Community level consequences of adaptive management through Climate Matching: oak galls as a model system

Frazer H. Sinclair

Submitted for the degree of Doctor of Philosophy

University of Edinburgh

2011

Declaration

This thesis is submitted to the University of Edinburgh in accordance with the requirements for the degree of Doctor of Philosophy in the College of Science and Engineering. Aspects of the presented work were made possible by collaboration and data sharing with individuals and institutions, details of which are presented below.

Chapter 2.

The French National Institute for Agricultural Research (INRA) provided various phenotypic and genotypic data from oak provenance trials that are under their management. All presented analyses of these data are my own.

Chapter 3.

INRA allowed access to their established oak provenance trial at the forest of Petite Charnie in Sarthe, Northwest France. Insect surveys at the trial were conducted by me, and by volunteers under my supervision. All presented analyses of these data are my own.

Chapter 4.

Insect specimens were collected by me from the oak provenance trial at Petite Charnie with the permission of INRA. Approximately 1/3 of DNA extractions and PCR reactions were conducted by Konrad Lohse, Julja Ernst, and Juan Carlos Ruiz Guajardo. All presented analyses are my own.

Chapter 5.

Insect specimens were sourced from the Stone laboratory collections at the University of Edinburgh. Unpublished DNA sequence data from 6 parasitoid individuals were provided by Konrad Lohse. All presented analysis of this data is my own.

Unless otherwise stated, the remaining work and content of this thesis are entirely my own.

Signed:

Frazer Sinclair 31/10/2011

Acknowledgments

I am extremely grateful to my supervisors – Karsten Schonrögge, Graham Stone, and Stephen Cavers – for introducing me to the fascinating world of oaks and galls, and for their support throughout this project. I fear I may be hooked, and will still be dreaming of gallwasps long after the dust has settled on this thesis.

This project benefited greatly from collaboration with the French National Institute for Agricultural Research (INRA), and so massive thanks must go to Alexis Ducousso, Rémy Petit, Antoine Kremer, Sophie Gerber, and Stefanie Wagner, for providing data from the Petite Charnie provenance trial and for their hospitality during various visits to Pierroton.

At various stages of this project I have been aided by members of the international gall community, whose knowledge and enthusiasm for tiny wasps is inspiring. Many thanks therefore to George Melika, György Csoka, Dick Askew, Jose-Luis Nieves-Aldrey, Juli Pujade-Villar, Chang-Ti Tang, and Man-Maio Yang.

Having started this PhD is blissful ignorance of phylogenetics, particular thanks must go to James Nicholls and Konrad Lohse for patiently teaching me from scratch. Thanks also to past and present Stone lab members for their advice and friendship along the way: Jack Hearn, Pablo Fuentes-Utrilla, Alex Hayward, Julja Ernst, Juan Carlos Ruiz Guajardo, Tracey Begg, Riikka Kaartinen, Laura Riggi, Marlous de Boer, Kasia Mikolajczak, and Sarah White.

One million French galls require a lot of counting and I could not have done it without the help of the following fieldwork heroes: Kat Bradley, Julja Ernst, Matt Wainhouse, Phil Butterill, Stefan Curtis, Katherine Rushton, and even on occasion - Karsten Schonrögge. Thanks also to the fine folk of Bernay and Conlie – especially the Busson brothers, Philippe & Pierrette, and the Tabot family - who tolerated my shameful ignorance of the French language and made the time spent in France so enjoyable.

Finally, my deepest gratitude to my parents, whose patience and support in the past year has been amazing, and to whom I dedicate this thesis.

Abstract

In the present century, ecosystems across the globe will be subject to profound changes in climate. Forests are expected to be particularly sensitive to such change as the long life span of trees limits the potential for rapid adaptation. In order to preserve commercial viability and the essential ecosystem services provided by forests, there has been much interest in strategies for managing the adaptation of trees to their climatic environment. Climate Matching has emerged as one such strategy, whereby climate models are used to identify provenances – tree populations at a particular locality - with seed expected to be well adapted to the future conditions of a particular planting site. Debate continues about the feasibility and merit of this and other approaches, but it has yet to be demonstrated that the underlying assumptions of Climate Matching are valid for focal European tree species. Furthermore, a potentially major omission thus far has been consideration of how the Climate Matching strategy might influence associated organisms. Given the widely demonstrated bottom-up effects of foundation species genotype that have emerged from the field of community genetics, it is possible that planting seed of non-local provenance could effect forest organisms such as insect herbivores. In this thesis, I investigate the underlying assumptions of Climate Matching and its community level consequences using a model system of cynipid oak galls on *Quercus petraea*.

Following a general introduction to Climate Matching and the study system, in Chapter 2 I use data from a provenance trial of *Q. petraea* in France to explore a central assumption of the Climate Matching strategy: that provenances of focal tree species show climate associated variation in adaptive phenotypic traits. In Chapter 3, I explore correlations between these phenotypic traits and the abundance, diversity, and community composition of an associated guild of specialist gall-inducing herbivores. Tree phenological traits in particular showed strong patterns of adaptation to climatic gradients, and influenced the abundance and community structure of galling species. However, as the response to non-local tree provenances was not strongly negative, it was considered unlikely that mixed planting of local and Climate Matched provenances would have sever impact on the gallwasp community.

Having assessed the bottom-up effects of provenance phenotypic variation on the galling community, my ultimate aim is to extend analysis to include associated hymenopteran inquilines and parasitoids. However, interpretation of effects at this level is hindered by taxonomic uncertainty, with a growing appreciation that morpho-taxa may not represent independently evolving lineages (i.e. 'true' species). In Chapters 4 & 5 I therefore develop approaches for addressing taxonomic uncertainty with this ultimate aim in mind. In Chapter 4, I apply a DNA barcoding approach to parasitoid and inquiline specimens reared from the provenance trial, and compare taxa based on barcodes with those based on morphology to identify points of taxonomic uncertainty. I also investigate the extent to which networks based on morphological and molecular taxa support contrasting conclusions of network properties. In Chapter 5 I explore the potential for molecular based resolution of species level taxonomic error in a challenging group of parasitoids: the genus *Cecidostiba*. Beginning with a framework of single locus DNA barcoding, I use data from multiple nuclear loci to reveal the existence of cryptic species.

Finally, in Chapter 6 I explore the practicalities of Climate Matching in light of my empirical results, and suggest fruitful avenues for further research.

Contents

Chapter 1 - Introduction

As a result of anthropogenic activity the global climate is experiencing a period of rapid change, characterised by an increase in global surface temperature. The inevitable continuation of this change throughout the 21st century is expected to test the resilience of many ecosystems (IPCC 2007), and the maintenance of key ecosystem functions may require innovative strategies for intervention and management. The development of an appropriate knowledge base on the effects of climate change and the effectiveness of potential management strategies represent one of the most pressing challenges in modern ecology.

In this thesis I identify and explore important gaps in our knowledge of Climate Matching, a recently proposed strategy for the adaptive management of forests in the UK. I begin in this opening chapter by describing the challenges that climate change poses for UK forests, the rationale that underlies Climate Matching, and those aspects of the strategy for which important knowledge is lacking. I then provide details of my chosen study system – the oak gall community associated with *Quercus petraea* – and outline how it is used in subsequent chapters to investigate the identified issues.

1.1. Forests and climate change in the UK

Forest cover in the UK is estimated at around 2.8 million hectares, constituting 11.6% of total land surface area (Forestry Commission 2004). Coniferous trees such as Sitka spruce (*Picea sitchensis*) and various species of pine (*Pinus* spp.) make up approximately 60% of the total, with the other 40% consisting of broadleaved species - principally oaks (*Quercus petraea* and *Q. robur* – 9.4% of total area), birch (*Betula* spp. – 6.7% of total) and ash (*Fraxinus excelsior* – 5.4% of total area, (Forestry Commission 2003). This forest represents a valuable national resource, supplying timber, enhancing biodiversity, protecting soil and water quality, and providing amenity. The sustainable management of existing forest and the creation of further forests is recognised as an important objective, supported by national and European

policy (Forestry Commission 2004, DEFRA and Forestry Commission 2005, EC 2005).

Climate change in the UK is already apparent, with a detected increase of 1°C in the Central England Temperature (CET) since the 1970's, and most regions experiencing increased winter rainfall and decreased summer rainfall (Jenkins et al. 2008). Projections based on various carbon emission scenarios indicate that these trends will continue throughout the 21st century, with hotter drier summers and milder wetter winters expected across the UK and much of Europe (Giorgi and Coppola 2009, Jenkins et al. 2009). The magnitude of UK changes will be greatest in the south of England where, by the 2080's, under a medium carbon emissions scenario, 50% probability estimates indicate an increase of 4-6 °C in summer mean daily temperature, an increase of 2-4 °C in winter mean daily temperature, a decrease of 20-40% in summer rainfall, an increase of 10-30% in winter rainfall, and a decrease by as much as 18% in summer cloud cover (Jenkins et al. 2009).

These changes may have varied effects on UK forests (Table 1.1). While some aspects of change could be beneficial, with for example elevated temperatures and CO_2 concentrations potentially increasing productivity, the overall impact on forests is likely to be negative with increased risk of severe mortality from summer drought, fire, and pest, and disease outbreaks (Broadmeadow et al. 2003, Broadmeadow and Ray 2005, Broadmeadow et al. 2005). There has consequently been increasing interest in management strategies and silvicultural techniques that can promote the adaptation and resilience of forests in the face of climate change (Broadmeadow et al. 2005, Hubert and Cottrell 2007).

Options include: the adoption of continuous cover or mixed stand forestry, which may provide a greater variety of microclimatic conditions that buffer susceptible young trees from climatic extremes (Castro et al. 2004, Barsoum et al. 2009); restocking of woodland through regeneration rather than planting, involving much higher initial stocking densities with consequently higher selection pressure for well adapted genotypes; regular gap creation and restocking (either through planting or regeneration), allowing selection to act on fresh material throughout the course of climate change; and increasing the connectivity of forests, allowing for increased

gene flow between populations and maintains higher levels of genetic diversity upon which selection can act (Hubert and Cottrell 2007).

Table 1.1. Summary of how predicted changes in the UK climate may positively and negatively effect forests and woodlands (adapted from Table 2 of Broadmeadow et al. 2003).

Change	Positive effects	Negative effects
Increase in atmospheric CO_2	• Increase in growth rate • Reduced stomatal conductance and lower water use per unit leaf area	• Reduction in wood density resulting in poorer quality timber and greater pest susceptibility • Increase in leaf area resulting in higher wind resistance and water use
Increase in temperature	• Increase in productivity • Longer growing season • Lower risk of winter cold damage • Less snow damage	• Delayed hardening and earlier bud-burst resulting in increased risk of autumn and spring frost damage • Longer growing season reducing winter soil recharge period • Reduced winter mortality of insect and mammalian pests • Increased fecundity, rapid development, and spread of pest species
Reduction in summer rainfall and increase in winter rainfall	• Drier summers resulting in reduced intensity of foliar herbivores and pathogens	• Summer drought-induced mortality • Water-stressed trees more susceptible to pests and pathogens • Increase in frequency of forest fires • Winter water-logging limiting access for forest operations and resulting in fine root death, thus increasing susceptibility soil pathogens and summer drought
Reduction in cloud cover	• Increase in productivity	• Increased diurnal temperature range resulting in increased risk of frost damage

A further potentially more radical strategy is to change the way that seed is selected for planting. Based on the rationale that naturally distributed plant populations are well adapted to their environments, national guidelines currently advocate the planting of seed of local provenance (i.e. from within the same defined seed zone as the planting site), ideally sourced from stands with demonstrated high performance (Hebert et al. 1999, Samuel 2003, Forestry Commission 2004, Hubert 2005, Hubert and Cundall 2006). While this practice is likely to be effective under stable environmental conditions, relatively rapid changes in the environment could lead to maladaptation of naturally distributed populations, with consequent reduction in the suitability of locally sourced seed. The strategy of Climate Matching aims to counter this effect by identifying and planting seed from provenances that are well adapted to the predicted future climatic conditions of a planting site.

In the principal study of Climate Matching by Broadmeadow et al. (2005), analyses were performed for four UK sites based on the UKCIP02 climate change scenarios (Hulme 2002) and interpolated global surface climate data (New et al. 2002). Predicted changes in winter (November-April) and summer (May-October) conditions, under both low and high CO_2 emission scenarios by the 2050's and 2080's, were applied to current conditions at the planting sites to obtain predicted monthly values for mean temperature, precipitation, and diurnal temperature range. A climatic difference index that matched the predicted climate of planting sites with 50 kilometer grid squares across Europe by minimising the sum of squared differences in the three climate variables was used to identify the best matched 0.2% of grid squares for the four prospective planting sites. There was a clear tendency for matching with sites from lower latitudes, but the locations of matched sites varied considerably between the planting sites and between projection times. For example, Kelty in eastern Scotland was matched with areas of Ireland and western Britain by the 2050s, and with southern Brittany by 2080s, while Alice Holt in the south of England was matched with Brittany by the 2050s, and areas of Italy, Sardinia, and Greece by the 2080s (Figure 1.1).

Climate Matching is an intuitive concept, and the planting of mixtures of native and matched provenances has been suggested as a no-regret option for forest management, together with the strategies that promote naturally occurring variation and encourage natural migration (Broadmeadow and Ray 2005). However, closer consideration of its underlying assumptions in relation to literature from the fields of genecology and community genetics indicates that it may not be completely without risk of adverse effects. In the following subsections, I identify two important aspects of Climate Matching that are in need of further attention, and describe how they might be investigated using established forestry trials.

