glyphomerus stigma* and *Torymus cyaneus*, in the phylogenies above suggests that entomophagy with a phytophagous component has evolved at least twice in the Toryminae. However, caution is required in interpreting this pattern in terms of the evolutionary dynamics of the family Torymidae: an analysis using COI data for a broader sample of Chalcidoidea (Chapter 6) suggests that *Glyphomerus stigma* is not a member of the Torymidae, but of the Ormyridae. Though once thought to lie within the Torymidae, this view is not supported by my data (see Chapter 6).

3. The relationship between Toryminae attacking hosts on rose and oaks

All of the individual gene trees and the combined analysis support *Glyphomerus* as basal within the Toryminae. Since this parasitoid attacks hosts on roses, this is supported as the basal plant association within the subfamily. The cyt*b* and COI trees and the combined evidence tree show *Torymus bedeguaris*, the only other rose-gall parasitoid in my dataset, to lie within a clade of *Torymus* attacking oak galls, suggesting a second shift from oaks to roses. The 28S tree is agnostic on this second transition, since the relationships in *Torymus* are not resolved. This result must be taken in the context of the possible misplacement of *Glyphomerus* within the Torymidae. If *Glyphomerus* is not a torymid, then the basal association with cynipids in my sample is with galls on oaks.

4. Phylogenetic status of subfamilies in the Torymidae

Both Megastigminae and Toryminae are supported as natural monophyletic groups for my samples, notwithstanding the placement of *Glyphomerus*.

5. Status of species groups within *Torymus*

Taxon sampling limited testing of monophyly to only two of the *Torymus* species groups –

the *T. erucarum*-species group, with 2 sampled species, is supported as monophyletic, while the *T. beduguaris*-group is paraphyletic on current sampling.

5.4 Discussion

Restrictions imposed by incomplete sampling

I had hoped to address broader questions in this Chapter, but analyses were limited by a lack of specimens. My analysis was unable to address the issue of alternative evolutionary pathways from entomophagy to gall induction. However, these are currently being addressed in both Torymidae and Eurytomidae using a larger set of taxa. I was also unable to test the status of the *Torymus cyaneus* species groups within *Torymus.*

Relative diversification of alternative trophic strategies in Megastigmus

It is clear that seed feeding has been associated with a higher rate of speciation (ie adaptive radiation) than entomophagy in *Megastigmus.* This pattern is replicated between the phytophagous Cynipidae and their Figitid parasitiod sister group (Ronquist 1994), and also the Agaonidae and their parasitoid sister group (Machado 2005). In contrast to the parasitoids, most of the seed feeders and gall inducers are host specific or nearly so, illustrating the general pattern that host specificity coupled with high host diversity may drive the evolution of species richness in the next trophic level.

Repeated evolution of partial phytophagy in the Toryminae

Whatever the taxonomic placement of *Glyphomerus*, there have been at least two independent evolutions of partial phytophagy in my set of taxa. These species are important because they illustrate the mixed trophic behaviour required for transition from an entomophagous to a wholly phytophagous diet, and perhaps illustrate the ancestral feeding behaviour for the seed eating clade in *Megastigmus.* No *Megastigmus* with a partially phytophagous diet are yet known, but such species may have existed in the past. This illustrates a weakness of all analyses of character evolution using living taxa: reconstructions of evolutionary changes might be quite diffeerent if we knew about clades that have gone extinct since their divergence within the tree of taxa under study. The rarity of mixed feeding strategies may itself be an artefact of our knowledge of torymid biology. The hosts of many Torymids remain unknown, and even for known parasitoids we do not know if the larval diet is purely carnivorous or partially phytophagous.

Basal host plant associations in Torymidae

The taxa sampled here suggest a basal association with roses in the Toryminae. Were this result robust to sampling a wider set of taxa it would be of interest since it parallels the host plant shift seen in the host cynipids (Ronquist 1994). However, before this conclusion can be supported the position of *Glyphomerus* must be resolved (see Chapter 6). If the more finely resolved picture supported by the cyt*b* and combined analysis trees is true, it suggests that unlike the cynipid hosts the parasitoids can return to roses from oaks.

Taxonomic considerations

The support of Megastigminae and Toryminae as natural monophyletic groups suggests that the morphological characters used to distinguish these subfamilies are reliable. This is of use since it means that further taxa for life-history evolution analyses can be reliably selected at least to the level of genus using current taxonomy. The species-groups of Graham and Gijswijt (1998) are only partially supported by the limited sampling provided here. The *bedeguaris*-group appears to be paraphyletic on the basis of molecular data, but further sampling is required to confirm whether this result is consistent with the addition of further species.

An issue to be raised here is whether *Glyphomerus stigma* is an appropriate taxon to include in this analysis. Although *Glyphomerus* species possess morphological characters that ally them with the Toryminae, and they were kept within the subfamily during a recent taxonomic revision by Grissel (1995), molecular data from Chapter 6 does not support inclusion of *Glyphomerus* within the Torymidae. This raises the general point that phylogenetic relationships within the Chalcidoidea require considerably more research. While analyses of character evolution are possible within strongly supported monophyletic groups (as *Torymus* and *Megastigmus* appear to be), extension of such analyses more broadly into the Torymidae or their putative sister groups in the Chalcidoidea must proceed with extreme caution! The impact of *Glyphomerus* on character reconstruction in *Torymus* indicates the impact that 'rogue' taxa can have.

A final result of note is the location of *Torymus azeureus* within the Toryminae. This species is the only one of the Toryminae taxa included here to attack a non-cynipid host (gall midges on pine). The analysis of cyt*b* sequences places *T. azeurus* with high posterior probability within a clade whose members attack cynipid hosts on oak. This illustrates the value of sampling further taxa on a greater variety of insect hosts and plants. Taken at face value my results suggest that shifts between both insect hosts and plants on very different groups are possible. Furthermore, were other Torymid species to lie within an apparent 'oak cynipid-host clade', monophyly of *Torymus* attacking these hosts would be challenged. I will address this issue in future using samples kindly donated by Dr RR Askew.

Plate 5.1. Western palaearctic torymids in oak cynipid galls, part 1. 1-3 *Megastigmus dorsalis*. 1 and 2 show colour variations in females, 3 is a male. 4,5 female and male *Megastigmus stigmatizans*. 6,7 female and male *Megastigmus synophri*. 8 female *Megastigmus almusiensis*. 9. female *Adontomerus crassipes*. Image 8 is a fresh alcohol-stored specimen.

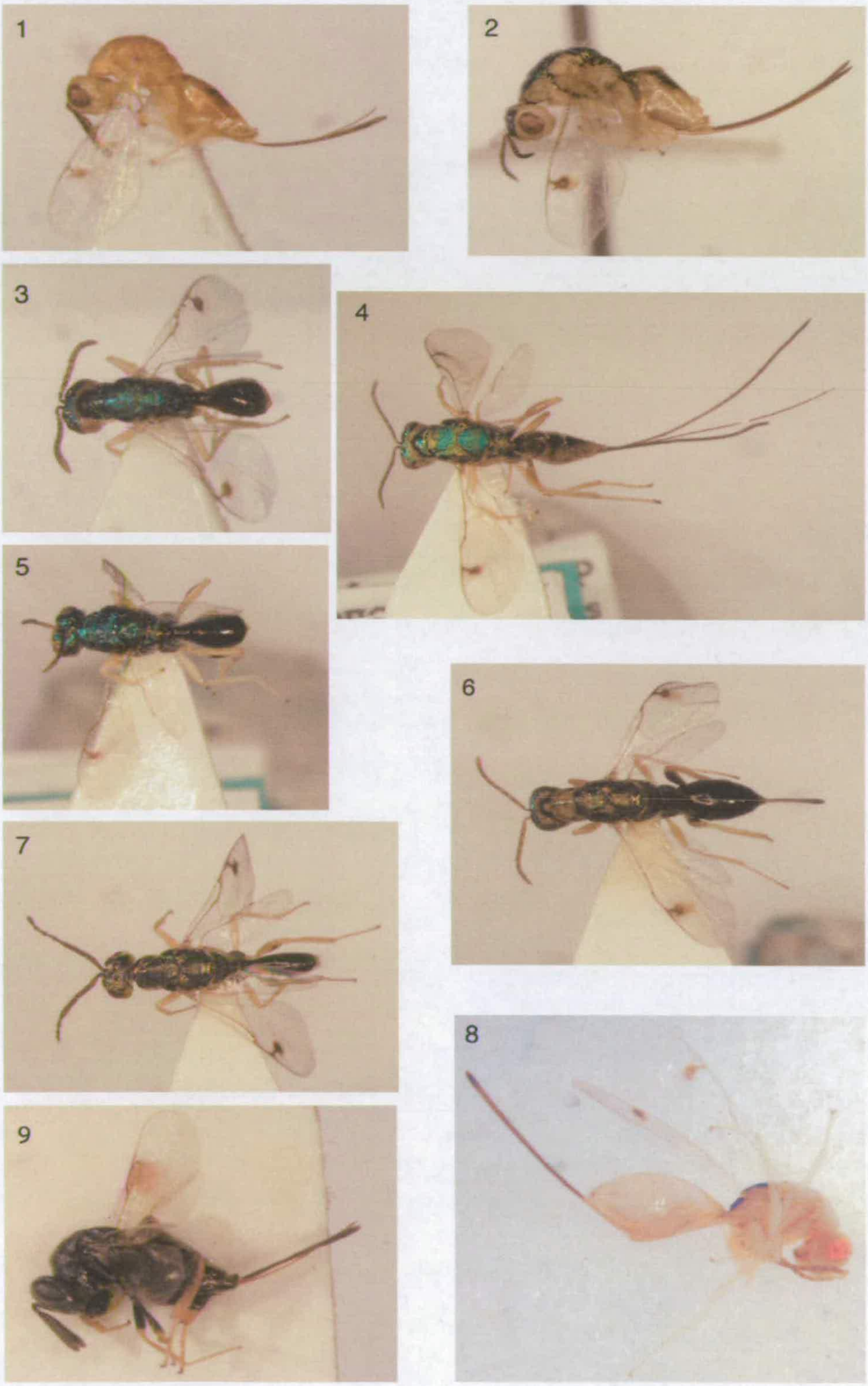


Plate 5.2. Western palaearctic torymids in oak cynipid galls, part 2: 1,2 male and female *Torymus affinis*. 3 male and 4,5,6 female *T. auratus*. 7 male and 8,9 female *T. cyaneus*. Images 2 (inset), 5 and 9 show fresh alcohol-stored material.

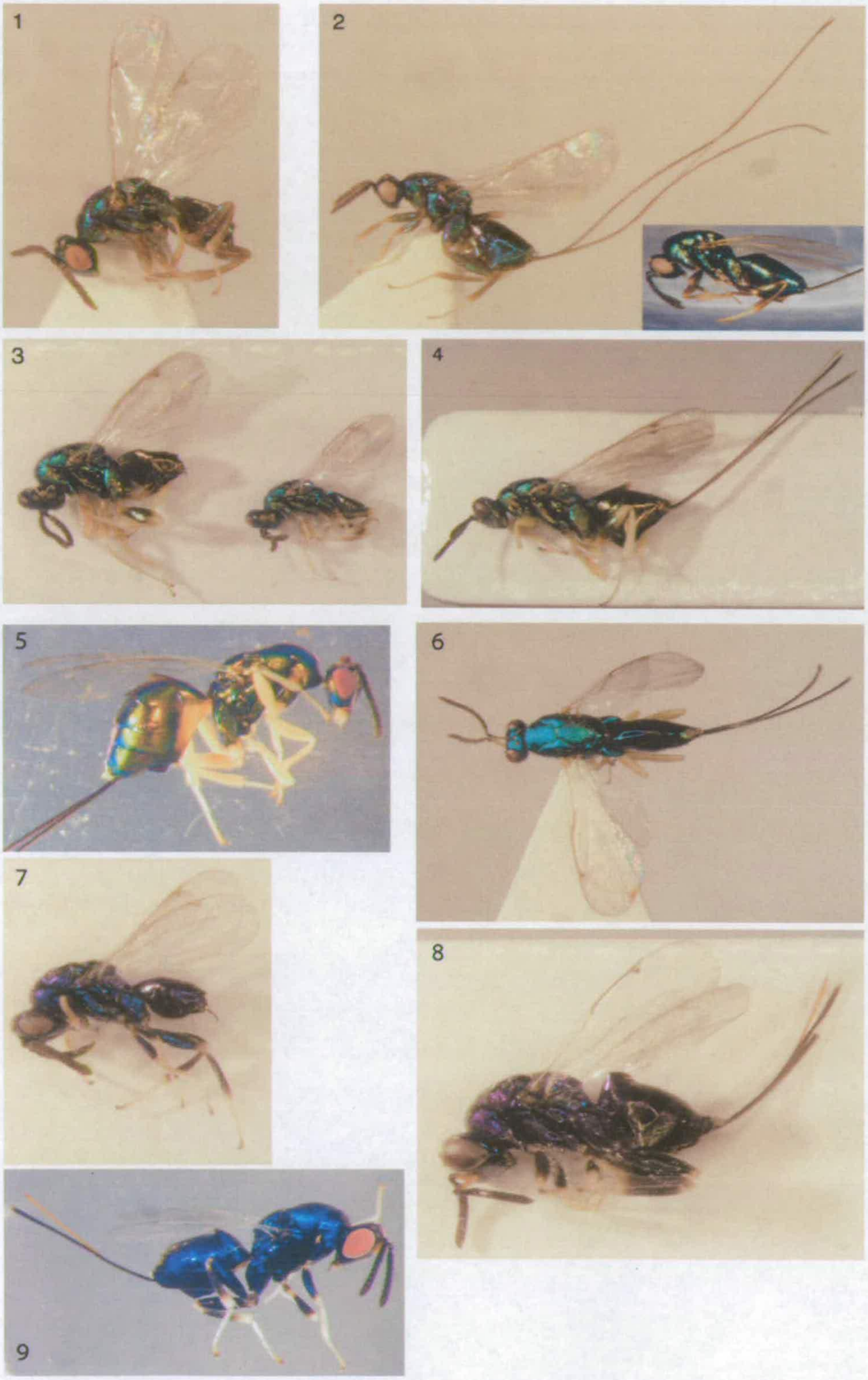
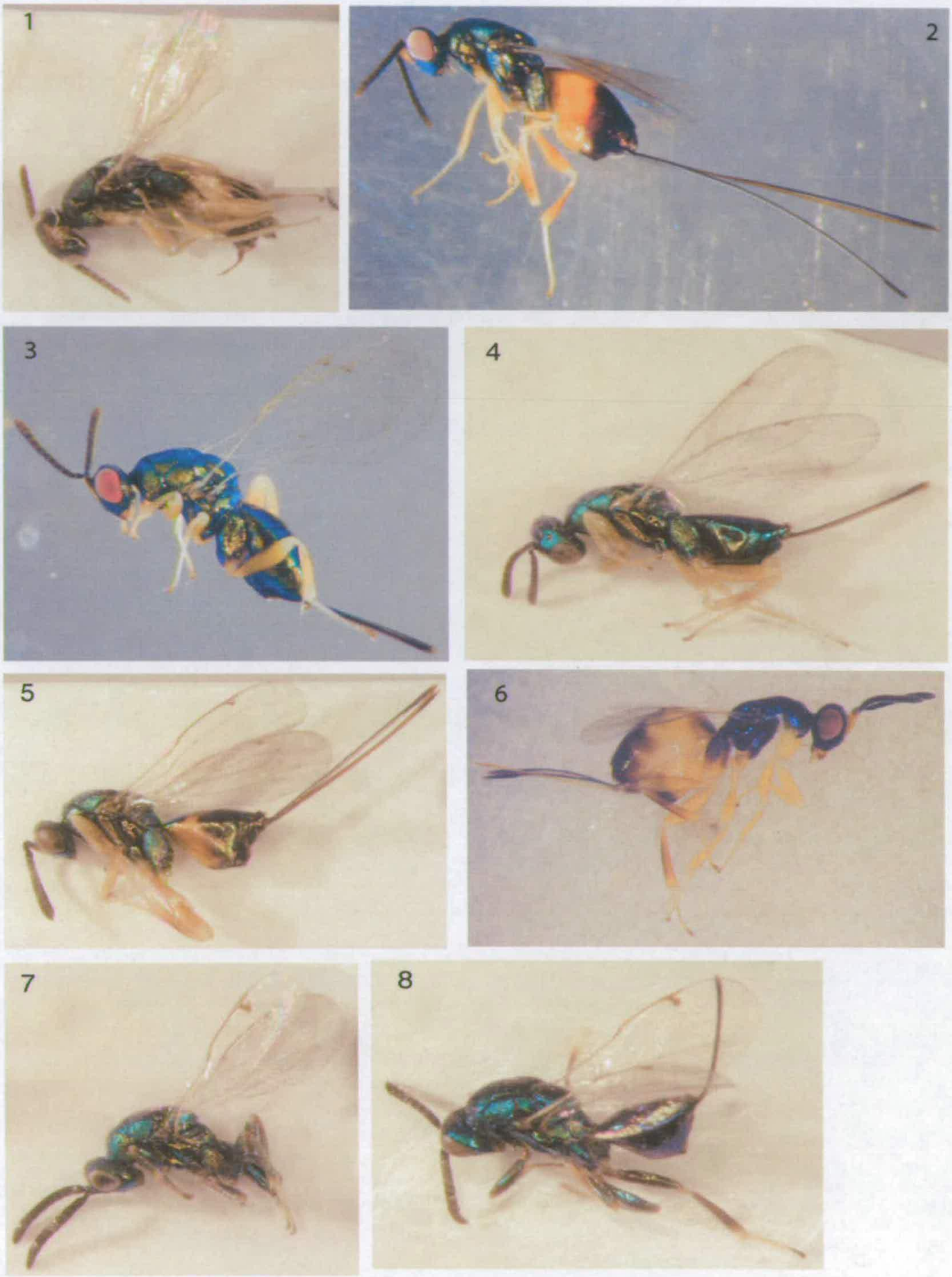