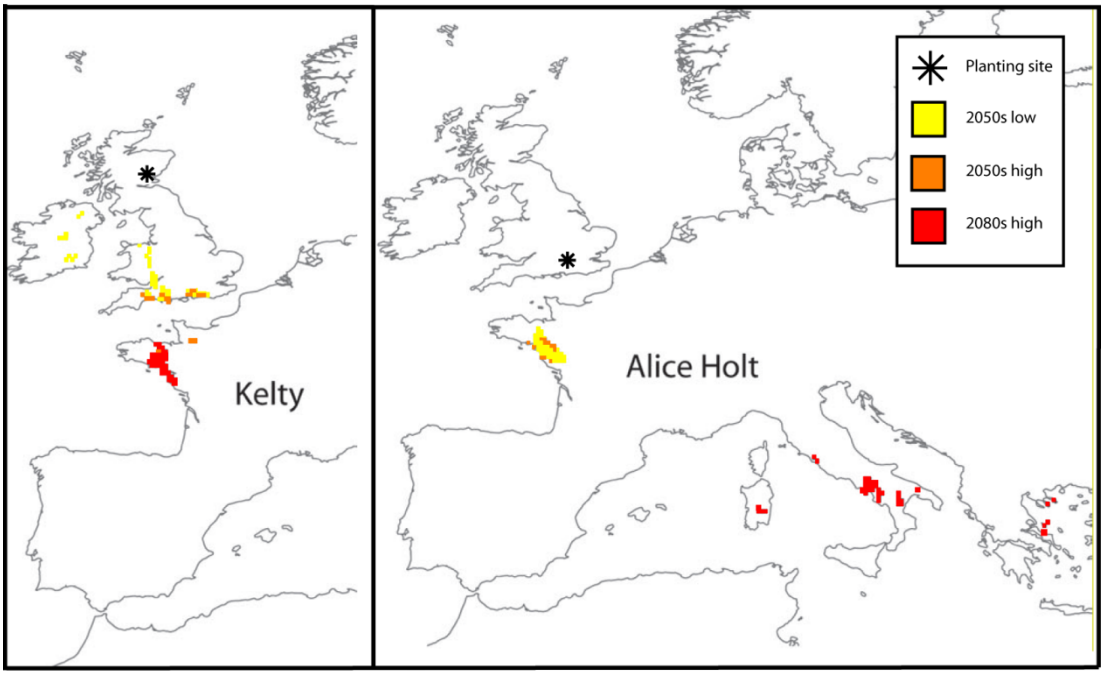


Figure 1.1. Illustration of Climate Matching analysis, showing the best matched 0.2% of 50 km grid-squares for planting sites at Kelty and Alice Holt, under low and high emission scenarios for the 2050s, and a high emissions scenario for the 2080s (adapted from Broadmeadow et al. 2005).

1.1.1. Local adaptation to climate

Through the process of natural selection, genotypes associated with relatively higher fitness under particular environmental conditions are expected to increase in frequency when and where those conditions arise. In the absence of constraints and opposing forces, variation in environmental conditions across the range of a species

can lead to differentiation of sub-populations as different genotypes are selected under different conditions, in a process that has come to be known as local adaptation (Levene 1953, Levins and MacArthur 1966, Hedrick et al. 1976, Kawecki and Ebert 2004). This process is important for many plant species, occurring in response to various abiotic and biotic influences including climate, soil conditions, and parasites (Sork et al. 1993, Linhart and Grant 1996, Kawecki and Ebert 2004, Macel et al. 2007, Wright 2007). However, local adaptation is not universal and its patterns can be complex, being constrained by available genetic variation and opposed by various ecological factors including high levels of gene-flow, temporal fluctuations in selective forces, differences in size and quality of habitats, and adaptive phenotypic plasticity (Kawecki and Ebert 2004). Even where a species exhibits strong patterns of local adaptation throughout a part of its range, these patterns may break down in particular regions, such as range margins, due to strong directional gene-flow (Savolainen et al. 2007).

By matching sites based on particular climate variables, such as the temperature, precipitation, and diurnal temperature range variables used by Broadmeadow et al. (2005), the effectiveness of Climate Matching hinges on the assumption that provenances of focal tree species are locally adapted to these aspects of their climatic environment. If this assumption is not met, either because local adaptation has not occurred or because it has occurred primarily in response to alternative influences, then Climate Matching is not likely to be successful in its objective of improving the resilience and adaptation of forests. Furthermore, if matched provenances are actually maladapted to their planting sites (e.g. through being less resistant to local pathogens), then there is a risk of out-breeding depression where gene-flow from maladapted individuals decreases the overall fitness of the local genepool (McKay et al. 2005). The question of whether provenances of focal tree species are locally adapted to their climatic environments is therefore central to Climate Matching, and it is important to establish the answer *a priori*, to avoid potentially negative consequences.

1.1.2. Community effects of tree provenance

Forests are biologically diverse ecosystems, containing multi-guilded communities that in addition to trees include a wide phylogenetic range of plants, microbes, and animals. The conservation and enhancement of this forest biodiversity is viewed as an essential element of sustainable forest management, and is reflected in UK and European policy (Forestry Commission 2004, EC 2005). The value of forest management strategies such as Climate Matching must therefore be considered not only in terms of their effectiveness in promoting forest adaptation and resilience, but also by any impacts they may have on associated biodiversity.

Trees play a key role in forest ecosystems, serving as 'foundation species' that structure the biotic environment for other forest organisms (Whitham et al. 2006). In recent years, research in the field of Community Genetics has widely demonstrated that variation in genetically controlled traits within foundation tree species can influence the structure of associated communities (Dungey et al. 2000, Wimp et al. 2005, Bangert et al. 2006), with effects that potentially span several trophic levels (Dickson and Whitham 1996, Bailey and Whitham 2003, Johnson 2008). Mechanisms for these effects include genetically determined variation in the concentration of plant defensive compounds such as tannins (Schweitzer et al. 2004, Bailey et al. 2005, LeRoy et al. 2006), plant vigour (Fritz and Price 1988), and plant phenology (Mopper and Simberloff 1995, Mopper 2005).

Given these widely demonstrated community level effects of plant genes, and that matched provenances must differ from local provenances in adaptive genetic traits for Climate Matching to be effective (as described in section 1.1.1), there is clearly the potential for Climate Matching to impact upon associated forest organisms. Empirical data are currently lacking, but the nature and extent of such impacts are likely to depend on precisely how matched provenances differ from local provenances, the relative abundance and spatial arrangement of introduced trees, and the ability of the native wildlife to respond to the novel trees (Hubert and Cottrell 2007). If differences are relatively subtle or associated organisms are plastic in their response, then the planting of matched and local provenances together could increase habitat heterogeneity and serve to promote biodiversity (Baldi 2008). If however the

differences are such that local organisms are less able to utilise introduced provenances, then the effect on biodiversity would be negative. Further investigation of just how matched provenances might influence associated organisms is of obvious relevance if the impacts of Climate Matching are to be accurately predicted and such a strategy implemented successfully.

1.1.3. Provenance research

As Climate Matching is a recently developed strategy, empirical data on its effectiveness and effects are currently not available. Experimental trials of matched provenances have been established at two sites in the UK midlands (N. Barosum, personal communication), and while these will be a valuable future resource, it will be some time before data are available. Fortunately, for almost 200 years foresters have been studying differences between tree provenances by establishing trials where trees of several provenances are grown together in a common environment and monitored for various traits, often with the intention of identifying well adapted provenances for commercial growth (Konig 2005). Although not necessarily designed for the purpose, these trials can offer a means for investigating both the patterns of local adaptation in focal tree species, and the effect of tree provenance on associated organisms.

The phenotype of individual organisms is determined by an interaction between their genotype and their environment. For provenance trials with adequate replication and blocking, environmental effects are controlled for so that observed variation in phenotypes relate directly to genotypic variation (Aitken 2004). The phenotypic variation can be analysed to identify the role of genetic differences between provenances, and also of provenance characteristics such as the environmental conditions at their source sites. Analyses of provenance trial data has traditionally involved analysis of variance (ANOVA) and linear regression (Konig 2005), but modern computing software for linear mixed-effect models allows for more appropriate treatment of nested and crossed experimental blocking effects, and unbalanced experimental data (Crawley 2007). Generalised versions of such models

allow for analysis of phenotypic traits that have non-normal errors, such as survival or count data (Bolker et al. 2009).

A common approach to investigating local adaptation from provenance trial data has been to compare the extent of differentiation between provenances for particular quantifiable phenotypic traits, with the variation in allele frequencies between provenances for neutral genetic markers. These can be described respectively by the statistics Q_{st} (Spitze 1993) and Wrights inbreeding coefficient F_{st} (Wright 1951, Weir and Cockerham 1984), and such comparisons are known as Q_{st}-F_{st} tests although they may in practice involve various statistical relatives of Q_{st} and F_{st} (Saether et al. 2007). Variation in allele frequencies between provenances for neutral markers arises through genetic drift, but if estimates of Q_{st} differ substantially from F_{st} then drift alone is not sufficient to explain the differentiation for the quantitative trait. Significantly higher values of Q_{st} suggest that the trait is experiencing spatially divergent selection and has become locally adapted, while significantly lower values of Q_{st} would suggest spatially stabilizing selection (Whitlock 2008).

While Q_{st}–F_{st} tests offer a useful means for exploring patterns of trait variation, the accurate estimation of both statistics is subject to various sources of bias that may be difficult to account for (Whitlock 2008), limiting the confidence in resulting inferences. A further potentially complementary approach to investigating spatially divergent selection is to relate the variation in particular traits observed within a provenance trial to gradients in environmental conditions between the sites of origin of the provenances. The detection of strong relationships, known as clines, suggests that the environmental factor is involved in driving local adaptation of the trait (Huxley 1955, Aitken 2004). In the context of Climate Matching, the identification of strong clines in adaptive phenotypic traits along gradients in temperature and precipitation would indicate that provenances are locally adapted to these aspects of their climatic environments.

With regards to investigating the community level effects of tree provenance, the interactions between trees and other organisms can be considered as an extended phenotype of the tree (Whitham et al. 2003), and can be analysed in much the same way as direct phenotypic traits such as tree size or phenology. Insect herbivores are a

convenient starting point for such investigation, as their interactions with trees can be easily established and quantified from field surveys within provenance trials. Variation in the abundance of individual species, or in multi-species parameters such as species richness or community similarity, can be analysed to assess the importance of differences between provenances. Where strong effects of provenance are apparent, further analysis in relation to direct phenotypic traits (i.e. tree size or phenology) may identify the mechanisms responsible for the effects of provenance. Such data on how and why herbivorous communities might vary in relation to tree provenance can be used to evaluate the impact that Climate Matching might have on associated biodiversity. Once appropriate datasets exist, this approach could be extended to consider wider community and ecosystem parameters (Whitham et al. 2006).

1.2. Study system

In order to explore these key issues empirically, I required a study system that: (i) centred on a tree species that is a candidate for Climate Matching, (ii) was represented in an appropriately scaled provenance trial, and (iii) involved a multi-trophic community of appropriate diversity that could be practically sampled. The combination of Sessile oak (*Quercus petraea*) and its associated oak gall community offered a suitable choice, being both a widespread and economically important European tree that is well represented in provenance trials, and a popular model system for field studies of community ecology. Here I provide details of the system and the selected study site.

1.2.1. *Quercus petraea*

Sessile oak (*Quercus petraea*) is a widespread European tree with a current natural distribution from Spain to Russia and from Scotland to Turkey, between sea level and 1600m elevation (Kleinschmit 1993). It is an economically important species and together with the closely-related Pedunculate oak (*Q. robur*), it represents almost 25% of broadleaved high forest in Great Britain (Forestry Commission 2003). The two species show some ecological differentiation, with *Q. robur* being more tolerant

of water-logging and favouring heavier alkaline soils, while *Q. petraea* occurs more commonly on acidic well-drained soils (Hubert 2005). Under recent climate change scenarios it has been predicted that oak productivity will increase in the north and west of the UK and decrease in the south and east, with increased risk of drought mortality (Broadmeadow et al. 2005). Given these predictions, and its economic significance, *Q. petraea* should be considered as a candidate species for Climate Matching, particularly in the south of the UK.

Following the last glacial period in the late Pleistocene era (approximately 10,000 years ago), *Q. petraea* is thought to have spread north from refugia in the Iberian peninsula, the Italian peninsula, and the Balkans, reaching its current distribution approximately 6000 years ago (Brewer et al. 2002, Petit et al. 2002). The genetic signature of refugial differentiation is observed in chloroplast genes (Le Corre et al. 1997, Kremer et al. 2002), but current patterns of nuclear genetic diversity are considered to also reflect a combination of the selection pressures acting on established populations, and wind mediated pollen flow between refugial lineages where they have met in central Europe (Kremer et al. 2002).

Quercus petraea is well represented in provenance trials, and the results of these indicate considerable variation in phenotypic traits between provenances. The phenological traits of budburst and leaf senescence have received particular interest due to their implications for frost resistance and the length of growing season. There is evidence of local adaptation in these traits from comparisons of F_{st} and Q_{st} (Kremer et al. 1997, Jensen and Hansen 2008), and from the identification of clines along gradients in latitude and altitude (Deans and Harvey 1995, Ducousso et al. 1996, Broadmeadow and Ray 2005, Alberto et al. 2011). Considerable phenotypic plasticity in the effect of temperature on budburst and leaf senescence has also been identified, with warmer temperatures resulting in earlier phenology for both traits (Vitasse et al. 2009, Vitasse et al. 2010). Variation in growth and architectural traits between provenances has also been demonstrated (Jensen 2000, Hubert 2005), again with evidence of local adaptation from comparisons of F_{st} and Q_{st} (Kremer et al. 1997, Jensen and Hansen 2008). However, in the context of Climate Matching it is the patterns of local adaptation of these traits along geographical gradients in

temperature and precipitation that are particularly relevant, and these have yet to be considered.

1.2.2. Petite Charnie provenance trial

The oak provenance trial in the forest of Petite Charnie in Sarthe, Northwest France (Figure 1.2), is one of four such trials established in France in the early 1990s by the French National Institute for Agricultural Research (INRA). They are intended as a resource for evaluating the range-wide genetic diversity of *Q. petraea*, to aid genetic conservation and management (Ducousso et al. 1996). The Petite Charnie trial contains almost 200,000 individual trees, representing 103 provenances of *Q. petraea* and 9 provenances of *Q. robur* spanning much of their natural range. For a putatively natural stand at each provenance, acorns were collected from at least 50 points with 30 meter spacing between them during 1986, 1987, 1989, and 1992, and were grown in the public nursery of Guemene-Penfao in Brittany, northwest France. At three years of age, all trees of a particular cohort were planted into a 'tranche' site that had been cleared and tilled during the previous year (A, Ducousso, personal communication). The trial thus contains four tranches, somewhat esoterically numbered 1, 2, 4, & 5, planted in the early months of 1990, 1991, 1993, and 1995 respectively. Each contains a unique combination of provenances and is further subdivided into several soil zones of approximately equal size, based on the soil description and associated plant communities prior to planting (Ducousso et al. 1996). Within each soil zone, provenances are represented by two or three distinct 'parcelles' of 24 trees each, planted in four rows of six trees, with spacing of 1.75 meters between trees and three meters between rows. Parcelles of 8 different randomly selected provenances were aggregated into blocks, with the position of blocks randomised within soil zones, to allow efficient statistical separation of parcelled, block, soil-zone, and provenance effects.

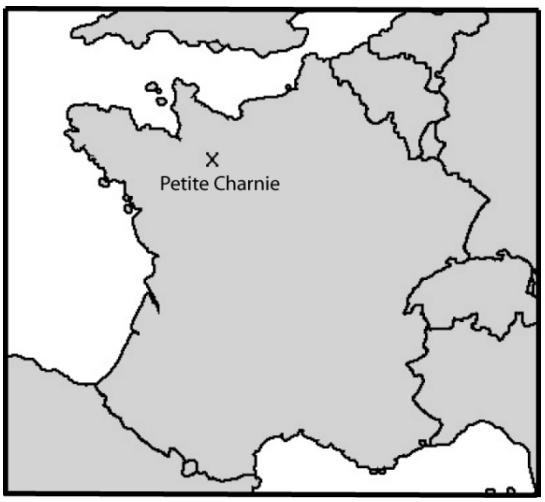


Figure 1.2. Location of the Petite Charnie provenance trials in Northwest France.

Petite Charnie has a mean elevation of 140m, and its climate is typically Atlantic, temperate and wet. The geological substratum is composed principally of red sandstone, schist and lens of clay (Bacilieri et al. 1995). The trial is surrounded by mature oak forest including both *Q. petraea* and *Q. robur*, mixed with beech (*Fagus sylvatica*), ash (*Fraxinus excelcior*) and hornbeam (*Carpinus betulus*).

Although *Q. petraea* is represented within various provenance trials in the UK, the trial at Petite Charnie was considered to have several advantages as a focal site for this study. Firstly, its scale is epic, both in terms of the number of trees and the variety of provenances, spanning a range of more than 2500 miles from Ireland to Georgia. Secondly, a wealth of existing data were made available by INRA, including measurements of tree phenotypic traits relating to phenology and growth taken at various times since establishment of the trial, and genotypic data (i.e. microsatellite markers) for a sample of trees from particular provenances. Thirdly the surrounding mature oak forest provides a source for populations of herbivores and other organisms, from which they may colonise the trial.

1.2.3. Oak gall communities

Oaks are important foundation species, supporting particularly rich phytophagous communities (Kennedy and Southwood 1984, Csóka 1998). In the Palaearctic region, these communities include approximately 200 species of oak gallwasps

(Hymenoptera; Cynipidae; Cynipini), the larvae of which induce complex galls within which they feed and pupate. The majority of these species have a complex cyclically parthenogenetic lifecycle with an alternating sexual generation that develops during spring, and an asexual generation that develops during the autumn of each year (Stone et al. 2002, Stone et al. 2008). Galls occur on various plant organs including buds, leaves, catkins, stems and roots, and may be single or multi-chambered (Stone et al. 2002, Csóka et al. 2005). Gall morphology and its location on the tree are generally diagnostic of a particular generation of a single species, and keys to Western Palaearctic species based on gall morphology are available (Buhr 1965, Ambrus 1974, Redfern and Shirley 2002). The distinctive morphologies and sessile nature of oak galls allow for straightforward identification and measurement of densities in the field, making them an attractive guild of herbivores for ecological study. They are also usually present in appropriate abundance and diversity for comparative study at various spatial scales, i.e. between sites (Schönrogge et al. 1995, Schönrogge et al. 1998, Schönrogge and Crawley 2000), or between individual trees within sites (Egan and Ott 2007, Kaartinen and Roslin 2011).

In addition to the gall-former, Western Palaearctic oak galls are often colonised by inquiline cynipids (Hymenoptera; Cynipidae; Synergini), whose phytophagous larvae are able to modify the tissue of existing galls but are unable to induce independently, and by hymenopteran parasitoids (of several families in the superfamily Chalcidoidea) that may feed on the larvae or pupae of the gall-formers, inquilines, or other parasitoids. Oak galls therefore encompass multi-trophic networks, with oak trees as primary producers, gallwasps and inquiline herbivores as primary consumers, and parasitoids and hyper-parasitoids as secondary and tertiary consumers (see Figure 1.3). These communities are relatively closed, in that individual species of oak gallwasp are specialised parasites of a limited range of oaks, and individual species of inquiline and parasitoid are generally specialised inhabitants of a limited range of oak galls (Askew 1961a, Askew 1980, Stone et al. 2002, Csóka et al. 2005).

Oak galls are a popular multi-trophic model system, and have recently been the focus of studies of biological invasions (Schonrogge et al. 1995, Schönrogge and Crawley

2000), comparative phylogeography (Hayward and Stone 2006), and habitat fragmentation (Kaartinen and Roslin 2011). If collected at an appropriately advanced phase of development the galls can be reared to establish associations with inquilines and parasitoids, and the closed nature of the communities means that populations can be studied by just sampling from oak, without the risk of bias from unsampled host plants. In the context of Climate matching, the *Q. petraea* oak gall community offers a means for investigating both how tree provenance may influence a guild of specialist herbivores, and also how these effects may cascade through a trophic association network.

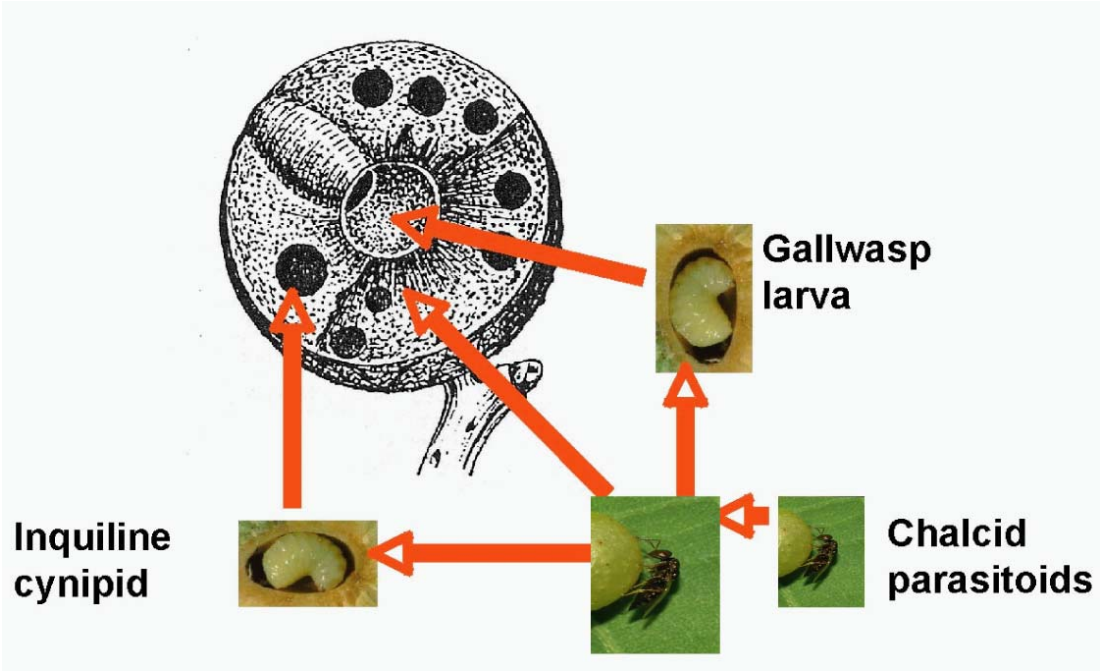


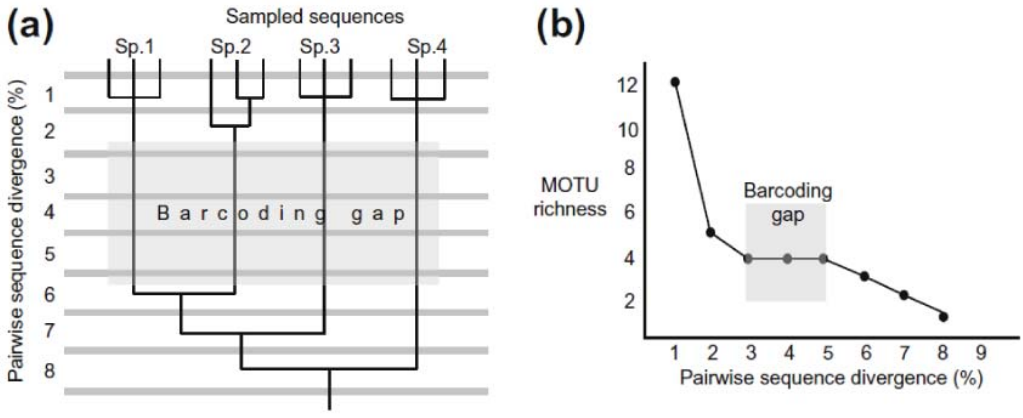
Figure 1.3. Illustration of a multi-trophic oak gall community, with the plant material (i.e. the gall) providing the primary resource, herbivorous gallwasp and inquiline larva feeding within chambers inside the gall, and parasitoids potentially targeting the gallwasp larva, the inquiline larva, or one another.

Western Palaearctic oak gall communities have a long history of detailed study (Askew 1961a, Askew 1962, 1980, Askew 1984, Schonrogge et al. 1995, Schönrogge and Crawley 2000, Csóka et al. 2005, Bailey et al. 2009), and the morphological taxonomy of gall inhabitants is well developed relative to other herbivorous insect guilds such as leaf-miners or externaly feeding caterpillars. However, integrated taxonomic approaches involving molecular techniques have recently revealed that all is not as it seems. While the gallwasp taxonomy has

remained relatively unchanged, due to the species specific structure of galls that provide reliable taxonomic characters (but see Stone et al. 2008 for the exceptional case of *Andricus burgundus*), assessment of the inquilines and parasitoids has revealed a high frequency of morphologically cryptic species (i.e. independently evolving lineages that are morphologically indistinguishable, Ács et al. 2007, Kaartinen et al. 2010, Nicholls et al. 2010). As species level taxa are a fundamental unit of ecological study, this potential taxonomic inaccuracy poses problems for studying the effect of host tree provenance on the multi-trophic oak gall community. Therefore, as an important prerequisite for further study, the later chapters of this thesis focus on using molecular techniques (particularly DNA barcoding - described in Box 1) to establish an accurate taxonomic framework for oak gall inquilines and parasitoids.

Box 1. DNA Barcoding as a tool for taxonomic assessment

The vast majority of known species level taxa have been described on the basis of differences in morphological characters, but there is a growing appreciation that such taxa may be discordant with modern species concepts that view 'existence as a separately evolving meta-population lineage' as the principal property of species (De Queiroz 2005, 2007). DNA barcodes - short sequences from a standardised region of DNA - can offer a means for assessing the accuracy of established morpho-species boundaries, based on the assumption that variation within species is less than and discrete from variation between species (Hebert et al. 2003). If a sample of barcode sequences from multiple species are grouped into molecular operational taxonomic units (MOTUs) based on their degree of sequence similarity (Blaxter et al. 2005), then the assumption that variation within species is less than variation between species will be characterised by a barcoding gap (Meyer and Paulay 2005, Ács et al. 2010). Where such a gap is apparent, MOTUs defined at thresholds within it are likely to represent meaningful independent lineages, and these MOTUs can be compared with morpho-species classifications to identify potential taxonomic error. MOTUs that contain all sequences from two or more distinct morpho-species are indicative of taxonomic over-splitting, whereas the presence of a single morpho-species in multiple MOTUs can be indicative of under-splitting (Ács et al. 2010).