Plate 5.3. Western Palaearctic torymids in oak cynipid galls, part 3: 1,2 male and female *T. erucarum*. 3,4 female *Torymus flavipes*. 5 female *T. geranii*. 6 Female *T. nobilis*. 7,8 male and female *T. notatus*. Images 2,3 and 6 show fresh alcohol-stored material.



Chapter 6

Barcoding of oak gall wasp communities

6.1 Introduction

This Chapter contains an investigation into the utility of COI barcoding in oak gall wasp communities. A discussion of the need for such an approach begins the chapter, followed by a brief review of the current state of molecular taxonomy. Barcoding of a group of 39 oak gall wasp parasitoid species and a group of putative Synerini species then follows in order to evaluate the potential utility of this approach in oak gall wasp communities.

6.2.1 Current impediments to data collection from oak galls

The study of oak gall wasp communities has provided valuable insights into a range of questions in ecology and evolution (Stone *et al* 2002, Csóka *et al* 2005). The system offers many further opportunities for research, and has the potential to become a powerful model for future analyses (Csóka *et al* 2005, Hayward & Stone 2005). However, the efficiency of large-scale ecological data collection from oak galls is currently limited by difficulties relating to the rearing and identification of their inhabitants, particularly the immature stages (eggs and larvae).

After collection, oak galls are placed in rearings for between 1-6 years to allow their inhabitants to develop to adulthood and emerge from the gall. The material obtained is then usually sent to specialist taxonomists for identification before further study can take place. Both the initial time burden whilst gall inhabitants emerge, and the subsequent time taken for identification can delay the onset of scientific study considerably.

An additional problem with rearing is that the adult specimens that emerge from a gall may not fully represent the richness or relative abundance of all the original inhabitants.

Specimens may perish due to imperfect conditions (Stone *el al*1995), and ecological interactions between occupants in a gall mean that individuals may be killed or predated during the rearing period (Shorthouse & Rohfritsch 1992, Stone *et al* 2002). It has not been possible to avoid problems associated with rearing by identifying larvae, since egg and larval taxonomy are generally little studied. Larval morphology is known to be a useful guide to relationships in the cynipid inquilines (Vårdal 2004, Vårdal & Nieves-Aldrey 2005), but the immature stages of the parasitoids are generally little-known. Although a basic key to larvae of some genera and families exists (Askew 1984), there is no key that allows the identification of immature parasitoids to species for the rich communities of southern Europe.

Furthermore, the time required to develop the skills necessary to accurately identify the species common in oak galls is considerable. In Europe the gall formers can generally be identified according to gall morphology, but this is much harder elsewhere and wasps can be very hard to identify even as adults (Stone et al 2002). Identification of parasitoids requires reference to six chalcid families, and identification of many inquiline species and some parasitoids (e.g. the cryptic *Eurytoma* types in Chapter 4) is not currently possible on the basis of morphology alone. Problems may be compounded in the future due to the current trend in the declining number of taxonomists being trained (Hopkins & Freckleton 2002), and by extension of research to new faunistic regions (*eg* the oak cynipid faunas associated with tropical *Cyclobalanopsis* oaks in regions such as Taiwan, with an estimated richness of 40 undescribed cynipid hosts) beyond the experience of existing taxonomists.

6.2.2 The advent of molecular taxonomy

The holy grail of species identification for ecological research would be a means of providing rapid and accurate identifications. A possible solution to the taxonomic impediment associated with traditional approaches is the use of modern molecular techniques to provide a new 'molecular taxonomy' (Godfray 2002, Tautz *et al* 2003, Blaxter 2003).

Part of the interest in molecular taxonomy is due to the realisation that we are still a very long way from cataloguing the true diversity of the planet (May 1988, Blaxter 2004). Estimates of the total number of earth's species typically range from 4 to 100 million, but only around 1.7 million of these have been formally described (Stork 1997). Furthermore, time may be running out to catalogue species diversity in many areas threatened by mass extinctions caused by man. There has also been a realisation that molecular taxonomy may provide the only means to identify many currently neglected taxa (Blaxter *et al* 2004), and also the often highly abundant and probably extremely speciose groups that lack a high degree of morphological differentiation (*eg* nematodes, mites, bacteria) (Hebert *et al* 2003a, Blaxter 2004)

The next section examines the considerations that apply to the selection of a suitable genetic marker for use in molecular taxonomy. The feasibility of one such marker (COI) is then investigated for use in oak gall wasp communities.

6.2.3 Selection of a suitable marker for molecular taxonomy

A molecular marker suitable for the separation of species by molecular taxonomy should possess the following characteristics (Hebert *et al* 2003a, Blaxter 2004a):

1. Universality - the marker should be applicable to as many groups as possible so that comparisons across taxa can be made.
2. Variability - the marker should contain sufficient variation to allow differentiation at the species level.
3. Accessibility - amplification of the marker should be as easy as possible.
4. Ease of use - alignment and analysis of the variability within the marker should be as easy as possible.

Markers so far commonly employed in animal molecular taxonomy include the nuclear small subunit ribosomal RNA gene (ssrDNA), the nuclear large subunit ribosomal gene (lsrDNA), the internal transcribed spacer section of the ribosomal array (ITS), and the mitochondrial cytochome *c* oxidase I gene (COI) (Blaxter 2004a). Of these, the ssrDNA and lsrDNA genes can lack the resolution necessary to differentiate between closely related species (Tautz *et al* 2003, Caterino 2000, and see Chapter 5 for Torymidae). In addition, variable regions within the genes can prove problematic to align (McHugh 1998, Winnepenninckx & Backeljau 1996). ITS sequences can also prove difficult to align due to indel events, and the presence of intra-genomic variation can result in difficulties during sequencing (Wesson *et al.* 1992, Rokas *et al* 2001, 2002). The COI gene also has various problems associated with its usage (discussed below). However, the gene has a number of attributes that mark it out as a good choice for use in molecular taxonomy, and its adoption as a general molecular taxonomic marker has been the focus of a recent and ongoing campaign by Herbert and collaborators (Hebert *et al* 2003a, 2003b, 2004).

6.2.4 The application of COI as a universal DNA barcode

As part of a world-wide effort to catalogue the earth's species, in the 'Barcoding Life' initiative (http://www.barcodinglife.org/), Hebert and collaborators have advocated the adoption of a COI gene fragment for use as a universal species 'barcode'. The term is used as a direct analogy to the Uniform Product Code (UPC) barcodes that feature arrangements of parallel bars and spaces of varying width used to uniquely identify manufactured goods. The Barcoding Life initiative aims to characterise as many of the earth's species as possible for COI in order to aid identification. Planned outputs of the project include a web-accessible database of DNA barcodes, and perhaps one day even a hand-held species identification device. The use of the term 'barcode' has been questioned by some since it implies that each species has a fixed and invariant characteristic (Moritz & Cicero 2004) and an alternative term is the 'molecular operational taxonomic unit' (Floyd *et al* 2002). However, the use of the word barcode is adopted here, since I do not share criticism of the term.

1 32013 70185 0

Figure 6.1 Example of a UPC standardised product barcode. A series of 12 digits and an associated series of bars and spaces coding for the same numbers

```
CAAAATAAATGTTGATATAAAATTGGATCTCCCCCTCCTGCTGGATCAAAAAAAGAT
GTATTTAAATTTCGATCAAACAATAATATAGTAATTGCTCCAGCTAAAACAGGTAAT
GATAATAATAATAAAATAGCTGTTAATAATATAGATCATGAAAATAATGGAATCAAA
TCAATCTTAAACAATTTTATATTTAAAATTGTAGAAATAAAATTAATAGAAGCTATA
ATTGATGAAACCCCAGCAATATGTAAAGAAAAAATAGATAAATCAACTGATGGACCT
CCATGAGATAAATTTGATGATAAAGGAGGATAAACAGTTCAACCTGTTCCAGTACCA
CTACCAATAAATATTCTAGAAATTAATAATATAATTCTCGGAGGAAGAAGTCAGAAA
CTTATATTATTCATTCGAGGAAATGCCATATCAGGAACTCCTATAATTAAAGGAATT
AAAAAATTTCCAAATCCCCCTATTATTACAGGCATAACAAAAAAAAAAATTATAATA
AAAGCATGAGTAGTTACTACTGAATTATAAATTTGATCATTTCCAATTAAAGAACCA
GGATTACCTAATTCTATTCGAATAATTAATCTTAATGATAATCCAATAATACCAGAT
CATATACCAAAAATAAAATACATAATT
```

Figure 6. Example of a COI 'barcode'. A 654 bp fragment of the COI gene for the oak gall wasp parasitoid *Megastigmus stigmatizans*. In effect, the sequence is similar to a conventional

barcode, but with the equivalent of 654 bars, where the variable states (A, T, C, G) at each site convey the information as opposed to bar width.

COI represents a good choice as a barcoding gene satisfying all of the criteria outlined in section 6.2.3 (Hillis *et al* 1996, Avise 2000):

1. Universality - in addition to cyt*b,* COI is the other protein coding mtDNA gene that is present in all eukaryotes.
2. Variability - as part of the mtDNA genome, COI evolves rapidly.
3. Accessibility - a wide range of primers exist for the amplification of COI from a diverse range of animal phyla making it a promising choice for species identification and comparisons across taxa (Folmer et el. 1994).
4. Ease of use - mtDNA genes generally pass unchanged from parent to offspring and contain no structural complications such as introns. In addition, as part of the mitochondrial genome COI is present in high copy number per cell. Typically, each cell only contains 2 copies of nuclear DNA sequences, but 100-10,000 mitochondrial genomes. This feature aids DNA recovery and PCR amplification, and is particularly useful when dealing with degraded specimens or small amounts of tissue.

A fragment of approximately 645 bp (the exact length varies depending on species) of the COI gene has been adopted as the standard barcoding fragment. As opposed to UPC barcodes, which use a decimal code, genetic material is based on a quaternary code, but in theory this still provides enormous potential variability. For example, a sequence only 15 nucleotides long with 4 character alternatives at each position has over one billion unique combinations. However, nucleotides are not random since they code for something of biological meaning (which eliminates some possible sequences) and share common ancestry (meaning that not all of ‘molecular barcode space is sampled by living organisms). A larger fragment is thus required. The 645 bp fragment represents a length of sequence that is relatively short and easy to work with, but should contain enough information for discrimination between closely related species.

Hebert *et al* (2003b) demonstrated that COI shows enough sequence divergence to separate a large number of animal species with a comparison of 13,000 pairs of congeneric species from 11 phyla using COI sequences. The results of the study indicated a mean sequence divergence of 11.3% between species pairs (equivalent to about 73 diagnostic substitutions for a 645 bp length of DNA). With the exception of the Cnidaria, 98% of the species pairs showed more than 2% divergence (13 diagnostic substitutions for 645 bp), and most pairs (79%) showed greater than 8% sequence divergence. The Cnidaria displayed far lower divergence with 94.1% of species possessing less than 2% sequence divergence, and the authors suggest that this result is probably due to the existence of an excision-repair system in the Cnidaria that is absent in other animal mitochondria.

The real test of the efficiency of a barcoding gene is its ability to discriminate between closely related taxa. Hebert conducted a study (Hebert *et al* 2004) to test the correspondence between species boundaries indicated by COI barcodes in a closely related group of species and those established by prior taxonomic research. A group of 260 North American bird species were chosen as the subject of the study, since they represent one of the largest and best-studied vertebrate groups and as a result morphological taxonomy is well established. All the species included were found to have different COI barcodes and differences between species (average 7.93%) were found to be 18 times greater in magnitude than those within species (0.43%). On the basis of this result, the authors argued that since patterns of intraspecific and interspecific variation in COI appear to be similar in various animal groups, a "standard screening threshold" of ten times the average intraspecific difference could be adopted in order to identify genetically divergent taxa as provisional species. The authors applied this threshold to the species examined in their study to give a 2.7% threshold for North American bird species (using the within species divergence average of 0.27% found, excluding four species that contained deep divergences within them, possibly indicating the presence of multiple species). Such a threshold would recognize over 90% of the 260 species.

However, the use of COI does not come without potential problems or critics. Mallet & Willmot (2003) object that DNA sequence differences between closely related species will often be too small to allow their discrimination, and that horizontal transfer between divergent lineages will confuse the issue. Indeed, cases of mitochondrial transfer by hybridisation (introgression) have been detected between closely related animal species (*eg* Glemet *et al* 1998), including gall wasps (Rokas *et al.* 2003 a,b). In addition, pseudogenes – nonfunctional copies of mtDNA genes usually containing multiple mutations and/or deletions can be a problem, and have been proven to exist in gall wasps (Rokas *et al.* 2003 a, b). This is a problem since the DNA under consideration must be known to be orthologous between species, since paralogues will define gene groups rather than species groups (Floyd *et al*

2002). The extent to which pseudogenes and introgression are present in oak gall wasp parasitoids and inquilines is unknown, but neither have yet been detected.

In addition, Moritz & Cicero (2004) also warn of the potential limitations of using mtDNA to infer species boundaries due to the retention of ancestral polymorphism, male-biased gene flow, and selection on any mtDNA nucleotide (as a result of the entire genome comprising one linkage group). These concerns are reiterated by Jiggins & Hurst (2005). Moritz & Cicero (2004) also warn of the potential pit-falls in using thresholds of intraspecific difference to infer species. Moritz & Cicero point out that previous studies that examined mtDNA genetic divergence in 39 sister species of north American birds revealed an average of 1.9% sequence divergence between species (Johnson & Cicero 2004), and that 29 pairs were at or below the 2.7% threshold suggested by Hebert *et al.* Thus, they would not be recognised as different species despite biological differences. Furthermore, paraphyly in mtDNA haplotypes among morphologically recognised species were also uncovered in these studies.