An idealised illustration of the barcoding gap - reproduced with permission from Ács et al. (2010). (a) A barcode gene-tree where each of four species is monophyletic and the variation between species exceeds variation within species. (b) The relationship between the number of

1.3. Thesis outline

This thesis focuses primarily on the forest management strategy of Climate Matching, and on two questions that I consider to be important in guiding where and how it may be successfully implemented: (i) are provenances of focal tree species locally adapted to their climatic environment? And (ii) how might introduced tree provenances influence associated organisms? These questions are investigated empirically using a model system of oak trees (*Quercus petraea*) and gall-forming herbivores, with data collected from an established provenance at the forest of Petite Charnie. Ultimately I aim to extend analysis to include associated communities of gall inquilines and parasitoids, but this is currently impeded by taxonomic uncertainty. A secondary theme of this thesis is therefore to use molecular taxonomic methods to establish an accurate taxonomic framework for oak gall communities.

The effectiveness of Climate Matching in promoting the adaptation of forests in the face of climate change hinges on the assumption that provenances of focal tree species are locally adapted to particular aspects of their climatic environment. If this assumption is not valid, then Climate Matching is unlikely to be successful and could have negative consequences for the fitness of local tree populations. In Chapter 2, this assumption is investigated for *Q. petraea* by assessing whether a sample of 17 widely geographically separated provenances show evidence of local adaptation to the climatic environments of their source sites. Differentiation in various quantified phenological and growth traits is measured by a minimum estimate of the statistic Q_{st} (Spitze 1993), and is compared with differentiation in neutral genetic markers measured by the statistic R_{st} (Slatkin 1995), to look for evidence of spatially divergent selection (as would be suggested where $Q_{st} > R_{st}$). For traits showing such a pattern, mixed-effect models are used to model trait variation in relation to climatic and geographic predictor variables taken from the provenance source sites, including the measures of temperature and precipitation that feature in Climate Matching. Strong clines along gradients in temperature and precipitation are considered to indicate that local adaptation has occurred in response to these influences.

The enhancement of forest biodiversity is an important aspect of sustainable forest management, and strategies such as Climate Matching must be evaluated not only by

their effectiveness in promoting forest adaptation and resilience, but also by any impacts they may have on associated biodiversity. Chapter 3 focuses on this issue by investigating how host-tree provenance and phenotype within the Petite Charnie provenance trial influences the abundance, richness, and structure of the associated community of herbivorous gallwasps. Generalised linear mixed models are used to model field survey data collected over two years of study in relation to host tree provenance and various tree phenotypic traits. Particular consideration is given to the roles of host-tree vigour, host-tree stress, and phenological synchronisation in structuring the herbivore community. Patterns of variation in the herbivore community are used to evaluate the expected impacts of Climate Matching.

Accurate species level identifications are essential for the appropriate interpretation of ecological data and analyses of the effects of host-tree provenance on gall associated inquilines and parasitoids are currently impeded by taxonomic uncertainty in these groups. Chapter 4 focuses on this issue by generating DNA barcodes for all gall inquilines and parasitoids reared from one year of study at the Petite Charnie provenance trials. The presence of a barcoding gap is investigated, and taxa based on barcodes are compared with morphological taxa to identify points of discordance. Various properties of ecological networks based on morphological and barcode identifications are compared to assess bias if reliant solely on morphological identifications.

While DNA barcodes can be useful for highlighting potential taxonomic error and developing taxonomic hypotheses, monophyly at the barcode locus alone is generally not considered as sufficient evidence for making taxonomic inferences. However, concordant patterns of monophyly at additional molecular markers can provide further support for taxonomic distinctiveness (Rosenberg 2007). In Chapter 5, a DNA barcoding approach is extended for gall parasitoids from the genus *Cecidostiba,* by incorporating data from 10 nuclear loci. Patterns of monophyly across loci are analysed to provide statistical support for taxonomic hypotheses drawn from barcode data.

In Chapter 6, I conclude by discussing how my empirical results might influence future decisions about Climate Matching. Finally, I suggest what I consider to be valuable avenues for future research.

Chapter 2 - Are provenances of *Quercus petraea* locally adapted to climate?

2.1. Introduction

The global climate is warming at a rate that is likely to exceed the natural resilience of many ecosystems (IPCC 2007). Forests are expected to be particularly sensitive to such climate change as the long life span of trees limits the potential for rapid adaptation (Lindner et al. 2010). There has consequently been much interest in adaptive management strategies that can preserve the productivity and ecosystem services of forests in the face of rapid climate change (Spittlehouse and Stewart 2003, Broadmeadow et al. 2005, Millar et al. 2007, Aitken et al. 2008, Bower and Aitken 2008, Bolte et al. 2009).

Following the increased availability of sophisticated regional climate models (Hulme 2002, Jenkins et al. 2009), Climate Matching has emerged as a potential strategy for promoting the adaptation of forests to the expected changes in climate. Developed by Broadmeadow et al (2005), Climate Matching involves the use of climate models to predict the future conditions of a planting site under various climate change scenarios, and to identify sites that have recently experienced similar conditions (Broadmeadow et al. 2005, Bolte et al. 2009). Based upon the assumption that tree populations are locally adapted to their climatic environments, climate matching offers a means for guiding the selection of seed that will perform well under future climates.

2.1.1. Local adaptation

Through the process of natural selection, genotypes that are associated with relatively high fitness under particular environmental conditions are expected to increase in frequency when and where those conditions arise. In the absence of constraints and opposing forces, variation in environmental conditions across the range of a species can lead to differentiation of sub-populations, as different genotypes are selected under different conditions. This process is known as local

adaptation and is of demonstrated importance for many plant species, occurring in response to various abiotic and biotic influences including climate, soil conditions, and parasites (Sork et al. 1993, Linhart and Grant 1996, Aitken 2004, Kawecki and Ebert 2004, Macel et al. 2007).

The potential significance of local adaptation for silviculture has long been recognised, and is widely studied through the establishment of provenance trials. In such trials, trees of various provenances - geographic locations from which tree seed or cuttings are collected - are planted together at one or more trial sites, and quantifiable traits relating to tree health and productivity are monitored (Aitken 2004). The intention of these trials has often been to identify provenances that would perform well for commercial purposes, but they have also come to be recognised as a valuable resource for studying forest genetics and adaptation (Ducousso et al. 1996, Konig 2005, Savolainen et al. 2007). Whilst local adaptation has been shown to be a widespread and important process for many forest tree species, it is not universal and its patterns can be complex. It is constrained by available genetic variation, and is opposed by various ecological factors including high levels of gene flow, temporal fluctuations in selective forces, differences in size and quality of habitats, and adaptive phenotypic plasticity (Kawecki and Ebert 2004). Even where a species exhibits strong patterns of local adaptation throughout a part of its range, these patterns may break down in particular regions, such as range margins, due to strong directional gene flow (Savolainen et al. 2007). Such variation in patterns of local adaptation limits the potential for extrapolating the results of provenance trials between species (Aitken 2004), or beyond the provenances represented in trials.

A common approach to investigating local adaptation from provenance trial data has been to compare the extent of differentiation between provenances for particular quantifiable traits, with the variation in allele frequencies between provenances for neutral genetic markers (Whitlock 2008). These can be described respectively by the statistics Q_{st} (Spitze 1993) and Wrights inbreeding coefficient F_{st} (Wright 1951, Weir and Cockerham 1984), and such comparisons are known as Q_{st}-F_{st} tests although they may in practice involve various statistical relatives of Q_{st} and F_{st} (Saether et al. 2007). Variation in allele frequencies between provenances for neutral markers arises

through genetic drift, but if estimates of Q_{st} differ substantially from F_{st} then drift alone is not sufficient to explain the differentiation for the quantitative trait. Significantly higher values of Q_{st} suggest that the trait is experiencing spatially divergent selection and has become locally adapted, while significantly lower values of Q_{st} suggest spatially stabilizing selection (Whitlock 2008).

While Q_{st}–F_{st} tests offer a useful means for exploring patterns of trait variation, the accurate estimation of both statistics is subject to various sources of bias that may be difficult to account for (Whitlock 2008), limiting the confidence in resulting inferences. A further potentially complementary approach to investigating spatially divergent selection is to relate the variation in particular traits observed within a provenance trial to gradients in environmental conditions between the sites of origin of the provenances. The detection of strong relationships, known as clines, can identify the environmental factors responsible for driving local adaptation. Where details of environmental conditions at provenance sites are lacking, geographical variables (i.e. latitude, longitude, and altitude), which are expected to correlate broadly with many environmental gradients, can be used as surrogates (Aitken 2004).

2.1.2. Climate Matching in the UK

In the UK, there has been an increase of 1°C in the Central England Temperature (CET) since the 1970's, with most regions experiencing an increase in winter rainfall and a decrease in summer rainfall (Jenkins et al. 2008). The recently released UKCP09 projections, which consider various atmospheric response variables under three scenarios of carbon emission by the 2050's and 2080's, indicate that further temperature increases are expected with hotter, drier summers and milder, wetter winters. The magnitude of these predicted changes is greatest in the south of England where, by the 2080's, under a medium carbon emissions scenario, 50% probability estimates indicate an increase of 4-6 °C in summer mean daily temperature, an increase of 2-4 °C in winter mean daily temperature, a decrease of 20-40% in summer rainfall, an increase of 10-30% in winter rainfall, and a decrease by as much as 18% in summer cloud cover (Jenkins et al. 2009).

In the principal study of Climate Matching by Broadmeadow et al. (2005), analyses were performed for four UK sites based on the UKCIP02 climate change scenarios (Hulme 2002) and interpolated global surface climate data (New et al. 2002). Predicted changes in winter (November-April) and summer conditions, under both low and high carbon emission scenarios by the 2050's and 2080's, were applied to current conditions at the sites to obtain predicted monthly values for mean temperature, precipitation and diurnal temperature range. A climatic difference index that matched the predicted climate of planting sites with 50 kilometer grid squares across Europe by minimising the sum of squared differences in climate variables was used to identify the best matched grid squares for the four prospective planting sites. There was a clear tendency for matching with sites from lower latitudes, but matched locations varied considerably between planting sites and between projection times. For example, Kelty in eastern Scotland was matched with areas of Ireland and western Britain by the 2050's, and with southern Brittany by 2080's, while Alice Holt in the south of England was matched with Brittany by the 2050's, and areas of Italy, Sardinia, and Greece by the 2080's.

Climate Matching is an intuitive concept, and the planting of mixtures of native and matched provenances has been suggested as a 'no-regret' option for forest management (Broadmeadow and Ray 2005). However, by matching sites based on particular climate variables - such as temperature, precipitation, and diurnal temperature range - the effectiveness of Climate Matching hinges on the assumption that provenances of focal tree species are locally adapted to these aspects of their climatic environment. If this assumption is not met, either because local adaptation has not occurred or because it has occurred primarily in response to alternative influences, then Climate Matching is not likely to be successful in its objective of improving the adaptation of forests to future climates. Furthermore, if matched provenances are actually maladapted to their planting sites (e.g. through being less resistant to local pathogens), then there is a risk of out-breeding depression where gene-flow from maladapted individuals decreases the overall fitness of the local gene-pool (McKay et al. 2005).

The question of whether provenances of focal tree species are locally adapted to their climatic environments is therefore central to Climate Matching, and it is important to establish the answer *a priori*, to avoid potentially negative consequences. Existing studies can provide some indication of patterns of local adaptation for various traits, but empirical data on the role of geographical gradients in temperature and precipitation are currently lacking for most European tree species. In this study, I make use of an established provenance trial to investigate this issue for *Quercus petraea* – a widespread and commercially important European tree species.

2.1.3. *Quercus petraea*

Sessile oak (*Quercus petraea*) is a deciduous broadleaved tree with a current natural distribution from Spain to Russia and from Scotland to Turkey, between sea level and 1600m elevation (Kleinschmit 1993). It is an economically important species and together with the closely related Pedunculate oak (*Q. robur*), it represents almost 25% of broadleaved high forest in Great Britain (Forestry Commission, 2003). Under recent climate change scenarios it was predicted that oak productivity will increase in the north and west of the UK and decrease in the south and east, with increased risk of drought mortality (Broadmeadow et al. 2005). Given these predictions, and its economic significance, *Q. petraea* should be considered as a candidate species for climate matching, particularly in the south of the UK.

Following the last glacial period in the late Pleistocene era (aproximately 10,000 years ago), *Q. petraea* is thought to have spread north from refugia in the Iberian peninsula, the Italian peninsula, and the Balkans, reaching its current distribution approximately 6000 years ago (Brewer et al. 2002, Petit et al. 2002). The genetic signiture of refugial differentiation is observed in chloroplast genes (Le Corre et al. 1997, Kremer et al. 2002), but current patterns of nuclear genetic diversity are considered to also reflect a combination of the selection pressures acting on established populations, and wind mediated pollen flow between refugial lineages where they have met in central Europe (Kremer et al. 2002).

Quercus petraea is well represented in provenance trials, and the results of these indicate considerable variation in phenotypic traits between provenances

(Kleinschmit 1993). The phenological traits of spring growth initiation and autumn growth cessation have received particular interest due to their implications for frost resistance and the length of growing season. There is evidence of local adaptation in these traits from comparisons of F_{st} and Q_{st} (Kremer et al. 1997, Jensen and Hansen 2008), and from the identification of clines along gradients in latitude and altitude (Deans and Harvey 1995, Ducousso et al. 1996, Broadmeadow and Ray 2005, Alberto et al. 2011). Considerable phenotypic plasticity in the effect of temperature on budburst and growth cessation has also been identified, with warmer temperatures resulting in earlier phenology for both traits (Vitasse et al. 2009, Vitasse et al. 2010). Variation in growth and architectural traits between provenances has also been demonstrated (Jensen 2000, Hubert 2005), again with evidence of local adaptation from comparisons of F_{st} and Q_{st} (Kremer et al. 1997, Jensen and Hansen 2008). However, in the context of Climate Matching it is the patterns of local adaptation of these traits along geographical gradients in temperature and precipitation that are particularly relevant, and these have yet to be specifically considered.