6.3 Application of COI barcoding to the oak gall wasp system

Due to the room for improvement in species identification in oak gall wasp communities and the suitable characteristics of COI described above (particularly its ease of use and potential ability to identify species based on small amounts of specimen material), a pilot study was undertaken to investigate its potential in this system. The application and results of this study are provided below, and a brief review of how the method might be employed to improve data collection from oak galls is provided in the discussion.

6.4 Methods

Specimen identification

All parasitoid specimens used in this Chapter were identified by taxonomists (Drs G. Melika & C. Thúroczy) at the Systematic Parasitoid Laboratory, Koszeg, Hungary. I also checked all specimens using the key constructed by Askew & Thúroczy (2005). Inquiline material was identified by Dr G Melika and Dr JL Nieves-Aldrey (Museo Nacional de Ciencias Naturales, Madrid). Specimens included in this study mainly originate from Hungary, but further geographic sampling will be necessary to generate a reliably useable set of European or western Palaearctic barcodes.

Most of the parasitoid species are well defined morphological species, although the extent to which groups of morphologically similar sibling species exist within current species is unknown (such as for *Eurytoma brunniventris*, Chapter 4). However, by comparison, the taxonomy of the *Synergini* is highly unstable and the specimens included here are part of a large ongoing study utilising both morphological and molecular markers to help resolve the issue. A list of the parasitoid species included in this study and their sampling locations is provided in Table 6.1 below, and for the inquiline species in Table 6.2. In many cases, the inquiline species are listed as putative members of multispecies complexes reflecting their current unknown status.

Data generation

DNA extraction, PCR amplification and sequencing were carried out as described in Chapter 2. Wherever possible two members of each species were included so that between species and within species sequence divergence could be compared.

Trees were constructed using the neighbour-joining approach and sequence divergence is expressed as uncorrected distances. The data are presented in this way since uncorrected distances most clearly express the actual physical differences between sequences, which is the critical factor when applying means of discriminating species on the basis of distinctive barcodes.

Table 6.1 Parasitoid species for which COI barcodes were obtained. 'Spec No.' refers to the specimen IDs used in this study.

Spec No.	Parasitoid species	Family	Host gall species	site	Country	Latitude
1, 2	*Mesopolobus xanthocerus*	Pteromalidae	*Andricus quercusramuli*	Matrafured	Hungary	47° 49' 60N
4	*Mesopolobus fuscipes*	Pteromalidae	*Cerris* bud gall 0	Szentkut	Hungary	47° 58' 60N
5	*Mesopolobus fuscipes*	Pteromalidae	*Cerris* bud gall 1	Matrafured	Hungary	47° 49' 60N
7, 8	*Mesopolobus amaenus*	Pteromalidae	*Andricus multiplicatus*	Matrafured	Hungary	47° 49' 60N
10	*Mesopolobus tibialis*	Pteromalidae	*Andricus singularis*	Godollo	Hungary	47°36'0"N
12	*Mesopolobus tibialis*	Pteromalidae	*Cynips longiventris*	Godollo	Hungary	47°36'0"N
13, 14	*Cecidostiba fungosa*	Pteromalidae	*Andricus hungaricus*	Godollo	Hungary	47°36'0"N
16, 17	*Cecidostiba semifascia*	Pteromalidae	*Biorhiza pallida*	Szentkut	Hungary	47° 58' 60N
18	*Caenacis lauta*	Pteromalidae	*Andricus lucidus*	Matrafured	Hungary	47° 49' 60N
19	*Caenacis lauta*	Pteromalidae	*Andricus kollari*	Godollo	Hungary	47°36'0"N
21	*Hobbya stenonota*	Pteromalidae	*Biorhiza pallida*	Godollo	Hungary	47°36'0"N
24	*Sycophila variegata*	Eurytomidae	*Andricus amblycerus*	Matrafured	Hungary	47° 49' 60N
26	*Sycophila variegata*	Eurytomidae	*Andricus caliciformis*	Matrafured	Hungary	47° 49' 60N
27	*Sycophila biguttata*	Eurytomidae	*Andricus mayri*	Godollo	Hungary	47°36'0"N
29	*Sycophila biguttata*	Eurytomidae	*Andricus coriarius*	Szentkut	Hungary	47° 58' 60N
30	*Eurytoma pistaciae*	Eurytomidae	*Andricus solitarius*	Matrafured	Hungary	47° 49' 60N
31	*Eurytoma pistaciae*	Eurytomidae	*Andricus mayri*	Godollo	Hungary	47°36'0"N
ebfi	*Eurytoma bruniventris*	Eurytomidae	*Trigonaspis synapis*	Varpalota	Hungary	47° 12' 0N
ebsi	*Eurytoma bruniventris*	Eurytomidae	*Cynips quercus*	Sopron	Hungary	47° 40' 60N
32	*Eupelmus annulatus*	Eupelmidae	*Andricus kollari*	Godollo	Hungary	47°36'0"N
33	*Eupelmus annulatus*	Eupelmidae	*Biorhiza pallida*	Matrafured	Hungary	47° 49' 60N
34	*Eupelmus urozonus*	Eupelmidae	*Cynips disticha*	Godollo	Hungary	47°36'0"N
35	*Eupelmus urozonus*	Eupelmidae	*Biorhiza pallida*	Szentkut	Hungary	47° 58' 60N
37	*Ormyrus nitidulus*	Ormyridae	*Andricus kollari*	Godollo	Hungary	47°36'0"N
/	*Ormyrus nitidulus*	Ormyridae	*Andricus kollari*	Godollo	Hungary	47°36'0"N
40	*Ormyrus pomaceus*	Ormyridae	*Andricus corruptrix*	Godollo	Hungary	47°36'0"N
/	*Ormyrus pomaceus*	Ormyridae	*Andricus mayri*	Godollo	Hungary	47°36'0"N
43, 44	*Minotetrastichus frontalis*	Eulophidae	*Neuroterus numismalis*	Matrafured	Hungary	47° 49' 60N
46	*Closterocerus trifasciatus*	Eulophidae	*Neuroterus numismalis*	Matrafured	Hungary	47° 49' 60N
48, 49	*Aprostocetus cerricola*	Eulophidae	*Andricus multiplicatus*	Matrafured	Hungary	47° 49' 60N
51, 52	*Aulogymnus gallarum*	Eulophidae	*Andricus quercusramuli*	Matrafured	Hungary	47° 49' 60N
54	*Aulogymnus testaceoviridis*	Eulophidae	*Andricus singularis*	Matrafured	Hungary	47° 49' 60N
55	*Aulogymnus testaceoviridis*	Eulophidae	*Andricus quercusramuli*	Matrafured	Hungary	47° 49' 60N
56	*Aulogymnus arsames*	Eulophidae	*Andricus schroeckingeri*	Sopron	Hungary	47° 40' 60N
57	*Aulogymnus arsames*	Eulophidae	*Andricus multiplicatus*	Sopron	Hungary	47° 40' 60N
58	*Aulogymnus trilineatus*	Eulophidae	*Andricus quercustozae*	Varpalota	Hungary	47° 12' 0N
59	*Aulogymnus trilineatus*	Eulophidae	*Andricus coriarius*	Sopron	Hungary	47° 40' 60N
60	*Baryscapus pallidae*	Eulophidae	*Biorhiza pallida*	Matrafured	Hungary	47° 49' 60N
61	*Baryscapus pallidae*	Eulophidae	*Cynips cornifex*	Szentkut	Hungary	47° 58' 60N
63, 64	*Baryscapus diaphantus*	Eulophidae	*Biorhiza pallida*	Godollo	Hungary	47°36'0"N
66	*Baryscapus anassilus*	Eulophidae	*Biorhiza pallida*	Godollo	Hungary	47°36'0"N
67	*Baryscapus anassilus*	Eulophidae	*Biorhiza pallida*	Godollo	Hungary	47°36'0"N
21, 22	*Glyphomerus stigma*	Torymidae	*Diplolepsis rosae*	Anatolikon	Greece	40° 39' 47N
3, 4	*Megastigmus almusiensis*	Torymidae	*Andricus seckendorfi*	Madenli	Turkey	40° 53' 60N
12	*Megastigmus dorsalis*	Torymidae	*Andricus hungaricus*	Godollo	Hungary	47°36'0"N
19	*Megastigmus dorsalis*	Torymidae	*Andricus kollari*	Teignmouth	Britain	50° 32' 60N
1, 2	*Megastigmus politus*	Torymidae	*Synophrus politus*	Pisodherion	Greece	40° 46' 60N
80	*Megastigmus stigmatizans*	Torymidae	*Andricus kollari*	Chiusi	Italy	43°1'0"N
95	*Megastigmus stigmatizans*	Torymidae	*Andricus quercustozae*	Sobreiro de Cima	Portugal	41° 51' 0N
7, 8	*Torymus auratus*	Torymidae	*Cynips quercusfolii*	Koszeg	Hungary	47° 22' 60N
18, 19	*Torymus beduguaris*	Torymidae	*Diplolepsis rosae*	Laroles	Spain	37° 1' 0N
1, 2	*Torymus cyaneus*	Torymidae	*Cynips quercus*	Szentkut	Hungary	47° 58' 60N
13	*Torymus erucarum*	Torymidae	*Cynips spp 1*	Godollo	Hungary	47°36'0"N
14	*Torymus erucarum*	Torymidae	*Andricus kollari*	Godollo	Hungary	47°36'0"N
4	*Torymus flavipes*	Torymidae	*Neuroterus quercusbaccarum*	Godollo	Hungary	47°36'0"N
5	*Torymus flavipes*	Torymidae	*Andricus curvator*	Matrafured	Hungary	47° 49' 60N
10, 11	*Torymus geranii*	Torymidae	*Cynips longiventris*	Godollo	Hungary	47°36'0"N
16, 17	*Torymus nobilis*	Torymidae	*Biorhiza pallida*	Miraflores. Madrid	Spain	40° 49' 0N

Table 6.2 Inquiline species for which COI barcodes were obtained. 'Spec No.' refers to the specimen IDs used in this study. Use of the symbol ' / ' indicates that exact locality is unknown for that specimen.

Spec No.	Synergini species	Host gall species	Host oak species	Site name	Country	Latitude	Longitude
1	*Synergus umbraculus*	*Andricus gallaetinctoriae*	*petraea*	Sopron	Hungary	47° 40' 60N	16° 36' 0E
2	*Synergus umbraculus*	*Andricus lucidus*	*pubescens*	Mátrafüred	Hungary	47° 49' 60N	19° 58' 0E
6	*Synergus hayneanus/reinhardi*	*Andricus quercustozae*	*pubescens*	Várpalota	Hungary	47° 12' 0N	18° 7' 60E
7	*Synergus hayneanus/reinhardi*	*Andricus caputmedusae*	*pubescens*	Szentkút	Hungary	47° 49' 60N	19° 58' 0E
8	*Synergus hayneanus*	*Andricus gallaetinctoriae*	*petraea*	Szentkút	Hungary	47° 49' 60N	19° 58' 0E
9	*Synergus hayneanus*	*Andricus quercustozae*	*pubesc.*	Várpalota	Hungary	47° 12' 0N	18° 7' 60E
10	*Synergus umbraculus/hayneanus*	*Andricus coriarius*	*pubescens*	Szentkút	Hungary	47° 49' 60N	19° 58' 0E
11	*Synergus umbraculus/hayneanus*	*Andricus lucidus*	*robur*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
12	*Synergus pallidipennis*	*Andricus hungaricus*	*robur*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
13	*Synergus pallidipennis*	*Andricus conificus*	*pubescens*	Várpalota	Hungary	47° 12' 0N	18° 7' 60E
14	*Synergus gallaepomiformis*	*Andricus hungaricus*	*robur*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
15	*Synergus gallaepomiformis*	*Andricus gallaetinctoriae*	*robur*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
16	*Synergus gallaepomiformis/pallicornis*	*Andricus hungaricus*	*robur*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
17	*Synergus gallaepomiformis/pallicornis*	*Andricus coriarius*	*pubescens*	Sopron	Hungary	47° 40' 60N	16° 36' 0E
18	*Synergus pallicornis/gallaepomiformis*	*Andricus hungaricus*	*robur*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
19	*Synergus pallicornis/gallaepomiformis*	*Andricus gallaetinctoriae*	*petraea*	Mátrafüred	Hungary	47° 49' 60N	19° 58' 0E
20	*Synergus pallicornis*	*Andricus gallaetinctoriae*	*petraea*	Szentkút	Hungary	47° 49' 60N	19° 58' 0E
21	*Synergus pallicornis*	*Andricus caputmedusae*	*robur*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
22	*Synergus pallicornis/nervosus-albipes*	*Andricus gallaetinctoriae*	*robur*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
23	*Synergus pallicornis/nervosus-albipes*	*Andricus coriarius*	*pubescens*	Sopron	Hungary	47° 40' 60N	16° 36' 0E
24	*Synergus gallaepomiformis/albipes*	*Andricus coriarius*	*pubescens*	Sopron	Hungary	47° 40' 60N	16° 36' 0E
25	*Synergus gallaepomiformis/albipes*	*Andricus coriarius*	*pubescens*	Sopron	Hungary	47° 40' 60N	16° 36' 0E
26	*Synergus nevosus/albipes*	*Andricus coriarius*	*pubescens*	Sopron	Hungary	47° 40' 60N	16° 36' 0E
27	*Synergus nevosus/albipes*	*Andricus hungaricus*	*cerris*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
28	*Synergus physocerus*	*Trigonaspis synaspis*	*robur*	Várpalota	Hungary	47° 12' 0N	18° 7' 60E
29	*Synergus physocerus*	*Trigonaspis synaspis*	*robur*	Várpalota	Hungary	47° 12' 0N	18° 7' 60E
30	*Synergus diaphanus*	*Andricus conificus*	*pubescens*	Várpalota	Hungary	47° 12' 0N	18° 7' 60E
31	*Synergus diaphanus*	*Andricus conificus*	*pubescens*	Várpalota	Hungary	47° 12' 0N	18° 7' 60E
32	*Synophrus politus*	*Synophrus politus*	*cerris*	Sopron	Hungary	47° 40' 60N	16° 36' 0E
33	*Synophrus politus*	*Synophrus politus*	*cerris*	Sopron	Hungary	47° 40' 60N	16° 36' 0E
34	*Ceroptres clavicornis*	*Andricus conglomeratus*	*robur*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
35	*Ceroptres clavicornis*	*Andricus lignicolus*	*petraea*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
36	*Ceroptres cerri*	*Neuroterus macropterus*	*cerris*	Sopron	Hungary	47° 40' 60N	16° 36' 0E
37	*Ceroptres cerri*	*cerris bud gall 1*	*cerris*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
38	*Synergus flavipes*	*Neuroterus lanuginosus*	*cerris*	Mátrafüred	Hungary	47° 49' 60N	19° 58' 0E
39	*Synergus flavipes*	*Neuroterus saliens*	*cerris*	Szentkút	Hungary	47° 49' 60N	19° 58' 0E
40	*Synergus apicalis/rotundiventris*	*Andricus qramuli*	*petraea*	Mátrafüred	Hungary	47° 49' 60N	19° 58' 0E
41	*Synergus apicalis/rotundiventris*	*Andricus gallaetinctoriae*	*robur*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
42	*Synergus thaumacerus*	*Andricus grossulariae*	*cerris*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
43	*Synergus thaumacerus*	*Chilaspis nitida*	*cerris*	Sopron	Hungary	47° 40' 60N	16° 36' 0E
44	*Synergus consobrinus*	*Andricus grossulariae*	*cerris*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
45	*Synergus consobrinus*	*Andricus grossulariae*	*cerris*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
46	*Saphonecrus undulatus*	*aphelonyx cerricola*	*cerris*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
47	*Saphonecrus undulatus*	*aphelonyx cerricola*	*cerris*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
48	*Saphonecrus haymi*	*Andricus testaceipes*	*petraea*	Mátrafüred	Hungary	47° 49' 60N	19° 58' 0E
49	*Saphonecrus haymi*	*Chilaspis nitida*	*cerris*	Sopron	Hungary	47° 40' 60N	16° 36' 0E
50	*Saphonecrus connatus*	*Andricus testaceipes*	*petraea*	Mátrafüred	Hungary	47° 49' 60N	19° 58' 0E
51	*Synergus apicalis/rotundiventris*	*Andricus quercusramuli*	*petraea*	Mátrafüred	Hungary	47° 49' 60N	19° 58' 0E
52	*Synergus apicalis/rotundiventris*	*Andricus testaceipes*	*petraea*	Mátrafüred	Hungary	47° 49' 60N	19° 58' 0E
53	*Synergus thaumacerus*	*Neuroterus saliens*	*cerris*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
54	*Synergus thaumacerus*	*Neuroterus saliens*	*cerris*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
55	*Synergus consobrinus*	*Andricus grossulariae*	*cerris*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
56	*Synergus consobrinus*	*Andricus grossulariae*	*cerris*	Gödöllő	Hungary	47°36'0"N	19°19'48"E
57	*Synergus clandestinus*	*? stunted acorns*	*pyrenaica*	El Ventorrillo	Spain	/	/
58	*Synergus clandestinus*	*? stunted acorns*	*pyrenaica*	El Ventorrillo	Spain	/	/
59	*Synergus incrassatus*	*A. quercusradicis agm.*	*pyrenaica*	El Escorial	Spain	40°34'60"N	4°9'0"W
60	*Synergus physoceras*	*T. synaspis agm.*	*faginea*	Valgallega	Spain	/	/
61	*Synergus physoceras*	*T. synaspis agm.*	*faginea*	Valgallega	Spain	/	/
62	*Synergus nervosus*	*A. quadrilineatus*	*petraea*	Dehesa de Somosierra	Spain	/	/
63	*Synergus nervosus*	*A. quadrilineatus*	*petraea*	Dehesa de Somosierra	Spain	/	/
64	*Synergus plagiotrochi*	*P. australis sex.*	*ilex*	Dehesa de Navacerrada	Spain	/	/
65	*Synergus plagiotrochi*	*P. australis sex.*	*ilex*	Dehesa de Navacerrada	Spain	/	/
66	*Saphonecrus lusitanicus*	*Plagiotrochus sp.*	*ilex*	Salamanca Cabrerizos	Spain	/	/
67	*Saphonecrus lusitanicus*	*Plagiotrochus sp.*	*ilex*	Salamanca Cabrerizos	Spain	/	/
68	*Saphonecrus barbotini*	*Plagiotrochus britaniae*	*coccifera*	Matadepera	Spain	/	/
69	*Synergus crassicornis*	*P. burnayi*	*ilex*	/	/	/	/
70	*Synergus crassicornis*	*P. burnayi*	*ilex*	/	/	/	/
71	*Periclistus brandti*	*Diplolepis rosae*	Rose sp.	/	/	/	/
72	*Periclistus brandti*	*Diplolepis rosae*	Rose sp.	/	/	/	/