2.1.4. Objectives

Where Climate Matching is based on predicted changes in temperature and precipitation, its potential for promoting adaptation of forests to future climates hinges on the assumption that tree provenances are locally adapted to the temperature and precipitation regimes of their provenance sites. The purpose of this study is to explore this issue by assessing patterns of local adaptation in *Q. petraea.* Specifically I investigate: (i) whether various quantifiable phenotypic traits relating to tree phenology and growth show evidence of local adaptation; (ii) whether locally adapted phenotypic traits show clines across gradients in temperature and precipitation; (iii) how such clines compare to other more general environmental gradients (i.e. latitude and longitude)? Based on the answers to these questions, the practicality of Climate Matching for *Q. petraea* is discussed.

2.2. Materials and methods

2.2.1. Petite Charnie provenance trial

The oak provenance trial in the forest of La Petite Charnie in Sarthe, Northwest France, was established in the early 1990s by the French National Institute for Agricultural Research (INRA) to aid genetic conservation and management of European white oaks (Ducousso et al. 1996). The trial contains 103 provenances of Sessile Oak (*Q. petraea*) and 9 provenances of pedunculate oak (*Q. robur*), planted in four age cohorts (further details of the trial design and establishment are provided in Chapter 1). To avoid the complication of comparing trees of different ages, a single age cohort (Tranch 4) was selected for further study. Tranche 4 is the largest at Petite Charnie, containing 680 parcelles that represent 57 different provenances (Figure 2.1). A set of 17 provenances of *Q. petraea,* for which genotypic data was also available, was selected from Tranche 4 to encompass the largest possible range of geographic distances from the trial site (Figure 2.2).

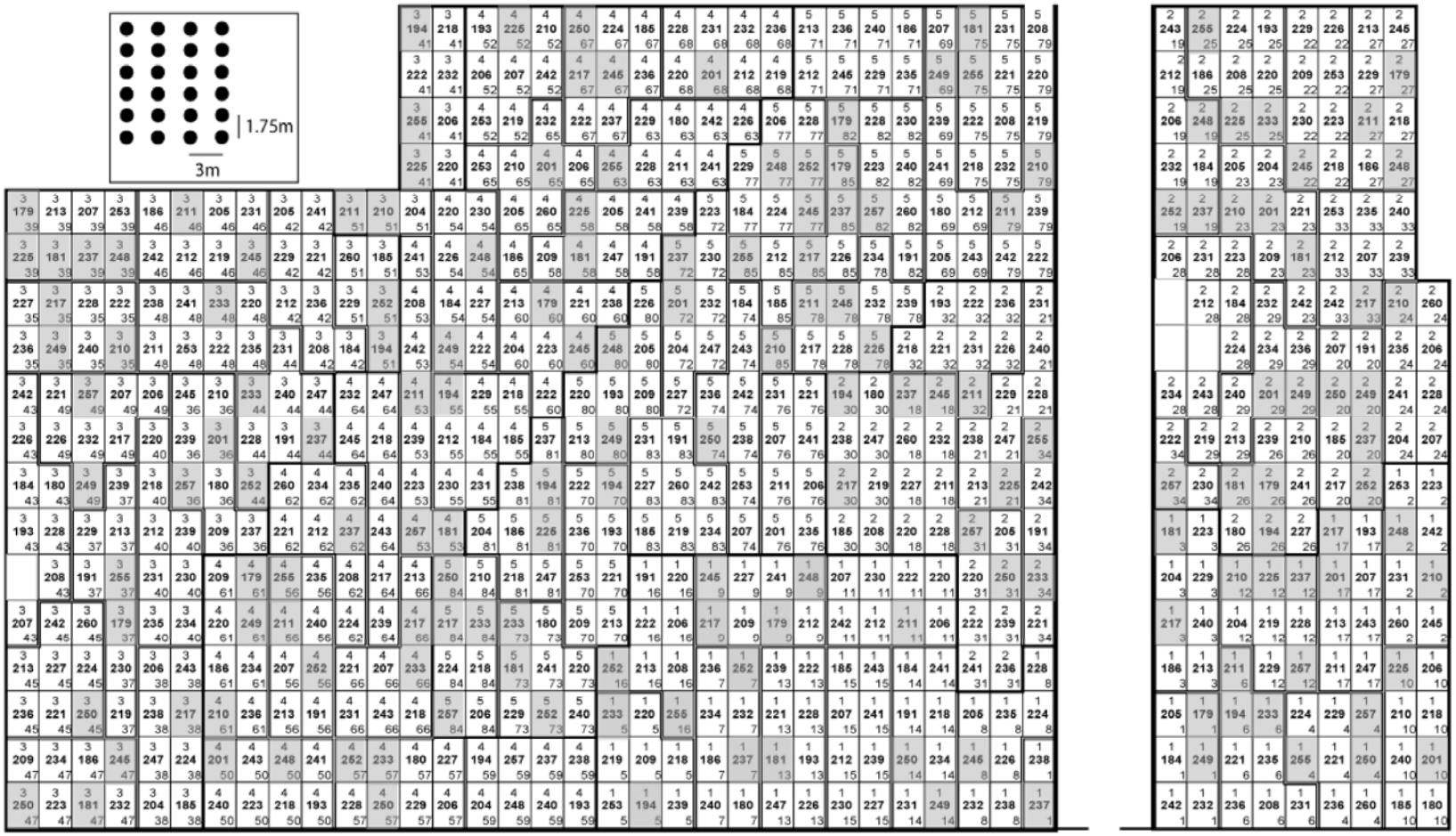


Figure 2.1. The layout of Tranche 4 at the Petite Charnie provenance trials. Each of the 680 squares represents a parcelle of 24 trees, identified by a soil zone number (1-5), a 3 digit provenance code, and a block number (1-85). Provenance codes follow Table 2.1. Parcelles of the studied provenances are shaded grey. The box in top left shows the arrangement of individual trees within a parcelle.

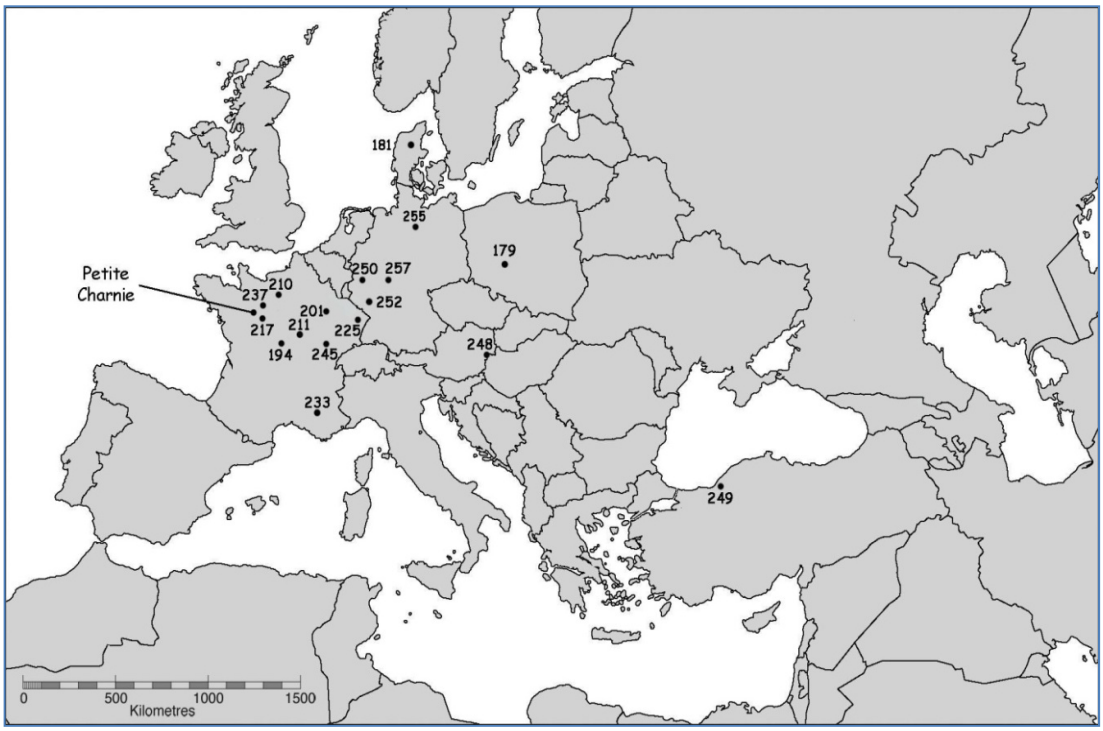


Figure 2.2. Location of the Petite Charnie provenance trial and the 17 study provenance sites. Provenance codes follow Table 2.1

2.2.2. Provenance source site climate data

For the selected provenances, data on precipitation and temperature were obtained from the WorldClim database (Hijmans et al. 2005), which interpolates to 1-km resolution based on weather station records from 1950-2000. Data were available for each month of the year based on averages across all available years. In the climate matching study of Broadmeadow et al (2005), sites were matched in terms of monthly precipitation, mean temperature, and diurnal temperature range, with the recent data for planting sites adjusted to incorporate predicted changes in summer (May-October) and winter (November-April) climate. In order to reflect the methods of Broadmeadow et al. (2005) while at the same time limiting the number of variables to be considered (thus avoiding over-parameterisation of models), monthly climate data for the study provenances were either summed (for precipitation) or averaged (for temperatures) to give the following summary variables:

- *SummerPrec* – The total precipitation (in mm) falling between May and October.

- *WinterPrec* – The total precipitation (in mm) falling between November and April.
- *SummerTemp* – Mean temperature (in C °) between May and October
- *WinterTemp* – Mean temperature (in C °) between November and April

Values of these environmental variables for each provenance are shown in Table 2.1. Diurnal temperature range could not be appropriately summarised in this way, and was therefore not considered in the analysis.

2.2.3. Tree phenotypic data

Measurements of the following phenotypic traits were provided by INRA for all trees of the study provenances in Tranche 4:

- *Budburst* – a measure of the timing of spring bud-burst, assessed in spring 1995 following a 6 stage scoring system (see Appendix 2.1.1). A high score represents early bud-burst.
- *Retention* – a measure of the timing of autumn leaf-fall, assessed autumn 2001 following 6 stage scoring system (see Appendix 2.1.2). A high score represents late leaf-fall.
- *Ht96* – top height in centimetres measured during winter 1996-97
- *Ht2001* – top height in centimetres measured during winter 2001-02
- *DBH* – the stem diameter in centimetres at a height of 1.3 m from the ground, measured during winter 2001-02.
- *Form* – a measure of overall tree shape, assessed during winter 2001-02 following a 10 stage scoring system (see Appendix 2.1.3). A high score indicates a well-formed tree.
- *NoBranches* – the number of branches, assessed during winter 2001-02 (see Appendix 2.1.4)
- *NoForks* – the number of forks in the stem, assessed during winter 2001-02 (see Appendix 2.1.5)

These phenotypic traits relate to various aspects of tree growth and health, and are widely used in the silvicultural industry. The phenological traits of *Budburst* and

Retention are considered to be particularly important for temperate and boreal trees, as they determine the length of growing season and influence the risk of cold injury, often showing strong patterns of local adaptation (Aitken et al. 2008). The traits relating to tree size (*DBH*, *Ht96*, and *Ht2001*) and architecture (*Form*, *NoBranches*, and *NoForks*) directly influence the value of timber and so are of particular interest for commercial forestry. Mean values of these traits for each study provenance are shown in Table 2.1.

Table 2.1. Summary of provenance source site and trial phenotypic data, showing: (a) the locality, geographic co-ordinates, and climate data for the 17 study provenances and the trial site, and (b) mean phenotypic trait values for all trees of each of the study provenances in Tranch 4 of the provenance trials.

	(a)								(b)							
Code	**Forest**	**Country**	**Longitude (DD)**	**Latitude (DD)**	***SummerPrec* (mm)**	***WinterPrec* (mm)**	***SummerTemp* (°C)**	***WinterTemp* (°C)**	***Budburst***	***Retention***	***Ht96***	***Ht2001***	***DBH***	***Form***	***NoBranches***	***NoForks***
179	Sycow	Poland	17.93	51.18	372	202	14.5	1.87	1.39	2.48	119	315	105	4.76	13.9	0.67
181	Horbylunde	Denmark	9.41	56.13	419	355	12.6	1.81	0.88	3.16	113	280	93.7	4.15	14.5	0.94
194	Soudrain	France	2.38	46.95	377	348	16.2	6.08	1.18	2.69	120	319	111	4.25	13.2	0.99
201	La Haie Renaut	France	4.95	48.67	369	310	15.2	4.79	1.64	1.80	123	325	108	4.18	16.0	0.77
210	Saint Germain	France	2.08	48.90	335	315	15.3	5.57	1.53	2.75	127	322	105	4.02	15.1	0.78
211	Prémery	France	3.60	47.20	402	350	15.7	5.51	1.47	2.58	120	319	102	4.31	14.5	0.82
217	Bercé	France	0.39	47.81	331	378	15.5	5.89	1.70	2.44	115	299	97.9	4.28	13.9	0.74
225	Still	France	7.25	48.58	460	386	13.8	2.72	1.42	2.03	116	294	98.8	4.54	14.5	0.69
233	Vachères	France	5.63	43.98	377	402	16.1	5.46	3.73	1.33	114	310	94.2	4.00	12.1	1.08
237	Réno Valdieu	France	0.67	48.50	332	354	14.9	5.51	1.49	2.75	124	329	114	4.19	14.1	0.82
245	Etangs	France	4.96	46.93	415	353	16.6	5.48	2.33	2.28	124	325	114	4.11	15.8	0.87
248	Klostermarienberg	Austria	16.57	47.41	424	219	15.7	2.85	3.16	1.65	113	317	106	3.88	13.2	0.94
249	Bolu	Turkey	31.67	40.92	281	470	13.8	2.13	1.63	0.45	104	288	89.6	4.30	13.4	0.82
250	Cochem	Germany	7.05	50.08	373	322	14.2	3.58	1.77	2.35	129	328	114	4.36	14.7	0.89
252	Johanneskreuz	Germany	7.83	49.40	414	363	13.5	2.5	1.06	2.34	127	322	114	4.44	14.5	0.99
255	Spakensehl	Germany	10.6	52.80	367	290	14.0	2.78	0.57	2.68	114	305	101	4.62	15.1	0.72
257	Wolfgang	Germany	9.05	50.15	360	281	15.6	4.17	1.54	2.38	122	301	97.5	3.92	12.4	0.91
	Petite Charnie	France	0.17	48.09	329	381	15.5	6.07								

2.2.4. Tree genotypic data and analysis

Nuclear genetic data for a sample of trees from 17 provenances were provided by INRA. Samples ranged in size from 21 to 29 trees. Data were in the form of genotypes for individual trees across 10 co-dominant diploid microsatellite loci, with alleles defined by their fragment lengths. These were determined by multiplex PCR reactions, conducted by an INRA researcher at the University of Bonn. The genotyped trees were selected from the provenance trial at Sillegny in Northeast France that was established concurrently with the trial at Petite Charnie. As the trees at these two trials (and at a further two trials) were selected randomly from a single pooled seed collection from each provenance, the genotyped trees are considered to be a representative sample of each provenance.

Tests for deviations from Hardy-Weinberg equilibrium were conducted for each locus in each population, and for all loci combined in each population, using the program FSTAT version 2.8.3.2 (Goudet 1995). Tests for linkage disequilibrium for each pair of loci in each population were also conducted in the same program. All loci have been tested for null alleles through progeny testing by researchers from INRA, and their effect in this dataset is considered to be negligible.

The genetic structure of the provenances was examined by calculating the following summary statistics:

- *No. of alleles & mean allelic richness* – The total number of alleles present at each locus across all populations, and the mean number of alleles per population, calculated using FSTAT version 2.9.3.2.
- H_o & H_e – The observed and expected levels of heterozygosity respectively, averaged across populations, calculated using the program using GenAlEx version 6.3 (Peakall and Smouse 2006).
- F_{it} , $F_{st,}$ & F_{is} - Wrights F-statistics (Wright 1951) estimated for each loci and summarised across loci following Weir and Cockerham (1984) using FSTAT version 2.9.3.2. Confidence intervals for global estimates were based on 1000 bootstrap replicates with re-sampling of loci.

- R_{st} – Measure of population subdivision based on microsatellite allele frequencies with a stepwise mutation model (Slatkin 1995), estimated for each locus and summarised across loci using FSTAT version 2.9.3.2.
- D_{est} – Jost's measure of genetic differentiation between populations (Jost 2008), estimated for each locus and summarised as the approximated harmonic mean across loci using the program SMOGD version 1.2.5 (Crawford 2010). Confidence intervals for each locus were based on 1000 bootstrap replicates, with re-sampling of individuals.

To summarise the relationship between provenances, a matrix of pair-wise D_{est} for all 17 provenances was generated using SMOGD version 1.2.5. Principal coordinate analysis (PCoA) was performed on this D_{est} matrix to identify the major axes of variation using GenAlEx version 6.3.

2.2.5. Phenotypic differentiation between provenances

The statistic Q_{st} offers a measure of the differentiation between populations for a quantitative trait, as described by the equation:

$$Q_{st} = \frac{V_p}{V_p + 2\,V_a} \qquad Equation\ 2.1$$

Where V_p is the population variance for the trait, and V_a is the additive genetic variance (Spitze 1993, O'Hara and Merila 2005). Q_{st} is most appropriately determined from controlled breeding experiments in a common garden or reciprocal transplant setting where V_p can be estimated as the variance component of the population effect, and V_a can be estimated as four times the variance component of the half-sibling family effect (O'Hara and Merila 2005, Jensen and Hansen 2008). Alternatively V_a can be estimated as the narrow sense heritability of the trait (h^2), multiplied by the within population variance (V_{wp}, Kremer et al. 1997). As such, Q_{st} can only be estimated directly when the design of an experiment involves pedigree information, or where the h^2 of a particular trait is known for the populations under consideration. However, in the absence of such data, an estimate of pseudo-Q_{st} (P_{st}) can be obtained by making certain assumptions (Saether et al. 2007). If h^2 is set to 1,

and environmental influences can be accounted for, such as in a common-garden or provenance trial setting with adequate blocking, P_{st} can be described by the equation:

$$P_{st} = \frac{V_p}{V_p + 2\,V_{wp}} \qquad Equation\ 2.2$$

As true values of h^2 will in practice always be less than 1, this P_{st} represents a minimum estimate of true Q_{st}. Hence, when compared to a measure of variation in allele frequencies it can provide evidence for spatially divergent selection because if P_{st} is greater than F_{st}, Q_{st} must also be greater than F_{st}. However, as the extent to which Q_{st} is greater than P_{st} cannot be established, instances where P_{st} is not significantly greater than F_{st} cannot be considered as evidence for stabilising selection ($Q_{st} < F_{st}$), or for drift ($Q_{st} \approx F_{st}$). It is also not advisable to compare values of P_{st} between traits, as would be possible for Q_{st}, because the difference between P_{st} and Q_{st} will vary between traits, relative to their values of h^2.

Given the design of the provenance trials, phenotypic measurements from individual trees were considered to be subject to the random effects of soil zone, parcelle nested within soil zone, and the crossed random effect of provenance. To incorporate this random effects structure, modelling was conducted using the program R version 2.11.1 (R Development Core Team 2011), with the *lmer* function from the *lme4* package (Bates et al. 2011). As traits were either continuous or ordinal with a moderate number of levels, a Gaussian error family was applied. Phenotypic measurements from individual trees were modelled separately for each trait using restricted maximum likelihood estimation, and values of V_p and V_{wp} were obtained for each trait as the variance component of the random effect of provenance, and the variance component of the random effect of parcelle nested within soil zone respectively. These were used to obtain estimates of Q_{st} following equation 2.2. Models were assessed for heteroscedasticity and normality of errors by plotting standardised residuals against fitted values, and ordered residuals against the quantiles of the normal distribution (Crawley 2007). Trees that had died by the time a trait was measured were excluded from the analysis of that trait.

The neutral genetic markers available in this study are from microsatellite loci, which are expected to follow a stepwise mutation model and to have a higher mutation rate

than other markers such as allozyme loci (Slatkin 1995). R_{st} was therefore used in preference to F_{st} as the appropriate measure of differentiation of allele frequencies. A method for comparing P_{st} with R_{st} was adapted from Whitlock & Guillaume (2009), requiring the simulation of a distribution for each of R_{st}, V_p, and V_{wp} for each trait, under the neutral hypothesis that P_{st} is equal to R_{st}.

The mean value of R_{st} across the 10 neutral microsatellite loci was estimated from variance components using the multi-locus method of Weir and Cockerham (1984) following the equation:

$$R_{st} = \frac{\sum sig\_a}{\sum(sig\_a + sig\_b + sig\_c)} \qquad Equation\ 2.3$$

where for each locus, *sig_a* is the component of variance for R_{st} among populations, *sig_b* is the component among individuals within populations, and *sig_w* is the component within individuals. The variance components were calculated using FSTAT version 2.9.3.2. To simulate random sampling of R_{st}, equation 2.3 was repeated for sets where the marker loci were randomly sampled with replacement until the number of loci in the simulated set equalled the number of loci in the real data set (i.e. n=10). Hereafter, the observed value of R_{st} calculated from the 10 loci is referred to as $R_{st(obs)}$, while simulated values are referred to as R'_{st}.

The neutral hypothesis assumes that P_{st} is equal to R_{st}, and so equation 2.2 can be rearranged to give a value of V_p under neutrality as:

$$V_{p(neutral)} = \frac{2\,R_{st}V_{wp}}{(1 - R_{st})} \qquad Equation\ 2.4$$

Estimates of $V_{p(neutral)}$ were obtained for each trait from the observed values of R_{st} and V_{wp} following equation 2.4. $V_{p(neutral)}$ is expected to vary due to stochastic heterogeneity of evolutionary history between populations, which can be approximated by the Lewontin-Krakauer distribution (Lewontin and Krakauer 1973, Whitlock 2008). Simulations of $V_{p(neutral)}$ (referred to as $V'_{p(neutral)}$) were therefore calculated by dividing the observed value of $V_{p(neutral)}$ by the degrees of freedom for provenance (the number of provenances minus 1), and multiplying by a number drawn at random from a chi-squared distribution with degrees of freedom equal to

the degrees of freedom for provenance (Whitlock and Guillaume 2009). Similarly, simulations of V_{wp} (referred to as V'_{wp}) were calculated by dividing the observed values of V_{wp} by the degrees of freedom for parcelle nested within soil zone (equal to the number of provenances), and multiplying by a number drawn at random from a chi-squared distribution with degrees of freedom equal to the degrees of freedom for parcelle nested within soil zone. Simulated values for neutral P_{st} (referred to as P'_{st}) were then generated as:

$$P'_{st} = \frac{V'_{p(neutral)}}{V'_{p(neutral)} + 2\,V'_{wp}} \qquad Equation\ 2.5$$

A distribution of the test statistic of $P_{st} - R_{st}$ assuming neutrality was generated by calculating 1000 repeats of P'_{st} - R'_{st}. The position of observed P_{st} minus $R_{st(obs)}$ within this distribution was assessed for each trait, to test for departure from neutrality.

2.2.6. Multi-model inference

In ecological studies, it is often of interest to examine the relationships between a response variable and several potential covariates to determine which, if any, are important predictors of variation in the response. Approaches to such questions have traditionally involved stepwise comparison of nested models to identify a single model that contains only predictor variables deemed to explain a significant amount of deviance in the response, as determined by null hypothesis testing (Crawley 2007). However, such approaches have been criticised on several grounds, including the dependency of the identified model on the employed selection algorithm (Calcagno and de Mazancourt 2010) and issues of multiple hypothesis testing (Whittingham et al. 2006). An alternative that is increasing in popularity is to use information criteria (IC) such as the Akaike information criteria (AIC, Akaike 1974) to compare the performance of multiple competing models.

The AIC and related ICs use deviance as a measure of the fit of a particular model to a given dataset, with a penalty applied for the number of estimated parameters. The AIC is generally used in its corrected form (AICc), to account for potentially small

samples. When multiple predictor variables are being considered, models containing all possible combinations can be ranked in order of performance by their IC score to identify the best approximating model. Various statistical software packages are available for automating the calculation of IC scores for potentially large model sets (e.g. Calcagno and de Mazancourt 2010, Barton 2011). Additional derived statistics such as model weights (the probability that a particular model is the best approximating model within a set) and evidence ratios (a measure of how much more likely one model is compared to another) can be used to assess model uncertainty (full details on the calculation of AIC and related statistics are provided by Symonds and Moussalli 2011). In situations where no single model is clearly superior to all others (i.e. the model weight of the best approximating model does not approach 1), model averaging can be employed to account for model uncertainty and obtain robust parameter and error estimates across multiple models, where the contribution of each model is weighted by its relative performance (Grueber et al. 2011, Symonds and Moussalli 2011). The model averaged parameter estimate for a particular predictor can either be based on all models, where it receives a value of zero for models that do not contain the predictor (termed 'full-model averaging'), or can be based only on those models that do feature the predictor (termed 'natural-averaging', Symonds and Moussalli 2011).

In this study, the relationships between various tree phenotypic traits (response variables) and tree provenance, or climatic conditions at provenance source sites (predictor variables) were investigated through an IC based approach. Differences in AICc scores and evidence ratios were used to compare the performance of models and make inferences about the importance of particular predictor variables. Where assessing the influence of multiple predictors, these were centred and standardized to allow for the interpretation of parameter estimates in models containing interaction terms (Gelman 2008, Schielzeth 2010). The effect size and the significance of individual predictors were assessed from estimates of their slope parameters and confidence intervals, obtained through model averaging. As there was potential for co-linearity of predictors, natural averaging rather than full-model averaging was applied, to avoid shrinkage towards zero.

The use of AICc and related ICs is subject to issues of boundary effects and uncertainty in the estimation of degrees of freedom in models that contain random effects (Bolker et al. 2009). Such estimation of degrees of freedom for random effects is not straightforward, and although there is no clear consensus approach, the default method adopted in the packages used here (i.e. using the minimum of 1 d.f for each random effect in the model) is potentially dubious (Bolker et al. 2009). However, as these issues also apply to alternative methods, such as likelihood ratio testing, the AICc approach was still considered to be the most appropriate option for the analysis presented here.

2.2.7. Non-independence of provenances

A common shortcoming in studies of intraspecies populations is that the populations are treated as statistically independent entities. This is unlikely to be true, as populations that are genetically similar through gene-flow or phylogeographic history, can be expected to co-vary in traits independently of any population-specific effects (Stone et al. 2011). Where the effects of non-independence are severe, failure to address this statistically will increase error rates in the inference of population-specific effects.