6.5 Results

Parasitoids

The average sequence divergence within species was found to be 1.58%. Therefore, if Hebert *et al*'s standard screening threshold of ten times the average intraspecific difference for separating species were to be applied to oak gall wasp parasitoids a threshold of 15.8% would be obtained. Within-species divergence showed considerable variation, ranging from 0-11.16%. Values at the upper end of this spectrum are extreme given commonly reported intraspecific divergences for mtDNA, which are rarely greater than 2% and usually less than 1% (Avise 2000). Species for which a within species distance of greater than 3% was found are listed in Table 6.3.

Table 6.3 Percentage within species divergence for parasitoid species that displayed a divergence of greater than 3%.

Parasitoid species	% within species distance
Baryscapus.pallidae	11.16
Ormyrus pomaceus	11.01
Auglogymnus trilineatus	4.43
Eurytoma brunniventris	3.95
Eupelmus annulatus	3.67
Auglogymnus arsames	3.06

Subsequent to the inclusion of specimens in this analysis, Dr RR Askew has informed the gall research group that specimens identified as *Baryscapus pallidae* can very easily be confused with a second species, *B. berhidanus*. Therefore, given the intraspecifc distance given for *B. pallidae* above, it seems likely that individuals from two different species of *Baryscapus* were included in this study. A similar explanation is probably responsible for the high within species divergence obtained for *Ormyrus pomaceus* and *Eurytoma brunniventris*. No closely-related species to *O. pomaceus* (other than *O. nitidulus*, from which both *O. pomaceus* specimens are clearly separated) is currently known. *A* study of European *Ormyrus* species is currently underway by Maria Hernandez, a student of José-Luis Nieves-Aldrey in Madrid. My data suggest the existence of cryptic sibling species. As described in Chapter 4, a similar situation is apparent for *E. brunniventris*, for which cyt*b* sequence data suggest the existence of at least four different species. Unfortunately, the results of Chapter 4 came too late to include members of each *E. brunniventris* 'Type' in the analyses for this chapter. However, I intend to continue the investigation of this issue in *E. brunniventris* using COI in the future. The high distances for *Auglogymnus trilineatus, A. arsames* and *Eupelmus annulatus* represent new findings and no explanation other than either the existence of a similar situation for these species, or the possibility of misidentifications can be given at this stage. *Eupelmus* is a genus which, like *Eurytoma*, includes extreme generalists (particularly *E. urozonus*), and cryptic taxa are thus perhaps likely. All the other parasitoid species had within species distances of approximately 2% or less, conforming to Avise's guidelines stated above.

The average between-species divergence observed for parasitoids was 15.19%, ranging from 0.31% (between *Auglogymnus testaceoviridus* and *A. gallarum*) to 21.41% (between *Ormyrus pomaceus* and *Mesopolobus tibialis*). The mean is very close to the 15.8% threshold for discrimination between species proposed by Hebert *et al.* and at first glance supports the application of the threshold. However, with further consideration it becomes apparent that the threshold is over optimistic. This study includes a relatively small number of species that originate from six different chalcid families, and the average evolutionary distance between most species must be expected to be relatively high. A more accurate estimate can only be obtained by the inclusion of more closely related species in the analysis. Whilst Hebert *et al.*'s standard screening threshold might be a useful guideline for identifying highly divergent taxa, distinguishing closely related species of chalcid wasp probably requires a reduction in the threshold value.

The COI phylogeny in Figure 6.1 indicates good pairing of conspecific parasitoids. Excluding *B. pallidae* for the reasons explained above (likelihood of the inclusion of two species), there is no evidence of introgression between different parasitoid species. Family groupings are also consistent, except for *Glyphomerus stigma* (Torymidae), which lies within the Ormyridae. As discussed in Chapter 5, it is possible that *G. stigma* does not belong in the Torymidae; further analysis is required before this issue can be clarified. With the exception of the Eulophidae, the remaining chalcid families are monophyletic. The Eulophidae are paraphyletic, with a split occurring between the Eulophinae (*Auglogymnus sp* in Figure 6.1) and the Tetrastichinae & Euledoninae (*Barysacpus, Minotetrastichus, Aprostocetus* and *Closterocerus sp* in Figure 6.1). However, COI is generally not regarded as a suitable gene for inferring deeper family level phylogenetic relationships due to its rapid rate of evolution, which leads to saturation of sequence data at higher taxonomic levels.

In summary, given the pairing of species visible in Figure 6.1 and the relatively high average between species distance, COI barcoding identification of known parasitoid species looks promising for oak gall wasp communities. However, the use of Hebert *et al*'s standard screening threshold is not supported as useful for discrimination between closely related species.

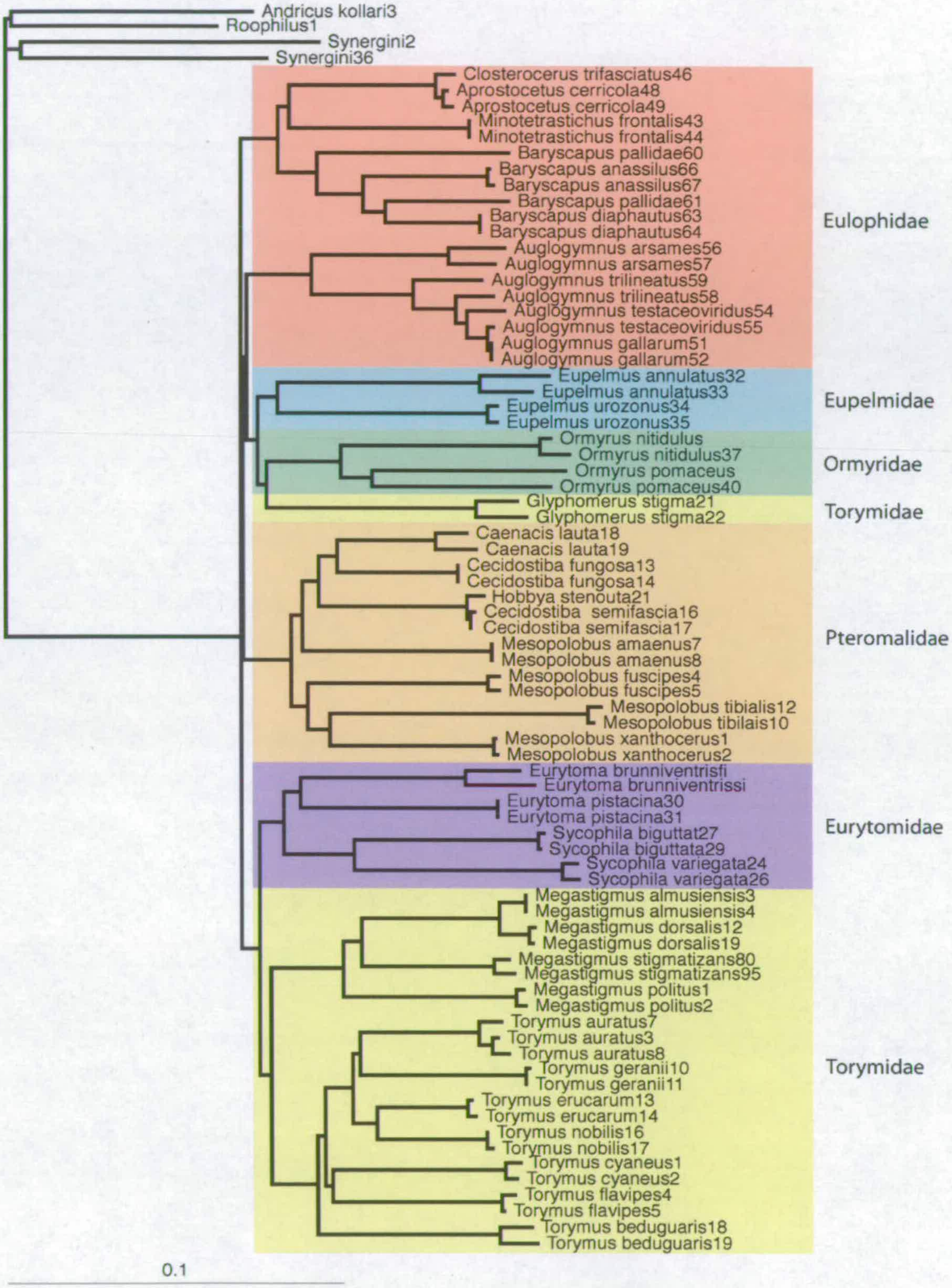


Figure 6.1 Neighbour-joining tree for a 654 bp fragment of the COI gene of the parasitoid species included in this study. Colours indicate parasitoid families: red - Eulophidae, blue - Eupelmidae, Green - Ormyridae, Yellow - Torymidae, Orange - Pteromalidae, Purple - Eurytomidae.

Inquilines

Figure 6.2 shows that COI sequences data indicate that significantly more work is needed on this group. Generally, it appears that whilst specimens of species in the genera *Saphronecrus, Synophrus* and *Ceroptres* are resolved as sister group pairs (with the exception of *Saphronecrus haymi*), specimens of some currently recognised *Synergus* species are not resolved into monophyletic lineages, while others (such as *S. hayneanus*) are so grouped. If the relationships in Figure 6.2 are taken at face value, the status of a number of taxa needs to be reassessed. Suggested reallocations of taxa supplied to be to molecularly-supported taxa are suggested in Table 6.4 below.

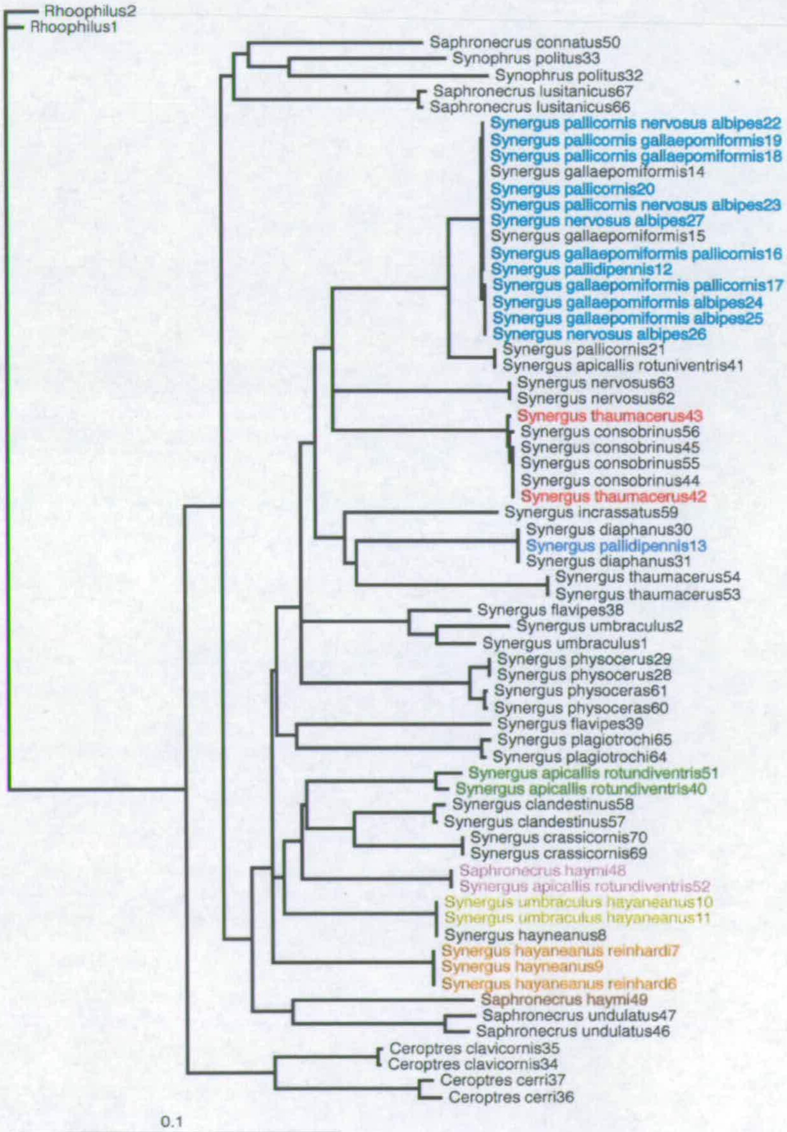


Figure 6.2 Neighbour-joining tree for a 654 bp fragment of the COI gene of the inquiline cynipid species included in this study. Colours correspond to modifications made on the basis of COI data, listed in Table 6.4.

Table 6.4 Preliminary modifications to *Synergus* species on the basis of COI barcoding data. 'Spec No.' refers to the project specimen ID.