Of the various methods that have been developed for addressing this issue (reviewed by Stone et al. 2011), a genetic autocorrelation approach was considered to be most compatible with the available genotypic data and the multi-model inference methods applied in this study. Such approaches attempt to detect and remove those portions of trait variation that are explained by genealogical correlation (Edwards and Kot 1995). To this aim, the primary axis of the principle coordinate analysis of Jost's D_{est} from analysis of the microsatellite data was adopted as a linear measure of neutral genetic distance between provenances. This linear variable – henceforth referred to as *PCoA* – was included as a fixed effect in models of phenotypic trait variation, alongside environmental predictor variables whose relationship with the trait was being assessed (see section 2.2.8). In this way, the portion of trait variation that correlated with *PCoA* was removed from that which could be attributed to the environmental predictors.

2.2.8. Environmental clines

For phenotypic traits that showed evidence of local adaptation, a multi-model inference approach was applied to investigate relationships with the environmental predictor variables *SummerPrec*, *WinterPrec*, *SummerTemp*, and *WinterTemp* (see section 2.2.2 for details of these variables). Using the *lmer* function of the *lme4* package in R, a global model was defined for each trait containing fixed effects for the four environmental variables, all pair-wise interactions between them, and the *PCoA* variable (see section 2.2.7). The random effects structure again included effects for soil zone, parcelle nested within soil zone, and a crossed random effect of provenance. As traits were either continuous or ordinal with a moderate number of levels, a Gaussian error family was applied. Models were assessed for heteroscedasticity and normality of errors by plotting standardised residuals against fitted values, and ordered residuals against the quantiles of the normal distribution (Crawley 2007). Global models were standardised to allow for model averaging using the *stdz.model* function of the *arm* R package (Gelman 2008, Gelman et al. 2011, Grueber et al. 2011). With the constraint that the *PCoA* variable must be included, all possible fixed effect sub-models were tested and ranked by their AICc scores using the dredge function of the *MuMIn* package in R (Barton 2011). Model averaged estimates of standardized fixed effect slope parameters with their standard errors and confidence intervals were obtained by natural averaging using the *model.avg* function of the *MuMIn* R package. Predictors were considered to be of significance where the 95% confidence intervals of their slope parameter estimate did not include zero. Significant relationships were visualised using the *plotLMER.fnc* function of the *LanguageR* R package (Baayen 2011), as applied to un-standardized models containing all significant predictors.

The identification of significant relationships between phenotypic trait variation and the environmental variables would indicate that such environmental factors are involved in driving local adaptation. However, there remains the possibility that other environmental variables are also involved in this process and to investigate this, the analyses were repeated to include the geographical variables Latitude and Longitude, which are expected to correlate broadly with many environmental

gradients (Aitken 2004). For practicality, two-way interactions were constrained to being between the four original variables, and between Latitude and Longitude.

2.3. Results

2.3.1. Population genetic structure

All 17 provenances were polymorphic for all loci, with mean allelic richness ranging from 4.29 to 15.31. The number of alleles at the 10 loci ranged from 8 to 28. After Bonferroni adjustment for the number of populations and loci, there was no significant ($P<0.05$) evidence of deviations from Hardy-Weinberg equilibrium, or linkage disequilibrium. The estimates of divergence (D_{est}) and differentiation of allele frequencies (F_{st} and R_{st}) were generally low (<0.2, Table 2.2), although the 95% confidence intervals for minimum divergence (D_{est}) were greater than 0 for all loci, indicating significant levels of differentiation between provenances within the sample.

In the principal coordinate analysis of D_{est}, the majority of the variation was summarised along a single axes (Figure 2.3, axis1=58.34%, axis2=14.55%, axis3=8.93%). While most of provenances were clustered together towards one end this axis, provenance 249 (from Bolu, Turkey) appeared relatively distinct. This likely reflects its large geographical distance from the other provenances (Figure 2.2), and the limiting influence that this has had on gene flow.

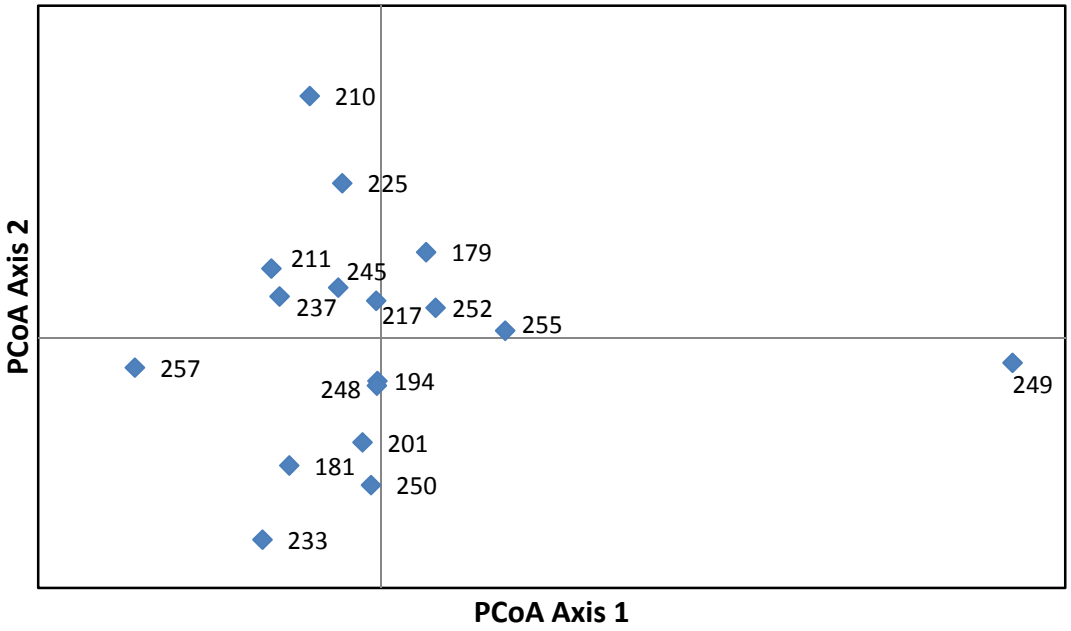


Figure 2.3. The location of provenances along the 1st and 2nd axes of the principal coordinate analysis of D_{est}.

Table 2.2. Estimates of population genetic parameters for 10 microsatellite loci. All parameters are described in the text (section 2.2.4).

Locus	**No. of alleles**	**Mean allelic richness**	H_o	H_e	F_{it}	F_{st}	F_{is}	R_{st}	**Min D_{est} 95% CI**	D_{est}	**Max D_{est} 95% CI**
C	16	4.93	0.282	0.272	0.031	0.045	-0.014	0.022	0.014	0.018	0.04
D31	15	8.75	0.766	0.733	-0.009	0.016	-0.025	-0.004	0.069	0.049	0.135
F	12	8.31	0.818	0.804	0.019	0.015	0.004	0.016	0.11	0.075	0.201
G	13	6.35	0.712	0.728	0.049	0.014	0.035	0.022	0.06	0.04	0.135
A15	8	4.29	0.607	0.605	0.045	0.028	0.017	0.018	0.053	0.049	0.109
A11	11	5.92	0.596	0.577	0.015	0.026	-0.011	0.029	0.042	0.038	0.09
AB	28	15.31	0.897	0.884	0.023	0.018	0.005	0.009	0.258	0.175	0.358
S19	23	10.8	0.830	0.805	0.004	0.015	-0.011	0.035	0.119	0.076	0.195
AK	12	8.09	0.785	0.772	0.027	0.023	0.005	-0.007	0.114	0.088	0.196
D20	11	6.36	0.667	0.659	0.028	0.026	0.002	0.037	0.062	0.059	0.14
Across loci	-	-	0.696	0.684	0.022	0.021	0.002	0.0138	-	0.047	-
95% CI	-	-	-	-	± 0.010	± 0.04	± 0.010	-	-	-	-

Table 2.3. Summary of P_{st} – R_{st} comparisons for phenotypic traits, showing the estimates of variance components between provenances (V_p) and within provenances (V_{wp}), estimates of P_{st}, the estimate of R_{st} across loci, and the probability that P_{st} is equal to R_{st}.

	Vp	Vwp	P_{st}	R_{st}	Probability $P_{st} \sim R_{st}$
Budburst	24.653	18.128	0.405	0.0138	0 ***
Retention	22.587	3.176	0.781		0 ***
Height 96	1.116	7.851	0.066		0 ***
Height 2001	0.943	16.172	0.028		0.066
DBH	1.035	11.946	0.042		0.002 **
Form	4.151	3.314	0.385		0 ***
No. Branches	2.000	10.325	0.088		0 ***
No. Forks	0.283	9.349	0.015		0.471

Confidence level codes: * = 95%, ** = 99%, *** = 99.9%

2.3.2. Phenotypic differentiation between provenances

Estimates of P_{st} ranged from 0.015 to 0.781, and were larger than R_{st} with a very high level of confidence for *Budburst, Retention, Ht96, Form*, *NoBranches* ($p < 0.001$), and to a lesser extent for *DBH* ($p < 0.01$, Table 2.3). Genetic drift is therefore insufficient to explain the extent of variation for these traits, with the implication that they have experienced spatially divergent selection resulting in local adaptation to particular environments. Estimates of P_{st} were not significantly greater than R_{st} for *NoForks* or *Ht2001* ($p > 0.05$), and there is therefore no clear evidence of these traits having been influenced by local adaptation. However, this should not be considered conclusive, as P_{st} represents a minimum estimate of Q_{st}, and the true value could be greater. In the case of *Ht2001*, the test statistic (P'_{st} minus R'_{st}) was well towards the higher end of the neutral distribution, and it is likely that true Q_{st} would be significantly greater than R_{st}.

2.3.3. Environmental clines

For the six phenotypic traits where P_{st} was significantly higher than R_{st}, modelling with combinations of the environmental predictor variables (i.e. *SummerPrec, WinterPrec, SummerTemp, WinterTemp*, and their pair-wise interactions) revealed best approximating models that performed better than those without any of the fixed effects in all cases except for *Ht96* (see Table 2.4). The degree of improvement in model performance was very slight for *DBH* (i.e. the evidence ratio of 1.11 indicated that the best approximating model was 1.11 times more likely to be better than the model with no environmental predictors), moderate for *NoBranches* and *Budburst* (i.e. with evidences ratios of 18.3 and 796 respectively), and large for *Retention* and *Form* (i.e. with evidence ratios of 1.02×10^4 and 8.59×10^4 respectively, Table 2.4).

Significant relationships between trait variation and the environmental predictors were apparent for *Budburst, Retention, Form* and *NoBranches*, but not for *DBH* and *Ht96*, as inferred where the 95% confidence intervals for the slope parameter of a predictor did not include zero (Table 2.4). Significant main effects were of most interest as these indicated relationships with a substantial and relatively consistent

slope. The co-occurrence of a particular variable as a significant main effect and as part of an interaction indicated that the slope of the main effect varied depending on the level of a second predictor, but that its direction was generally consistent. Predictors that featured in significant interactions but not as a significant main effect indicated that the slope of its relationship varied substantially depending on the value of a second predictor, but that its direction was inconsistent.

For *Budburst*, the significant positive main effects of *SummerTemp* and *WinterPrec* indicated that higher provenance summer temperature and winter precipitation corresponded with early bud-burst phenology at the trial site (Figure 2.4 a_i and a_{ii}). The significant negative main effect of *WinterTemp* indicated that higher provenance winter temperature corresponded with later bud-burst phenology (Figure 2.4 a_{iii}). The significant interaction between *WinterPrec* and *SummerPrec* where only *WinterPrec* also featured as a main effect indicated that the slope of the relationship between *Budburst* and *WinterPrec* varied in relation to the value of *SummerPrec,* but was generally positive (Figure 2.4 a_{iv}). The significant interaction between *SummerTemp* and *WinterTemp* where both also featured as main effects indicated that their slopes varied in relation to the values of each other, but were consistently positive for *SummerTemp*, and negative for *WinterTemp* (Figure 2.4 a_v and a_{vi}).

For *Retention* the significant negative main effects of *SummerTemp* and *WinterPrec* indicated that higher provenance summer temperature and winter precipitation corresponded with early leaf-fall phenology at the trial site (Figure 2.4 b_i and b_{ii}). The significant positive main effect of *WinterTemp* indicated that higher provenance winter temperature corresponded with later leaf-fall phenology (Figure 2.4 b_{iii}). The significant interaction between *SummerTemp* and *WinterTemp* where both also featured as main effects indicated that their slopes varied in relation to the values of each other, but were consistently negative for *SummerTemp*, and generally positive for *WinterTemp* (Figure 2.4 b_v and b_{vi}).

For *Form,* the significant negative main effect of *SummerTemp* indicated that higher provenance summer temperature corresponded with poorer tree form at the trial site (Figure 2.4 c_i). The significant positive main effect of *WinterTemp* indicated that higher provenance winter temperature corresponded with better tree form (Figure 2.4

c_{ii}). The significant interaction between *SummerTemp* and *WinterPrec* where only *SummerTemp* also featured as a main effect indicated that the slope of the relationship between *Form* and *SummerTemp* varied in relation to the value of *SummerPrec,* but was generally negative (Figure 2.4 c_{iii}). The significant interactions between *WinterTemp* and *SummerPrec* and between *WinterTemp* and *WinterPrec* where only *WinterTemp* also featured as a main effect indicated that the slope of the relationship between *Form* and *WinterTemp* varied in relation to the values of both *SummerPrec* and *WinterTemp,* but was generally positive (Figure 2.4 c_{iv} and c_{v}). The significant interaction between *WinterPrec* and *SummerPrec* where neither also featured as significant main effects indicated that the slopes of their relationships with *Form* varied substantially depending on the value of each other, but were inconsistent in their direction (Figure 2.4 c_{vi}).