Spec No.	Previous species identity	Name after analysis
12	*Synergus pallidipennis*	*Synergus "gallaepomiformis -like"*
16	*Synergus gallaepomiformis/pallicornis*	*Synergus "gallaepomiformis -like"*
17	*Synergus gallaepomiformis/pallicornis*	*Synergus "gallaepomiformis -like"*
18	*Synergus pallicornis/gallaepomiformis*	*Synergus "gallaepomiformis -like"*
19	*Synergus pallicornis/gallaepomiformis*	*Synergus "gallaepomiformis -like"*
20	*Synergus pallicornis*	*Synergus "gallaepomiformis -like"*
21	*Synergus pallicornis*	*Synergus "gallaepomiformis -like"*
22	*Synergus pallicornis/nervosus-albipes*	*Synergus "gallaepomiformis -like"*
23	*Synergus pallicornis/nervosus-albipes*	*Synergus "gallaepomiformis -like"*
24	*Synergus gallaepomiformis/albipes*	*Synergus "gallaepomiformis -like"*
25	*Synergus gallaepomiformis/albipes*	*Synergus "gallaepomiformis -like"*
26	*Synergus nevosus/albipes*	*Synergus "gallaepomiformis -like"*
27	*Synergus nevosus/albipes*	*Synergus "gallaepomiformis -like"*
40	*Synergus apicalis/rotundiventris*	*Synergus apicalis*
51	*Synergus apicalis/rotundiventris*	*Synergus apicalis*
42	*Synergus thaumacerus*	*Synergus consobrinus*
43	*Synergus thaumacerus*	*Synergus consobrinus*
13	*Synergus pallidipennis*	*Synergus diaphanus*
10	*Synergus umbraculus/hayneanus*	*Synergus hayneanus*
11	*Synergus umbraculus/hayneanus*	*Synergus hayneanus*
S6	*Synergus hayneanus/reinhardi*	*Synergus reinhardi*
7	*Synergus hayneanus/reinhardi*	*Synergus reinhardi*
9	*Synergus hayneanus*	*Synergus reinhardi*
48	*Saphonecrus haymi*	*Synergus tibialis*
52	*Synergus apicalis/rotundiventris*	*Synergus tibialis*
39	*Synergus flavipes*	*Synergus variabilis*

However, it is also possible that some of the confusion is the result of the drawbacks in using mitochondrial sequence described earlier in this chapter. Errors may be due to mistakes in the original morphology-based identification of specimens, or to introgression or lineage sorting between closely related species. That reassessment of inquiline morphological taxonomy is necessary is supported by the fact that collaborative work with Zoltan Ács and George Melika shows nuclear sequence data (unaffected by introgression) broadly to support my conclusions. The inquilines may thus represent a key test of molecular markers to resolve the status of taxa unresolvable using morphology.

6.4 Discussion

The results of the parasitoid barcoding study suggest that COI can provide a useful means by which to distinguish between species of oak gall wasp parasitoid. To quote Hebert (2004): "the simplest test of species identification by DNA barcode is whether any sequences are found in two species; none was in this study". What is more, most species were separated by relatively large sequence divergence, so most sequences should be diagnostic enough that identifications can be made unequivocally. Further sampling is now required from additional locations across the species' ranges, and from larger numbers of species within each family, until the entire community is adequately sampled. I aim to complete this work in the near future.

No evidence of introgression was found for the parasitoids. However, introgression is strongly implicated for the gall wasps (Rokas *et al* 2003b) and if gall wasps were to be barcoded use of a nuclear gene in addition to COI would be sensible. Such an approach is advocated by Blaxter (2004a), who recommends use of at least one nuclear and one organelle gene for any barcoding system. The close relationship between the cynipid gall-inducers and inquilines underlines the possibility of introgression-related problems in the inquilines. We have adopted a combined mitochondrial and nuclear gene approach in our ongoing work in the inquiline cynipids, combining COI with data for the 28S RNA nuclear gene. In general, such an approach seems particularly appropriate in groups of closely-related taxa whose status as biological species remains unexamined.

The issue of how to derive a species boundary from barcode sequence data is not clear. Although Hebert *et al* (2003a, 2003b, 2004) have conducted studies in which a large number of species were investigated, a true test of the precision of a COI barcode will require all members of a genus to be investigated, rather than a sample of poorly defined close relatives (Mortiz & Cicero 2004). This issue has been demonstrated by Johnson & Cicero (2004), who compared sister species of North American bird, and obtained results that contrasted strongly with those of Hebert *et al* (2004). A similar study is not possible with sole reference to oak gall wasp parasitoids, since many of the genera also contain many species that attack other non-cynipid hosts. Problems associated with species delineation appear to have no simple solution, and "the species problem" (Mayr & Ashlock 1991, Hull 1997), persists in the absence of a unifying method of identifying what constitutes a species. Perhaps the best way to proceed is to define empirical limits within specific taxa (families, genera) as available datasets grow.

As stated above, the results provided in this study suggest that COI has the potential to discriminate between different species of gall inhabitant. However, an additional question is whether it might be possible to utilise the tool in order to avoid practical problems associated with gall rearing (see above). The potential value of extracting DNA from a gall in order to gain good quality ecological data regarding the abundance and diversity of inhabitants without the need to rear them is obvious. However, it is not currently clear whether such a goal is possible. Extraction of DNA from whole galls would probably be complicated by the high levels of phenolic secondary plant compounds in gall tissues, making careful extraction and clean-up an absolute necessity if PCR inhibition is not to be a problem in barcoding. If DNA can be successfully extracted from whole galls, species recognition via chip-based DNA arrays should be possible fairly cheaply. Such an approach would involve binding to silicon chips a catalogue of oligonucleotides specific to parasitoid barcodes, and then using this to probe the products of PCR amplified in the presence of fluorescently labelled DNTPs, so that bound sequences could be read by a computer. The pattern of annealing would then report which species were present. However, the limitation of this approach is that it would only provide information on presence/absence of species and not on abundance. Furthermore, unknown species, or those for which barcodes had not been developed, would also not be identified. It is not currently apparent how a cheap method capable of reporting both presence/absence and relative abundance would work. However, such a technique has the potential to revolutionise the collection of data from not only oak galls, but from analogous habitats, and it is likely that efforts to provide a solution will continue.

Chapter 7

Concluding remarks

At first glance, an oak gall may appear rather unremarkable. However, after an appreciation has been gained for the complex interactions that occur within oak galls, and a glimpse taken of the wide variety of creatures that inhabit them, they are anything but ordinary. As a child, when I used to occasionally find oak galls during fishing trips or visits to the local park, I imagined they were the result of some kind of plant disease, and was slightly repulsed. However, years later when I began the research for this PhD, these feelings were replaced by a sense of amazement. The primary cause for this amazement was the discovery that the galls were home to an incredible diversity of parasitic wasps. These parasitoids come in a wide range of shapes and sizes and many are vibrant shades of metallic green, gold or blue. The idea that there were miniature ecosystems often consisting of relatively large numbers of these wasps in single galls no larger than the size of a marble was astonishing.

Up until beginning this PhD, I had paid little attention to the existence of parasitic wasps, but although there is generally scant awareness of their existence outside of entomological circles, they are highly abundant and important components of terrestrial ecosystems. Rivalling the weevil in terms of species richness, and perhaps ultimately the most diverse animal group on the planet, parasitic wasps more generally play a key ecological role. To a large extent, they are responsible for the regulation of innumerable invertebrate host populations, including those of many significant pest species. The impact of the parasitic wasp is due almost entirely to the feeding habits of their immature stages. Female parasitic wasps oviposit their eggs either onto or into the host, often simultaneously injecting powerful venom that results in either partial or total host incapacitation. The fate of the host is a protracted death, during which it is consumed bit by bit by the developing parasitic wasp larva. The host is subject to such a grisly demise, that contemplation of its fate has at various times contributed to Darwin's loss of faith, and acted as the inspiration for the Hollywood blockbuster movie 'Alien'. Since death of the host usually occurs, parasitic wasps would actually be more correctly referred to as parasitoid wasps. It is hard not to be impressed by the parasitic wasp when observed under the microscope, where they resemble tiny machines that appear to have been perfectly engineered for their role. Consideration of how such tiny creatures are capable of locating their hosts at all is equally impressive.

Oak gall wasps have received a fair amount of scientific attention in recent years, however, much remains to done, and some of the biggest questions regarding their biology remain unanswered. Perhaps the most obvious question still yet to be resolved is how gall wasps induce the formation of the gall. Despite the many advances in modern biology, we still seem some way off finding out just how gall wasps hijack the oaks developmental pathway for their own means, resulting in the formation of characteristic and often incredibly complicated novel structures.

The primary focus of this thesis was an examination into the phylogeography of oak gall wasp parasitoids and how this relates to that of their hosts. In addition to extending our knowledge of oak gall wasp communities, this is part of a move to expand the study of phylogeography to the level of whole communities as opposed to single species. Determining the extent to which communities show concordant phylogeographic origins has important implications for the study of community ecology and the search for so called community assembly rules. Truly independent comparisons between communities can only be made if they are phylogenetically distinct, otherwise the patterns observed may be due to shared ancestry and not as a result of some underlying rule of association. The results uncovered for the parasitoids in this thesis are concordant with what has so far been found for the gall wasp hosts, and suggest a possible origin of oak gall wasp communities somewhere in the east. These findings are exciting, however, it will be necessary to increase sample size significantly by extension to a much larger sample of species before real conclusions can be drawn.

In addition to examining the European phylogeography of oak gall wasp parasitoids, an investigation was made into the possible structuring effect of oak gall traits on parasitoid populations. It is often thought that these traits must have evolved to provide gall wasps with some degree of protection against parasitoids, given the high rate of mortality which they suffer. However, the majority of parasitoids that attack oak galls are generalists and are able to successfully attack galls possessing a wide range of different structural features. One possible solution to this apparent contradiction is that seemingly generalist parasitoids may actually be composed of a number of distinct host races. No evidence of this was found in this study, however, it is entirely possible that the use of higher resolution nuclear markers will provide evidence of their existence.

All the methods employed in this thesis have used molecular tools as a basis for investigating evolutionary and ecological questions. The use of such tools has revolutionised the way in which biological research is conducted. However, they have also changed the way in which biologists work, and perhaps if Darwin had been a contemporary biologist, it would be unlikely that he would have spent 5 years travelling the world exploring and collecting specimens. On the other hand, these are exciting times for biological research and the use of molecular methods enables the examination of questions that just several decades ago were beyond the wildest dreams of biologists.

Future directions for the study of oak gall wasp communities include the extension of phylogeographic studies to include inquiline species, in order to discover whether they also show evidence of a shared origin in the east. Also of particular interest is the extent to which inquilines will be structured by oak section, given their closer association with the oak host. Of course the big issues of the explanation of gall morphology and the mystery of gall induction remain, and will no doubt be the subject of much further research in years to come. It is also clear that much work remains to be done on oak gall wasp parasitoids, and after completion of this PhD I hope to spend at least a little more time investigating further some of the fascinating aspects of their biology.

References

Abe Y (1992) The advantage of attending ants and gall aggregation for the gall wasp *Andricus symbioticus* (Hymenoptera: Cynipidae) *Oecologia* **89**: 166–167.

Abrahamson WG, Blair CP, Eubanks MD, Morehead SA (2003) Sequential radiation of unrelated organisms: the gall fly Eurosta solidaginis and the tumbling flower beetle *Mordellistena convicta. Journal of Evolutionary Biology* **16**: 781-789.

Acs Z, Melika G, Kalo P, Kiss GB (2002) Molecular analysis of the *Eurytoma rosae* species group (Chalcidoidea, Eurytomidae). In: *Parasitic Wasps: Evolution, Systematics, Biodiversity and Biology Control* (ed. Melika G, Thúroczy C), pp. 234-242. Agroinform, Budapest.

Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**: 716-723.

Araujo MB, Pearson RG, Thuiller W, et al (2005) Validation of species-climate impact models under climate change. *Global Change biology* **11**, 1504-1513.

Askew RR (1961) On the biology of the Inhabitants of oak Galls of Cynipidae (Hymenoptera) in Britain. *Transactions of the Society for British Entomology* **14**: 237-268.

Askew RR (1965) The biology of the British species of the genus *Torymus* Dalman associated with galls of Cynipidae on oak, with special reference to alternation of forms. *Transactions of the Society for British Entomology* **16**: 217-232.

Askew RR (1984) The biology of gall wasps. In: *The Biology of Galling Insects* (ed. Ananthakrishnan TN), pp. 223-271. IBH Publishing Co, New Delhi.

Askew RR (1999) Parasitoid oviposition in the false chamber of the gall of *Cynips disticha* Hartig (Hym. Eurytomidae, Cynipidae). *Entomologist's Monthly Magazine* **135**: 215-216.

Askew RR, Nieves-Aldrey JL (2005) A new genus and species of Pteromalidae

(Hymenoptera, Chalcidoidea) from Spain, parasitic in cynipid galls on Fumaria. *Journal of Natural History* **39**: 2331-2338.

Askew RR, Shaw MR (1986) Parasitoid communities: their size, structure and development. In: *Insect parasitoids*. (ed. Wagge J, Greathead D), pp. 225-264. Academic Press, London, UK.

Askew & Thuroczy (In Press) Key to the parasitoids associated with oak cynipid galls. In: *The oak gallwasps of the Western Palaearctic: ecology, evolution and systematics* (ed. Stone GN, Csóka G, Melika G). The Ray Society, London.

Atkinson RJ, McVean GAT, Stone GN (2002) Use of population genetic data to infer oviposition behaviour: species-specific patterns in four oak gall wasps (Hymenoptera: Cynipidae). *Proceedings of the Royal Society Series B* **269**: 383–390.

Atkinson RJ, Brown GS, Stone GN (2003) Skewed sex ratios and multiple founding in galls of the oak apple gall wasp Biorhiza pallida (Hymenoptera: Cynipidae). *Ecological Entomology* **28**: 14–24.

Atkinson RJ, Rokas A, Stone GN (In press) Longitudinal patterns in species richness and genetic diversity in European oaks and oak gallwasps. In: Phylogeography in southern European refugia (ed. Weiss S, Ferrand N). Kluwer Academic Press, Dordrecht, The Netherlands.

Auger-Rozenberg M, Kerdelhué C, Magnoux E, Turgeon J, Rasplus J-Y, Roques A (2005) Molecular phylogeny and evolution of host-plant use in conifer seed chalcids in the genus *Megastigmus* (Hymenoptera: Torymidae) *Systematic Entomology* In press.

Avise JC (1987) *Identification and interpretation of mitochondrial DNA stocks in marine species*. In: *Proceedings of the stock identification workshop*. (ed. H. Kumph and E. L. Nakamura), pp. 105–136. National Oceanographic and Atmospheric Administration, Panama City, Florida, USA.

Avise JC (2000) *Phylogeography: the History and Formation of Species*. Harvard University Press, Cambridge.

Avise JC (2004) *Molecular markers, natural history, and evolution.* (Second Edition) Sinauer Associates, Sunderland, Massachusetts.

Avise JC, Giblin-Davidson C, Laerm J, Patton JC, Lansman RA (1979) Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis. Proceedings of the National Academy of Sciences of the United States of America* **76**: 6694-6698.

Bayes T (1763) An Essay towards solving a Problem in the Doctrine of Chances. *Philosophical Transactions of the Royal Society of London*, **53**: 370-418.

Beebe NW, Ellis JT, Cooper RD, Saul A (1999) DNA sequence analysis of the ribosomal DNA ITS2 region for the *Anopheles punctulatus* group of mosquitoes. *Insect Molecular Biology* **8**: 381-390.

Benayas JMR, de la Montana E (2003) Identifying areas of high-value vertebrate diversity for strengthening conservation. *Biological Conservation* **114**: 357-370.

Bennett KD (1986) The rate of spread and population increase of forest trees during the postglacial. *Philosophical Transactions of the Royal Society of London, Series B* **314**: 523-531.

Bensasson D, Zhang DX, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. *Trends in Ecology and Evolution* **16**: 314–321.

Blaxter M (2003) Counting angels with DNA. *Nature* **421**: 122-124.

Blaxter M (2004) The promise of a DNA taxonomy. *Philosophical Transactions of the Royal Society of London B Biological Sciences* **359**: 669–679.

Blaxter M, Elsworth B, Daub J (2004). DNA taxonomy of a neglected animal phylum: an unexpected diversity of tardigrades. *Proceedings of the Royal Society of London, Series B.* **271**: 189-192.

Boivin S, Kerdelhué C, Rozenberg MA, Cariou ML, Roques A (2003) Characterization of microsatellite loci in a seed chalcid, *Megastigmus spermotrophus* (Hymenoptera: Torymidae). *Molecular Ecology Notes* **3**: 363-365.

Bonsall MB, Hassell MP (1997) Aparent competition structures ecological assemblages. *Nature* **388**: 371-373.

Boucek Z (1988). *Tamarixia leucaenae*, new species (Hymenoptera: Eulophidae) parasitic on the leucaena psyllid *Heteropsylla cubana* Crawford (Hemiptera) in Trinidad. *Bulletin Of Entomological Research* **78**: 545-547.

Braig HR, Zhou W, Dobson S, O'Neill SL (1998) Cloning and characterisation of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia*. *Journal of Bacteriology* **180**: 2373-2378.

Braune HJ (1992) Eichengallen als ökologische Kleinsysteme: Analytische Studien zur strukturellen und funktionellen Organisation ihrer Bewohner. *Faunistisch-Ökologische Mitteilungen* **6**: 299–318.

Brooks TM, Mittermeier RA, Mittermeier CG, da Fonseca GAB, Rylands AB, Konstant WR (2002) Habitat loss and extinction in the hotspots of biodiversity. *Conservation Biology* **16**: 909-923.

Brown WM, Vinograd J (1974) Restriction endonuclease cleavage maps of animal mitochondrial DNAs. *Proceedings of the National Academy of Sciences of the United States of America* **71**: 4617-4621.

Brewer S, Cheddadi R, de Beaulieu JL, Reille M, Data contributors (2002) The spread of deciduous *Quercus* throughout Europe since the last glacial period. *Forest Ecology and Management*, **156**: 2748.

Brower AVZ (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, U.S.A.*, **91**, 6491-6495.

Buckler ES, Ippolito A, Holtsford TP (1997) The evolution of ribosomal DNA: divergent paralogues and phylogenetic implications. *Genetics* **145**: 821–832.

Bush GL (1969) Sympatric host race formation and speciation in frugivorous flies of the genus Rhagoletis (Diptera, Tephritidae). *Evolution* **23**: 237-241

Bush GL (1975) Modes of animal speciation. *Annual Review of Ecology and Systematics* **6**: 339-364.

Camin JH, Sokal RR (1965) A method for deducing branching sequences in phylogeny. *Evolution* **19**: 311-326.

Campbell B, Heraty J, Rasplus JY, Chan K, Campbell JS, Babcock C, (2000) Molecular systematics of the Chalcidoidea using 28S-D2 rDNA. In: *Hymenoptera: Evolution, Biodiversity and Biological Control.* (Ed. Austin AD, Dowton M), pp. 59-73. CSIRO.

Campbell BC, Steffen-Campbell JD, Werren JH (1993) Phylogeny of the Nasonia species complex (Hymenoptera: Pteromalidae) inferred from an internal transcribed spacer ITS2) and 28S rDNA sequences. *Insect Molecular Biology* **2**: 225-237.

Castro LR, Dowton M, Austin AD (2002) Contrasting rates of mitochondrial molecular evolution in parasitic Diptera and Hymenoptera. *Molecular Biology and Evolution* **19**, 1100-1113.

Caterino MS (2000) Descriptions of the first Chlamydopsinae (Coleoptera: Histeridae) from Wallacea. *Tidschrift voor Entomologie* **43**: 267-278.

Caterino MS, Cho S, Sperling FAH (2000) The current state of insect molecular systematics: a thriving Tower of Babel. *Annual Review of Entomology* **45**: 1-54.

Cavalli-Sforza LL, Edwards AWF (1967) Phylogenetic analysis: models and estimation procedures. *Evolution* **32**: 550-570.

Chen Y, Hui XA, Fu JZ, Huang DW (2004) A molecular phylogeny of eurytomid wasps inferred from DNA sequence data of 28S, 18S, 16S, and COI genes. *Molecular*

Phylogenetics and Evolution **31**: 300-307.

Chown SL, Gaston KJ (2000) Areas, cradles and museums: the latitudinal gradient in species richness. *Trends in Ecology and Evolution* **15**: 311-315.

Claridge MF, Askew RR, (1960) Sibling species in the Eurytoma rosae group (Hym., Eurytomidae). *Entomophaga* **5**: 141-53.

Clark KR (1993) Non-parametric multivariate analysis of changes in community structure. *Australian Journal of Ecology* **18**: 117-143.

Clark KR, Gorley RN (2001) *PRIMER v5: User manual/tutorial. PRIMER-E*, pp. 91. Plymouth,UK.

Collura RV, Stewart CB (1995) Insertions and duplications of mitochondrial DNA in the nuclear genomes of Old World monkeys and hominoids. *Nature* **378**: 485-489.

Cook JM, Rokas A, Pagel M, Stone GN (2002) Evolutionary shifts between host oak sections and host plant organs in *Andricus* gallwasps. *Evolution* **56**: 1821-1830.

Compton SG, Ellwood MDF, Davis AJ, *et al.* (2000) The flight heights of chalcid wasps (Hymenoptera, Chalcidoidea) in a lowland bornean rain forest: Fig wasps are the high fliers. *Biotropica* **32**: 515-522.

Cornell HV (1983) The secondary chemistry and complex morphology of galls formed by the Cynipinae. Why and how? *American Midland Naturalist* **110**: 225–234.

Craig TP, Itami JK, Price PW (1990) The window of vulnerability for a shoot-galling sawfly to attack by a parasitoid. *Ecology* **71**: 1471-1482.

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution* **15**: 290-295.

Cronin JT, Abrahamson WG, (2001) Do parasitoids diversify in response to host-plant shifts by herbivorous insects? *Ecological Entomology* **26**: 347-355.

Csóka G, Stone GN, and Melika, G (2005) Biology, Ecology and evolution of gall-inducing cynipidae. In: *Biology, ecology and evolution of gall-inducing arthropods* (ed. Raman A, Shaeffer CW, Withers TM), Pp. 569-636. Science Publishers, inc, USA, (2 volumes).

Davis PH (1965). *Flora of Turkey and the east Aegean islands.* Vol.1-9, Edinburg University Press, Edinburgh, Great Britain.

Davis P (1971) Distribution patterns in Anatolia with particular reference to endemism. In: *Plant life of South-West Asia.* (Ed. Davis PH, Harper PC, Hedge IC) Botany Society of Edinburgh, Great Britain.

Diamond JM (1975) *Assembly of Species Communities, in Ecology and Evolution of Communities* (ed. Cody ML, Diamond JM), pp. 342-444 Belknap, Harvard.

Douady CJ, Delsuc F, Boucher Y, Doolittle WF, Douzery EJP (2003) Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. *Molecular Biological and Evolution* **20**: 248-254.

Dowton M, Austin AD (2001) Simultaneous analysis of 16S, 28S, COI and morphology in the Hymenoptera: Apocrita - evolutionary transitions among parasitic wasps. *Biological Journal of the Linnean Society* **74**: 87-111.

Doyle JJ (1992) Gene trees and species trees: molecular systematics as one-character taxonomy. *Systematic Botany* **17**: 144–163.

Drummond AJ, Rambaut A (2003) BEAST v1.0, Available from http://evolve.zoo.ox.ac.uk/beast/.

Dumolin-Lapègue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. *Genetics* **146**: 1475-1487.

Eck RV, Dayhoff MO (1966) Evolution of the structure of ferredoxin based on living relics of primitive amino acid sequences. *Science* **152**: 363-366.

Edwards AWF, Cavalli-Sforza LL (1964) Re-construction of evolutionary trees. In: *Phenetic and phylogenetic classification* (ed. Heywood VH, McNeill J), pp. 67-76. Systematics Association, London.

Felsenstein J (1973) Maximum-likelihood and minimum-steps methods for estimating evolutionary trees from data on discrete characters. *Systematic Zoology*, **22**: 240-249.

Felsenstein J (1978) Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27**: 401-410.

Felsenstein J (1979) Alternative methods of phylogenetic inference and their interrelationship. *Systematic Zoology* **28**: 49-62.

Felsenstein J (1981) Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* **17**: 368-376.

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

Felsenstein J (1989) PHYLIP-phylogeny inference package (Version 3.2). *Cladistics*, **5**: 164-166.

Felsenstein J (2004) *Inferring Phylogenies*. Sinauer Associates, Sunderland, Massachusetts.

Fernandes GW, Fagundes M, Woodman RL, Price PW (1999) Ant effects on three-trophic level interactions: plant, galls, and parasitoids. *Ecological Entomology*, **24**: 411-415.

Ferris SD, Sage RD, Prager EM, Ritte U, Wilson AC (1983) Mitochondrial DNA evolution in mice. *Genetics* **105**: 681-721.

Ferris C, Oliver RP, Davy AJ, Hewitt GM (1993) Native oak chloroplasts reveal an ancient divide across Europe. *Molecular Ecology* **2**: 337-344 .

Fitch AE (1880) Insects bred from *Cynips kollari* galls. *Entomologist* **13**: 252–263.

Floyd R, Eyualem A, Papert A, Blaxter M (2002) Molecular barcodes for soil nematode identification. *Molecular Ecology* **11**: 839–850.

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

Friedrich M, Tautz D (1995) Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. *Nature* **376**: 165-167.

Gaut BS, Lewis PO (1995) Success of maximum likelihood phylogeny inference in the four-taxon case. *Molecular Biology and Evolution* **12**: 152-162.

Gilbert FS, Astbury C, Bedingfield J, Ennis B, Lawson S, Sitch TA (1994) The ecology of the pea galls of Cynips divisa. In: *Plant Galls: Organisms, Interactions, Populations* (ed. Williams MAJ), pp. 369 390. Clarendon Press, Oxford.

Godfray HCJ (2002) Challenges for taxonomy. *Nature* **417**: 17-19.

Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular Biology and Evolution* **11**: 725-736.

Glemet, H, Blier P, Bernatchez L, (1998) Geographical extent of Arctic char (*Salvelinus alpinus*) mtDNA introgression in brook char populations (*S. fontinalis*) from eastern Quebec, Canada. *Molecular Ecology* **7**: 1655–1662.

Graham MWR, de V Gijswijt MJ (1998) Revision of the European species of *Torymus* Dalman (s. lat.) (Hmyenoptera: Troymidae). *Zoologische Verhandelingen* **317**:1-202.

Grissell EE (1995) Toryminae (Hymenoptera: Chalcidoidea: Torymidae): A redefinition, generic classification, and annotated world catalog of species. *Memoirs on Entomology, International* 2: 1-470.

Grissell EE (1999) An annotated catalog of world Megastigminae (Hymenoptera: Torymidae). *Contributions of the American Entomological Institute*, **31**: 1-92.

Hails RS, Crawley MJ, (1992) Spatial density dependence in populations of a cynipid gall-former Andricus quercuscalicis. *Journal of Animal Ecology* **61**: 567-583.

Hale ML, Acs Z, Stone GN (2004) Polymorphic microsatellite loci in *Eurytoma brunniventris*, a generalist parasitoid in oak cynipid galls. *Molecular Ecology Notes* **4**: 197-199.

Harrison RG (1989) Animal mtDNA as a genetic marker in population and evolutionary biology. *Trends in Ecology and Evolution*, **4**: 6-11.

Harrison RG, Rand DM, Wheeler WC (1987) Mitochondrial DNA variation in field crickets across a narrow hybrid zone. *Molecular Biology and Evolution* **4**: 144-158.

Harry M, Solignac M, Lachaise D (1998) Molecular evidence for parallel evolution of adaptive syndromes in fig-breeding *Lissocephala* (Drosophilidae). *Molecular Phylogenetics and Evolution* **9**: 542-551.

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**: 160-174.

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**: 97-109.

Hayward A, Stone GN (2005) Oak gall wasp communities: ecology and evolution. *Basic and Applied Ecology* **6**: 435-443.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identifications through DNA barcodes. *Proceedngs of the Royal Society of London B-Biological Sciences* **270**: 313–321.

Hebert PDN, Ratnasingham S, deWaard JR (2003b) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedngs of the Royal Society of London B (Supplement)* **270**: S96-S99.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004) Identification of birds through DNA barcodes. *PLoS Biology* **2**: e312.

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

Hewitt GM (1999) Post-glacial recolonisation of European biota. *Biological Journal of the Linnean Society* **68**: 87–112.

Hillis DM, Bull JJ (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Systematic Biology* **42**: 182-192.

Hillis DM, C Moritz, BK Mable (1996) Molecular Systematics. 2nd edn. Sunderland, MA: Sinauer Associates, Inc.

Holt RD (1977) Predation, apparent competition, and the structure of prey communities. *Theoretical Population Biology* **12**: 197-29.

Holt RD, Lawton JH (1993) Apparent competition and enemy free space in insect host parasitoid communities. *American Naturalist* **142**: 623-645.

Hopkins GW, Freckleton RP (2002) Declines in the numbers of amateur and professional taxonomists: implications for conservation. *Animal Conservation* **5**: 245-249.

Huelsenbeck JP, Crandall KA (1997) Phylogeny estimation and hypothesis testing using maximum likelihood. *Annual Review of Ecology and Systematics* **28**: 437-466.

Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogeny. *Bioinformatics*, **17**: 754–755.

Huelsenbeck JP, Larget B, Miller RE, Ronquist F (2002) Potential applications and pitfalls of Bayesian inference of phylogeny. *Systematic Biology* **51**: 673-688.

Hull DL (1997) The Ideal Species Definition and Why We Can't Get It. In: *Species: The Units of Biodiversity,* (ed. Claridge MF, Dawah, HA Wilson MR), pp. 357-380.London: Chapman Hall

Huntley, B. & Birks, H. J. (1983) *An Atlas of Past and Present Pollen Maps for Europe, 0-13000 years ago.* Cambridge: Cambridge University Press.

Hurst GDD, Jiggins FM (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 series B* **272**: 1525-1534.

Irestedt M, Fjeldså J, Nylander JAA, Ericson P (2004) Phylogenetic relationships of typical Antbirds (Thamnophilidae) and test of incongruence based on Bayes factors. *BMC Evolutionary Biology* **4**: 23.

Jalas J, Suominen J (1987) *Atlas Florae Europaeae. Distribution of vascular plants in Europe II. Angiosperms (part) 3: Salicaceae to Balanophoraceae; 4: Polygonaceae; 5: Chenopodiaceae to Basellaceae.* Cambridge: Cambridge University Press.

Jeffries MJ, Lawton JH (1984) Enemy free space and the structure of ecological communities. *Biological Journal of the Linnean Society* **23**: 269-286.

Jermin LS, Crozier RH (1994) The cytochrome b region in the mitochondrial DNA of the ant *Tetraponera rufoniger*: sequence divergence in Hymenoptera may be associated with nucleotide content. *Journal of Molecular Evolution* **38**: 282–294.

Johnson NK, Cicero C (2004) New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. *Evolution* **58**: 1122–1130.

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: *Mammalian Protein Metabolism* (ed. Munro HN), pp21-123. Academic Press, New York.

Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association* **90**: 773-795.

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**: 111-120.

Kondoh M (2003) Foraging adaptation and the relationship between food-web complexity and stability. *Science* **299**: 1388-1391.

Kraytsberg Y, Schwartz M, Brown TA, Ebralidse K, Kunz WS, Clayton DA, Vissing J, Khrapko K (2004) Recombination of human mitochondrial DNA. *Science* **304**: 981.

Kremer A, Petit RJ (1993) Gene diversity in natural populations of oak species. *Annales des Sciences Forestières*, **50**: 186-202.

Kwast E (2001) Range expansion of Andricus aries in Europe. *Cecidology* **16**: 62–67

Lanave C, Preparata G, Saccone C, Serio G (1984) A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution* **20**: 86-93.

La Salle J, Gauld ID (1991) Parasitic Hymenoptera and the biodiversity crisis. Insect Parasitoids, 4th European Workshop. *Redia* **74**: 3315-334.

La salle, J (2005) biology of gall inducers and evolution of gall induction in chalcidoidea (Hymenoptera: Eulophidae, Eurytomidae, Pteromallidae, Tanaostimatidae, Torymidae). In: *Biology, ecology and evolution of gall-inducing arthropods* (ed. Raman A, Shaeffer CW, Withers TM), pp. 569-636. Science Publishers inc, USA, (2 volumes).

Lawton JH (1986) The effect of parasitoids on phytophagous insect communities. In: *Insect parasitoids.* (ed. Waage J, Greathead D), pp. 265–289. Academic, London.

Lehman N, Wayne RK (1991) Analysis of coyote mitochondrial DNA genotype frequencies: estimation of the effective number of alleles. *Genetics* **128**: 405-416.

Li S (1996) *Phylogenetic Tree Construction Using Markov Chain Monte Carlo.* Ph.D. thesis, Ohio State University, Columbus.

Liao D (2000) Gene conversion drives within genic sequences: concerted evolution of ribosomal RNA genes in bacteria and archaea. *Journal of Molecular Evolution* **51**: 305-317

Lill JT, Marquis RJ, Ricklefs RE (2002) Host plants influence parasitism of forest caterpilllars. *Nature* **417**: 170-173.

Lindley J, Moore T (1870) *A treasury of botany: a popular dictionary of the vegetable kingdom.* Part11, 2nd edn. Longmans, London.

Lockhart PJ, Larkum AWD, Steel MA, Waddell PJ, Penny D (1996) Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequences analysis. *Proceedings of the National Academy of Sciences of the USA* **93**: 1930-1934.

Loxdale HD, Lushai G (1998) Molecular markers in entomology. *Bulletin of Entomological Research* **88**: 577-600.

Machado CA, Robbins N, Gilbert MTP, Herre EA, (2005) Critical review of host specificity and its coevolutionary implications in the fig/fig-wasp mutualism. *Proceedings of the National Academy of Sciences of the USA* **102**: 6558-6565.

Mallet J, Willmott K (**2003**) Taxonomy: renaissance or Tower of Babel? *Trends in Ecology and Evolution* **18**:57-59.

Mau B (1996) *Bayesian phylogenetic inference via Markov chain Monte Carlo methods.* Ph. D. dissertation, University of Wisconsin, Madison.

May RM (1988) How many species are there on earth? *Science* **241**: 1441-1449.

Mayr E, Ashlock PD (1991) Principles of systematic zoology, 2nd ed. McGraw-Hill, New York, New York, USA.

Medail F (2001) Biodiversity hot spots: A tool for conservation. *Biofutur* **211**: 37-37.

Melika G, Abrahamson WG, (2003) Review of the world genera of oak Cynipid wasps (Hymenoptera: Cynipidae: Cynipini) In: *Parasitic wasps, evolution, systematics, biodiversity and biological control.* (Ed. Melika G, Thuroczy CS), pp. 150-192. Agroinform Budapest.

Metropolis N, Rosenbulth AW, Teller AH, Teller E (1953) Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**: 1087.

Manos PS, Doyle JJ, Nixon KC (1999) Phylogeny, Biogeography, and processes of molecular differentiation in *Quercus* subgenus *Quercus* (Fagaceae). *Molecular Phylogenetics and Evolution* **12**: 333-349.

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

Mantel N, and Valand RS (1970) A technique of nonparametric multivariate analysis. *Biometrics* **26**: 547-55.

Mardulyn P, Cameron SA (1999) The major opsin in bees (Insecta: Hymenoptera): a promising nuclear gene for higher level phylogenetics. *Molecular Phylogenetics and Evolution* **12**: 168–176.

Mardulyn P, Whitfield JB (1999) Phylogenetic signal in the COI, 16S and 28S genes for inferring relationships among genera of Microgastrinae (Hymenoptera: Braconidae): evidence of a high diversification rate in this group of parasitoids. *Molecular Phylogenetics and Evolution* **12**: 282-294.

McHugh D (1998) Deciphering metazoan phylogeny: the need for additional molecular data. *American Zoology* **38**: 859-866.

Moritz C, Cicero C (2004) DNA barcoding: Promises and pitfalls. *PLoS Biology* **2**: 1529–1531.

Moore WS (1995) Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. *Evolution* **49**: 718-726.

Morris RJ, Lewis OT, Godfray HCJ (2004) Experimental evidence for apparent competition in a tropical forest food web. *Nature* **428**: 310-313.

Nee S, May RM (1997) Extinction and the Loss of Evolutionary History. *Science* **278**: 692-694.

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

Neigel JE, Avise JC (1986) Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In: (ed. Karlin S, Nevo E) *Evolutionary Processes and Theory,* pp. 515–534. Academic Press, New York.

Newton MA, Raftery AE (1994) Approximate Bayesian Inference with the weighted likelihood boot-strap. *Journal of the Royal Statistical Society. Series B (Methodological)* **56**: 3-48.

Nieves-Aldrey JL (1983a) Sobre las especies de Sycophila Walker, asociadas con agallas de cinipidos en la Peninsula Iberica, con descripcion de una nueva especie (Hym., Eurytomidae). *Eos* **59**: 179–191.

Nieves-Aldrey JL (1983b) Sobre las especies del genero Mesopolobus (Hym., Pteromalidae) asociadas con agallas de cynipidos en Quercus sp. en Salamanca. *Boletín de la Asociacíon Española de Entomología* **7**: 9–18.

Nieves-Aldrey JL (1983c) Contribucion al conocimiento de los eulófidos (Hym., Chalcidoidea, Eulophidae) parásitos en las agallas de cinipidos producidas sobre especies de Quercus. *Boletín de la Asociacíon Española de Entomología* **6**: 43–54.

Nieves-Aldrey JL (1984a) Primeros datos sobre los representantes de la familia Ormyridae en Espana, con descripción de una nueva especie. *Graellsia* **40**: 119–127.

Nieves-Aldrey JL (1984b) Observaciones sobre los Torimidos (Hym., Chalcidoidea, Torymidae) asociados con agallas de Cinipidos (Hym., Cynipidae) sobre Quercus spp en la zona centro-occidental de Espana. *Boletín de la Asociacíon Española de Entomología* **8**:

121–134.

Nieves-Aldrey JL (2001) *Hymenoptera; Cynipidae. Fauna Ibérica. Museo Nacional de Ciencias Naturales,* Consejo Superior de Investigaciones Cientificas, Madrid, Spain.

Nieves-Aldrey JL Pujade-Villar J (1985) Sobre las especies ibéricas de la seccion 1 (Mayr, 1872) del género Synergus Htg. *Eos* **61**: 219–237.

Nieves-Aldrey JL Pujade-Villar J (1986) Sobre las especies ibéricas de la seccion 2 (Mayr, 1872) del género Synergus Htg con descripción de una especie nueva. *Eos* **62**: 137–165.

Nordlander G (1984) [What do we know about parasitic cynipoids (Hymenoptera)?] *Entomologiske Tidskrift* **105**: 36-40 [In Swedish with English summary].

Noyes JS (1998) *Catalogue of the Chalcidoidea of the world. World Biodiversity Database CD-ROM Series.* Expert Center for Taxonomic Identification, Amsterdam, The Netherlands.

Nylander JAA (2004) *MrModeltest v2.* Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Nylander JAA, Ronquist F, Huelsenbeck JP, Nieves-Aldrey JL (2004) Bayesian phylogenetic analysis of combined data. Systematic Biology **53**: 47-67.

Ohno, S (1970) *Evolution by Gene Duplication.* Springer-Verlag, Heidelberg.

Page RDM, Holmes EC (1998). *Molecular Evolution: A phylogenetic approach.* Blackwell Scientific, Oxford.

Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences* **12**: 357-358.

Pagel M, Meade A, Barker D (2004) Bayesian estimation of ancestral states on phylogenies. *Systematic Biology* **53**: 673-684.

Palumbi SR (1989) Rates of molecular evolution and the fraction of nucleotide positions free to vary. *Journal of Molecular Evolution* **29**: 180-187.

Pamilo P, Nei M (1988) Relationships between gene trees and species trees. *Molecular Biology and Evolution* **5**:568-583.

Parfitt E (1856) Note on *Cynips lignicola* and a description of its parasite *Callimone flavipes*. *Zoologist* **14**: 5075–5076.

Pearson RG, Dawson TP (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography* **12**: 361-371.

Petit RJ, Csaikl UM, Bordacs S, *et al.* (2002) Chloroplast DNA variation in European white oaks: phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management* **156**: 526.

Plantard O, Rasplus JY, Hochberg ME (1996) Resource partitioning in the parasitoid assemblage of the oak galler Neuroterus quercusbaccarum L. (Hymenoptera: Cynipidae). *Acta Oecologica* **17**: 1–15.

Plantard O, Hochberg ME (1998) Factors affecting parasitism in the oak galler Neuroterus quercusbaccarum (Hymenoptera: Cynipidae). *Oikos* **81**: 289–98.

Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. *Systematic Biology* **53**: 793-808.

Posada D, Crandall KA (1998) Model test: testing the model of DNA substitution. *Bioinformatics* **14**: 817-818.

Posada D, Crandall KA (2001) Selecting the best-fit model of nucleotide substitution. *Systematic Biology* **50**: 580-601.

Powell JR (1983) Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: evidence from *Drosophila*. *Proceedings of the National Academy of Sciences of the USA* **80**:

492-495.

Price PW, Fernandes GW, Waring GL (1987) Adaptive nature of insect galls. *Environmental Entomology* **16**: 15-24.

Price PW, Pschorn-Walcher H (1988) Are gall insects better protected against parasites than exposed feeders? *Ecological Entomology* **13**: 195-205.

Prüser F, Mossakowski D (1998) Low substitution rates in mitochondrial DNA in Mediterranean carabid species. *Insect Molecular Biology* **7**: 121-128.

Pujade-Villar J, Folliot R, Bellido D (2003) The life cycle of *Andricus hispanicus* (Hartig, 1856) n. stat., a sibling species of *A. kollari* (Hartig, 1843) (Hymenoptera: Cynipidae). *Butlletí de l' Institució Catalana d'História Natural* **71**: 83-95.

Pujade- Villar J, Nieves-Aldrey JL (1990) Revisión de las especies europeas del género Saphonecrus Dalla Torre and Kieffer, 1910 (Hymenoptera: Cynipidae: Cynipinae). *Bulletín de la Institució Catalana d'Historia Natural* **58**: 45–55.

Pujade-Villar J, Nieves-Aldrey JL (1993) Revisión de las especies europeas del género Ceroptres Htg. 1840 (Hymenoptera: Cynipidae). *Boletin de la Asociacion Española de* Entomologia **17**: 49–63.

Purvis A, Agapow PM, Gittleman JL, Mace GM (2000) Nonrandom Extinction and the Loss of Evolutionary History. *Science* **288**: 328-330.

Raftery AE (1996). Hypothesis testing and model selection. In: *Markov Chain Monte Carlo in Practice* (ed. Gilks WR, Spiegelhalter DJ, Richardson S), pp. 163-188. Chapman and Hall, London.

Rambaut A, Drummond A (2003) *Tracer: MCMC Trace Analysis Tool.* Oxford: University of Oxford.

Rannala B, Yang Z (1996) Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. *Journal of Molecular Evolution* **43**: 304-311.

Richardson BJ, Bavenstock PR, Adams M (1986) *Allozyme Electrophoresis: A Handbook for Animal Systematics and Population Studies.* New York: Academic Press.

Ritchie AJ (1984) *A review of the higher classification of the inquiline gall wasps (Hymenoptera: Cynipidae) and a revision of the Nearctic species of Periclistus Förster.* Ph.D. dissertation Carleton University, Ottawa.

Robinson-Rechavi M, Huchon D (2000) RRTree: Relative-Rate Tests between groups of sequences on a phylogenetic tree. *Bioinformatics* **16**: 296-297.

Rodriguez FL, Oliver J, Martin L, Medina JR (1990) The general stochastic model of nucleotide substitution. *Journal of Theoretical Biology* **142**: 485-501.

Rokas A, Atkinson R, Brown G, West SA, Stone GN (2001). Understanding patterns of genetic diversity in the oak gallwasp *Biorhiza pallida*: demographic history or a *Wolbachia* selective sweep? *Heredity* 87: 294-305.

Rokas A, Nylander JAA, Ronquist F, Stone GN (2002) A maximum likelihood analysis of eight phylogenetic markers using gallwasps (Hymenoptera: Cynipidae); implications for insect phylogenetic studies. *Molecular Phylogenetics and Evolution* **22**: 206-219.

Rokas A, Atkinson RJ, Webster L, Stone GN (2003)a. Out of Anatolia: longitudinal gradients in genetic diversity support a Turkish origin for a circum-Mediterranean gallwasp *Andricus quercustozae*. *Molecular Ecology* **12**: 2153-2174.

Rokas A, Melika G, Abe Y, Nieves Aldrey J-L, Cook JM, Stone GN (2003)b. Lifecycle closure, lineage sorting and hybridisation revealed in a phylogenetic analysis of European oak gallwasps (Hymenoptera: Cynipidae: Cynipini) using mitochondrial sequence data. *Molecular Phylogenetics and Evolution* 26: 36-45.

Roques A, Skrzypczynska M (2003) Seed-infesting chalcids of the genus Megastigmus Dalman, 1820 (Hymenoptera : Torymidae) native and introduced to the West Palearctic region: taxonomy, host specificity and distribution. *Journal of Natural History* **37**: 127-238.

Ronquist F (1994) Evolution of parasitism among closely related species: Phylogenetic relationships and the origin of inquilinism in gall wasps (Hymenoptera, Cynipidae). *Evolution* **48**: 241–266.

Ronquist F (1999) Phylogeny, classification and evolution of the Cynipoidea. *Zoologica Scripta* **28**: 139-164.

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

Rott AS, Godfray HCJ (2000) The structure of a leafminer-parasitoid community. *Journal of Animal Ecology* **69**: 274-289.

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

Sarich VM, Wilson AC (1973) Generration time and genomic evolution in primates. *Science* **179**: 1144-1147.

Schneider S, Roessli D, Excoffier L (2000) *Arlequin: A software for population genetics data analysis*. Version 2.000. Geneva: Genetics and Biometry Laboratory, University of Geneva.

Schonrogge K, Crawley MJ (2000) Quantitative webs as a means of assessing the impact of alien insects. *Journal of Animal Ecology* **69**: 841-868.

Schönrogge K, Stone GN, Crawley MJ (1995) Spatial and temporal variation in guild structure: parasitoids and inquilines of *Andricus quercuscalicis* (Hymenoptera: Cynipidae) in its native and alien ranges. *Oikos* **72**: 51–60.

Schönrogge K, Stone GN, Crawley MJ (1996a) Alien herbivores and native parasitoids: rapid development of guild structure in an invading gall wasp, *Andricus quercuscalicis* (Hymenoptera: Cynipidae). *Ecological Entomology* **21**: 71–80.

Schönrogge K, Stone GN, Crawley MJ (1996b) Abundance patterns and species richness of the parasitoids and inquilines of the alien gall former *Andricus quercuscalicis* Burgsdorf (Hymenoptera: Cynipidae). *Oikos* 77: 507–518.

Schönrogge K, Walker P, Crawley MJ (1998) Invaders on the move: parasitism in the galls of four alien gall wasps in Britain (Hymenoptera: Cynipidae). *Proceedings of the Royal Society of London, Series B* **256**: 1643–1650.

Schönrogge K, Askew RR (In press) Oak cynipid communities. In: *The oak gallwasps of the Western Palaearctic: ecology, evolution and systematics* (ed. Stone GN, Csóka G, Melika G). The Ray Society, London.

Schwarz G (1978) Estimating the dimension of a model. *Annals of Statistics* **6:** 461–464.

Settle WH, Wilson LT (1990) Variation by the variegated leafhopper and biotic interactions: parasitism, competition, and apparent competition. *Ecology* **71**: 1461-1470.

Shao R, Dowton M, Murrell A, Barker SC (2003) Rates of gene rearrangement and nucleotide substitution are correlated in the mitochondrial genomes of insects. *Molecular Biology & Evolution* 20: 1612-1619.

Shorthouse JD, Rohfritsch O, (1992) *The Biology of Insect-induced Galls.* Oxford University Press, Oxford.

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

Singer-Sam J, Tanguay RL, Riggs AD (1989) Use of chelex to improve the signal from a small number of cells. *Amplifications: A forum for PCR Users* **3**: 11.

Smith, F. (1854) A new British *Cynips* and the galls made thereby. *Transactions of the Proceedings of the Entomological Society London* **3**, 35.

Sorenson MD, Quinn TW (1998) Numts: A challenge for avian systematics and population biology. *The Auk* **115**: 214-221.

SPSS Inc. (1999) *SPSS Base 10.0 for Windows User's Guide.* SPSS Inc., Chicago IL.

Stone GN, Atkinson R, Rokas A, Csóka G, Nieves-Aldrey J-L (2001) Differential success in northwards range expansion between ecotypes of the marble gallwasp *Andricus kollari*: a tale of two lifecycles. *Molecular Ecology* **10**: 761-778.

Stone GN, Cook JM (1998) The structure of cynipid oak galls: patterns in the evolution of an extended phenotype. *Proceedings of the Royal Society of London, Series B* **265**: 979-988.

Stone GN, Schönrogge K, Crawley MJ, Fraser S (1995) Geographic variation in the parasitoid community associated with an invading gallwasp, *Andricus quercuscalicis* (Hymenoptera: Cynipidae). *Oecologia* **104**: 207-217.

Stone GN, Schönrogge K, Atkinson RJ, Bellido D, Pujade-Villar J (2002) The population biology of oak gallwasps (Hymenoptera: Cynipidae). *Annual Review of Entomology* **47**: 633-668.

Stone GN, Schönrogge K (2003) The adaptive significance of insect gall morphology. *Trends in Ecology and Evolution* **18**: 512-522.

Stone GN, Sunnucks PJ (1993) The population genetics of an invasion through a patchy environment: the cynipid gallwasp *Andricus quercuscalicis. Molecular Ecology* **2**: 251-268.

Stork NE (1997) Measuring global biodiversity and its decline. In: *Biodiversity II: Understanding and protecting our natural resources,* (ed. Wilson EO, Reaka-Kudla ML, Wilson DE), pp. 41-68. Washington, D.C., Joseph Henry/National Academy Press.

Sullivan J, Swofford DL (1997) Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. *Journal of Mammal Evolution* **4**: 77–86.

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM (1996) Phylogenetic inference. In: *Molecular Systematics*, 2nd edn (ed. Hillis DM, Moritz C, Mable BK), pp. 407-514. Sinauer Associates

Sunderland, Massachusetts.

Swofford DL (2002) *paup*. Phylogenetic Analysis Using Parsimony (*and Other Methods).* Sinauer Associates, Sunderland, MA.

Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF (1998) Comparative phylogeography and postglacial colonization routes in Europe. *Molecular Ecology* **7**: 453–464.

Takabayashi J, Dicke M (1996) Plant-carnivore mutualism through herbivore-induced carnivore attractants. *Trends in Plant Science* **1**: 109-113.

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution **10**: 512-526.

Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. *Trends in Ecology and Evolution* **18**: 70–74.

Taylor JS, Raes J (2004) Duplication and Divergence: The evolution of new genes and old ideas. *Annual Review of Genetics* **9**: 615-643.

Templeton AR (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. *Evolution* **37**: 221-244.

Templeton AR (1998) Nested clade analysis of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology* **7**: 381-397.

Thompson JN (1999) Specific hypotheses on the geographic mosaic theory of coevolution. *American Naturalist* **153**: S1-S14.

Thompson JN, Cunningham BM (2002) Geographic structure and dynamics of coevolutionary selection. *Nature* **6890**: 735-738.

Toumi L, Lumaret R (1998) Allozyme variation in cork oak (*Quercus suber* L.): the role of phylogeography and genetic introgression by other Mediterranean oak species. *Theoretical and Applied Genetics* **97**: 647-656.

Turlings TCJ, Loughrin JH, McCall PJ, *et al.* (1995) How caterpillar-damaged plants protect themselves by attracting parasitic wasps. *Proceedings of the National Academy of Sciences USA* **92**: 4169-4174.

Tzedakis PC (1993) Long-term tree populations in northwest Greece through multiple Quaternary climatic cycles. *Nature* **364**: 437-440.

Van Benthem F. et al. (1984) Fagaceae. The Northwest European Pollen Flora, 33. *Review of Paleobotany and Palynology* **42**: 87-110.

Vårdal H (2004) *From Insect Parasitoid to Gall Inducers and Inquilines. Morphological Evolution in Cynipoid Wasps.* Ph.D. Thesis.

Vårdal H, Nieves-Aldrey JL (2005) Chapter 4. Oak Cynipid morphology and taxonomy. Immtature stages of oak gall wasps and their inquilines. In: *The Oak Cynipids of the Western Palearctic,* (ed. Stone G, Melika G, Csóka G). Ray Society *In press.*

Walker, F. (1874) The Devonshire gall, *Cynips kollari* (by the late F. Walker). *Entomologist* **9**, 52–54.

Walker P, Leather SR, Crawley MJ (2002) Differential rates of invasion in three related alien oak gall wasps (Cynipidae : Hymenoptera). *Diversity and Distributions* **8**: 335-349.

Wallis GP (1999) Do animal mitochondrial genomes recombine? *Trends in Ecology and Evolution* **14**: 209-210.

Walsh PS, METZGER DA, HIGUCHI R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**: 506-513.

Washburn JO (1984) Mutualism between a cynipid wasps and ants. *Ecology* **65**: 654–656.

Washburn JO, Cornell HV (1981) Parasitoids, patches, and phenology: their possible role in the local extinction of a cynipid gall wasp population. *Ecology* **62**: 1597–1607.

Weiher E, Keddy PA (1999) Relative abundance and evenness patterns along diversity and biomass gradients. *Oikos* **87**: 355-361.

Wiebes-Rijks AA, Shorthouse JD (1992) Ecological relationships of insects inhabiting cynipid galls. In: *Biology of Insect-Induced Galls*. (ed. Shorthouse JD, Rohfritsch O), pp. 238–257. Oxford University Press, New York.

Wesson DM, Porter CH, Collins FH (1992) Sequence and secondary structure comparisons of ITS rDNA in mosquitoes (Diptera: Culicidae). *Molecular Phylogenetics and Evolution* **1:** 253-269.

Wilcox TPD, Zwickl J, Heath T, Hillis DM (2002) Phylogenetic relationship of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Molecular Phylogenetics and Evolution* **25**: 361-371.

Wimp GM, Martinsen GD, Floate KD, *et al.* (2005) Plant genetic determinants of arthropod community structure and diversity. *Evolution* 59: 61-69.

Winnepenninckx B, Backeljau T (1996) 18S rRNA alignments derived from different secondary structure models can produce alternative phylogenies. *Journal of Zoolological Syststems and Evolutionary Research* **34**: 135-143.

White T, Bruns T, Lee S, Taylor J (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, In: *PCR protocols: A guide to methods and applications* (ed. Innis MA, Gelfand DH, Sninsky JJ, White TJ), pp. 315-322. San Diego, California, Academic Press.

Whittaker RJ, Araujo MB, Paul J *et al.* (2005) Conservation Biogeography: assessment and prospects. *Diversity and Distributions* **11**: 3-23.

Wilcox TP, Zwickl DJ, Heath TA, Hillis DM (2002) Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Molecular Phylogenetics and Evolution* **25**: 361–371.

Wu CI (1991) Inference of species phylogeny in relation to segregation of ancient polymorphisms. *Genetics* **127**: 429-435.

Yang Z (1993) Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Molecular Biology and Evolution* **10**: 1396-1401.

Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *Journal of Molecular Evolution* **39**: 306-314.

Yang Z (1996) Among-site rate variation and its impact on phylogenetic analysis. *Trends in Ecology and Evolution* **11**: 367-372.

Zanetto A, Kremer A (1995) Geographical structure of gene diversity in *Quercus petraea* (Matt.) Liebl. 1. Monolocus patterns of variation. *Heredity* **7**: 506517.

Zhang DX, Hewitt GM (1996) Nuclear integrations: challenges for mitochondrial DNA markers. *Trends in Ecology and Evolution* **11**: 247-251.

Appendix 1

Range of nuclear markers explored

Allozyme analysis

Allozyme analysis has been used successfully to generate nuclear marker data for gall wasps (Stone & Sunnucks 1993, Stone *et al* 2001, Atkinson, *et al* 2003). It is a quick and cost-effective method of generating data for multiple nuclear loci, and therefore this thesis began with an attempt to utilise the same allozyme loci that had proved useful in studies on gall wasps.

Methods

It is necessary to use specimens that have been preserved at -70°C or lower for successful allozyme analysis, since the proteins involved are highly unstable and decompose rapidly at higher temperatures. Cellulose acetate electrophoresis was adopted since it does not require a large amount of specimen material, which can act as a limiting factor when analysing chalcid wasps due to their small size. Only females were used, which are diploid in most chalcids, so results were comparable across specimens. Individual insects were homogenized in 50μL of homogenization buffer, 15μL of which was diluted with another 15μL of the buffer and 0.5-1μL of this was run on each gel. Staining reactions were as in Richardson *et al* (1986). The following running buffers were used depending on which loci were screened:

0.1M Tris malate (pH8.6) ethylenediaminetetraacetic acid (EDTA): Malate Dehydrogenase (MDH), Hexokinase (HK), Adenylate Kinase (AK), 6-Phosphogluconate Dehydrogenase (6PGD).

0.1M Tris glycine (pH8.6): Glucos-6-Phosphate Isomerase (GPI), Fumarate Hydratase (FUM).

40mM phosphate (pH6.1): mitochondrial Glutamate-Oxalate-Transaminase (GOT-m), Glycerol-3-Phosphate Dehydrogenase A (α-GPD1), Peptidase B (PEP-β,).

Results

In the initial stages of research for this thesis, a range of parasitoid species was targeted for study. These included: *Ormyrus nitidulus* and *O. pomaceus* (Ormyridae), *Sycophila biguttata* and *Eurytoma biguttata* (Eurytomidae), and *Torymus auratus, T. cyaneus Megastigmus stigmatizans* and *M. dorlsalis* (Torymidae). These species display a range of interesting biologies and allow comparisons within and between different genera and families to be made. In the end it was necessary to focus on a subset of these for various limiting factors.

For the allozyme analysis, I screened 50 specimens consisting of ten of each of the following species: *Ormyrus nitidulus, Sycophila biguttata, Torymus auratus, T. cyaneus* and *Megastigmus stigmatizans*. These specimens were assayed for variability in the following loci: 6PGD, GOT-m, PEP-β, AK, α-GPD1, and HK. Variation between species was present, however, very little within species variation was found.

22 specimens of *Megastigmus stigmatizans* sampled from across the European range were screened for 6PGD, MDH, GOT-m, PEP-β, HK, AK, α-GPD1, and FUM. These loci revealed very low variability between specimens: HK, PEP-β, α-GPD1 and MDH were fixed for a single allele across all specimens screened, while 6PGD, GOT-m, FUM, and AK displayed one, or in some cases two heterozygous specimens. In all cases the polymorphic specimens originated from Turkey.

Further allozyme analysis had to be abandoned due to the very limited availability of specimens suitable for testing. Most available parasitoid material had been preserved in alcohol and could not be used for allozyme analysis. Due to the low levels of diversity uncovered during the pilot study, attempts at collecting material with the use of allozymes in mind was not continued. This requires taking a cryoshipper into the field, which is only possible for a very limited period. Instead, focus was shifted towards the use of microsatellites and nuclear sequence markers. Both of which can be scored for alcohol preserved material.

Microsatellite loci

The value of microsatellite markers is discussed in Chapter 2. Microsatellite loci were investigated for two species: *Megastigmus stigmatizans* and *Eurytoma brunniventris*.

Megastigmus stigmatizans

Fluorescent dye-labelled primers were obtained for the primers described by Boivin *et al* (2003). Boivin *et al* (2003) developed a panel of six microsatellite loci in order to study the invasion pattern of the seed chalcid *Megastigmus spermotrophus*. They described the following six loci: MS1-alpha, MS1-43, MS2-89, MS2-162, MS3-148, MS3-180. All loci were reported to display high allelic diversity, with 17-30 alleles per locus. Furthermore, it was reported that all six loci were successfully amplified in 15 congeneric species. Since *M. stigmatizans* is also congeneric and success had been demonstrated for a range of similar species, these loci presented an obvious choice for analysis in *M. stigmatizans*.

A panel of *M. stigmatizans* specimens from across the European range were screened for polymorphism using these loci. The loci amplified well, however, virtually no polymorphism was detected, with most loci fixed for all specimens screened. Financial and temporal resources for the development of additional loci were not available. Therefore, focus was shifted onto *Eurytoma brunniventris* for which a group of polymorphic loci had recently been developed.

Eurytoma brunniventris

Microsatellite loci for *E. brunniventris* were for developed by Hale *et al* (2004) for use in the work of my supervisor Dr GN Stone and the gall wasp research group at the University of Edinburgh. The following nine polymorphic loci were described by Hale *et al* (2004): EB1, EB6, EB13, EB15, EB21, EB36, EB47, EB49, with 2-5 alleles per locus reported. Fluorescent dye-labelled primers were obtained for these loci, and a panel of *E. brunniventris* from across the European range were screened for polymorphism. However, results were extremely poor and it was clear that optimisation had not been carried out effectively. Attempts were made at optimisation; however, time prevented further progress from being made using these primers.

Nuclear sequence markers

Nuclear sequence markers are advantageous for the reasons discussed in Chapter 2. However, their use is often restricted by a lack of well-conserved primer sites across species, and limited mutation rates. Attempts to amplify nuclear sequence markers with sufficient variability for the purposes required in this thesis are summarised below.

The internal transcribed spacer regions (ITS1&2)

The nuclear ribosomal gene cluster contains the genes responsible for synthesizing the molecules that form the ribosomal subunits. Between these genes lie two internal transcribed spacers. These non-coding regions are free from functional constraint and have proven highly variable in a number of studies on insects (Campbell *et al* 1993, Beebe *et al* 1999, Rokas *et al* 2001), and are suggested as useful markers for lower-level phyologenies by Rokas *et al.* (2002).

A panel of *M. stigmatizans* specimens from across the European range were analysed for variation in ITS sequences using the primers ITS4/ITS5 from White *et al* (1990). Two internal primers (ITS5.8F/ITS5.8R) designed by Rokas *et al* (2002) were used for direct sequencing. PCR amplification was successful, however, sequencing results revealed a large amount of intra-individual variation. The rDNA array typically exists as a set of tandem repeats and the sequence in each repeat is usually believed to be the same due to gene conversion (Liao 2000). However, intra-genomic variation in ITS sequences has been detected previously in insects (Wesson *et al.* 1992, Rokas *et al* 2001, 2002), and including sequences containing such variation in phylogenetic analyses has been reported to cause confounding results (Buckler *et al.* 1997).

Due to the discovery of intra-genomic variation, multiple cloning of individual PCR products per specimen would have been necessary in order to obtain a clear signal after sequencing. Cloning large numbers of sequences is highly expensive and time consuming and the resources necessary to do so prevented further consideration of ITS here.

Long-wavelength opsin (LWRh)

The LWRh gene has been used in phylogenetic studies on both bees (Mardulyn & Whitfield

1999) and gall wasps (Rokas *et al* 2002, Cook *et al* 2002). Amplification of LWRh using the primers LWRhF/LWRhR from Mardulyn & Cameron (1999) and the PCR conditions recommended by the same authors was carried out. However, this proved unsuccessful and although attempts made to amplify under a range of PCR conditions were made, these failed also.