For *NoBranches*, the significant negative main effects of *SummerTemp* and *WinterPrec* indicated that higher provenance summer temperature and winter precipitation corresponded with fewer branches at the trial site (Figure 2.4 d_{i} and d_{ii}). The significant positive main effects of *SummerPrec* and *WinterTemp* indicate that higher provenance summer precipitation and winter temperature corresponded with more branches (Figure 2.4 d_{iii} and d_{iv}).

Table 2.4. Summary results for modelling of tree phenotypic traits in relation to environmental predictor variables, showing the difference in AICc scores ($Diff_{AICc}$) and the evidence ratios (ER) between the best approximating models and those without any fixed effects for environmental predictors. Estimates of slope parameters, their standard errors (SE), and confidence intervals, are presented for significant relationships (i.e. where 95% confidence intervals did not include zero). Interaction terms are indicated by two predictor names separated by a colon (:). 'NS' indicates that no significant relationships were identified.

Trait	$Diff_{AICc}$	ER	Predictor	Estimate	SE	Lower 95% CI	Upper 95% CI
Budburst	-13.36	796	*SummerTemp*	1.76	0.66	0.47	3.05
			WinterPrec	1.25	0.39	0.48	2.02
			WinterTemp	-1.99	0.75	-3.45	-0.52
			WinterPrec : SummerPrec	-2.55	0.93	-4.38	-0.73
			SummerTemp : WinterTemp	-1.49	0.57	-2.60	-0.37
Retention	-18.46	1.02×10^4	*SummerTemp*	-1.43	0.28	-1.97	-0.88
			WinterPrec	-1.19	0.23	-1.63	-0.74
			WinterTemp	1.61	0.31	1.00	2.22
			SummerTemp : WinterTemp	1.57	0.36	0.86	2.27
DBH	-0.21	1.11	*NS*				
Ht96	0.00	1	*NS*				
Form	-22.72	8.59×10^4	*SummerTemp*	-0.53	0.18	-0.88	-0.18
			WinterTemp	0.45	0.21	0.05	0.86
			SummerPrec : WinterPrec	1.07	0.27	0.54	1.59
			SummerPrec : WinterTemp	0.63	0.26	0.13	1.14
			SummerTemp : WinterPrec	-1.11	0.46	-2.02	-0.21
			WinterPrec : WinterTemp	1.28	0.37	0.55	2.00
NoBranches	-5.81	18.3	*SummerPrec*	1.51	0.66	0.22	2.80
			SummerTemp	-3.16	1.22	-5.56	-0.76
			WinterPrec	-1.86	0.79	-3.42	-0.31
			WinterTemp	3.57	1.28	1.05	6.09

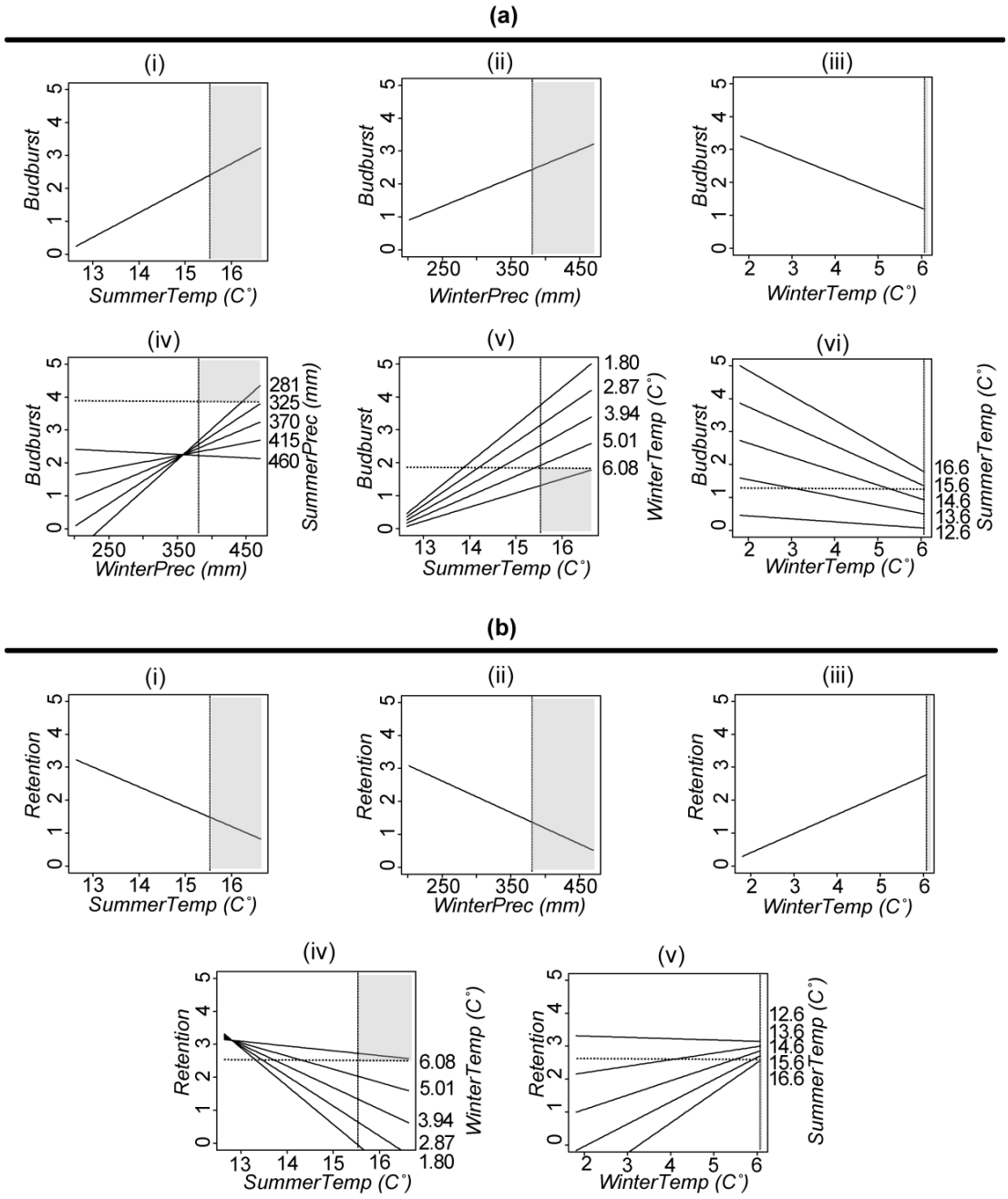


Figure 2.4. Graphical illustration of significant relationships between environmental predictor variables and the phenotypic traits (a) *Budburst*, (b) *Retention*, (c) *Form*, and (d) *NoBranches* (see Table 2.4 for parameter estimates). For interaction terms, the relationship between the response variable (y axis) and the first predictor variable (x axis) is plotted for 5 values of the second predictor variable (values shown to the right of plots), corresponding to the maximum, minimum, and 25 percentiles from the observed range of the second predictor. Values of the environmental variables for the trial site are indicated for the first and second predictors by horizontal and vertical dashed lines respectively. Grey shaded areas indicate the expected direction of change in predictor variables under general climate change predictions (e.g. increased summer and winter temperatures, increased winter precipitation, and decreased summer precipitation).

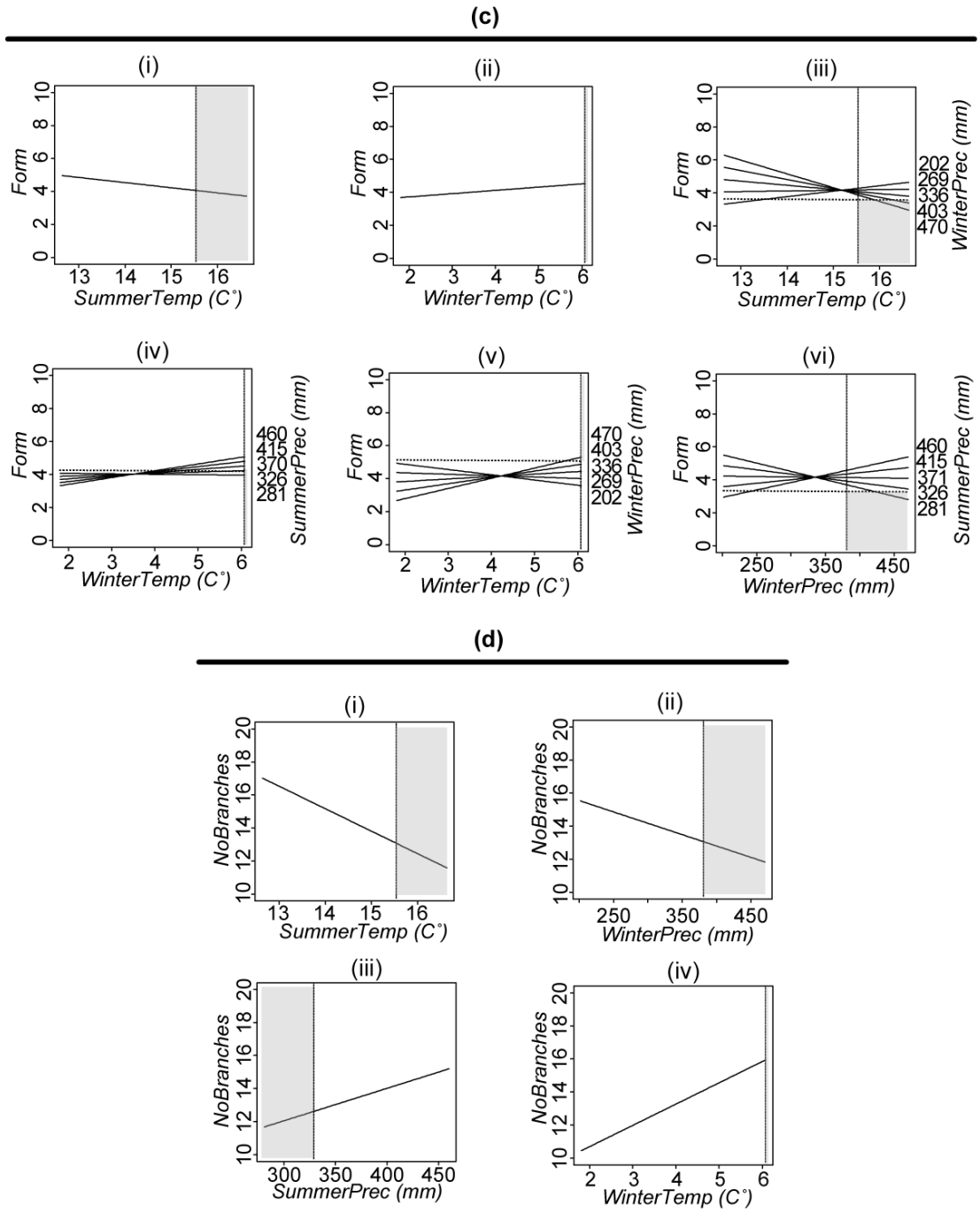


Figure 2.4. Continued from previous page.

Although relationships with temperature and precipitation were apparent for several phenotypic traits when only these variables were considered, inclusion of the geographic variables latitude and longitude substantially altered the inference. Slightly better approximating models were identified for *Retention, DBH, Ht96, Form,* and *NoBranches,* with evidence ratios indicating that the best approximating models that could include latitude and longitude were respectively 4.1, 4.63, 5.84, 1.47, and 4.57 times more likely to be better than the best approximating models with just the environmental variables. For *Budburst*, the difference was much greater, with an evidence ratio of 1.32×10^4 in favour of the best approximating model that could include latitude and longitude (Table 2.5). Latitude appeared to be the most influential of the geographic variables, with significant main effects of latitude identified for *Budburst*, *Retention*, and *Ht96*. When the geographical variables were included for modelling of *Budburst,* the significant main effects of *WinterPrec* and *WinterTemp* remained, but the significant positive main effect of *SummerTemp* was no longer apparent, and an additional significant positive main effect of *SummerPrec* was identified. For *Retention, DBH, ht96,* and *Form*, no significant main effects for temperature or precipitation remained, although various significant interaction terms were apparent. For *NoBranches*, no significant main effects or interactions remained (Table 2.5).

Table 2.5. Summary results for modelling of tree phenotypic traits in relation to environmental and geographic predictor variables, showing the difference in AICc scores ($Diff_{AICc}$) and the evidence ratio (ER_1) between the best approximating models and those without any fixed effects for environmental or geographic predictors, and the evidence ratios (ER_2) between the best approximating models that could contain latitude and/or longitude and those that could contain only temperature and precipitation variables. Estimates of slope parameters, their standard errors (SE), and confidence intervals, are presented for significant relationships (i.e. where 95% confidence intervals did not include zero). 'NS' indicates that no significant relationships were identified. The table is continued on the following page.

Trait	$Diff_{AICc}$	ER_1	ER_2	Predictor	Estimate	SE	Lower 95% CI	Upper 95% CI
Budburst	-32.33	1.05×10^7	1.32×10^4	*Latitude*	-1.80	0.57	-2.91	-0.68
				SummerPrec	1.40	0.39	0.63	2.17
				WinterPrec	-2.22	0.83	-3.84	-0.60
				WinterTemp	-3.33	1.59	-6.44	-0.21
				SummerPrec : WinterPrec	-3.37	0.66	-4.66	-2.08
				SummerPrec : WinterTemp	-3.58	0.76	-5.08	-2.09
				SummerTemp : WinterPrec	10.94	1.91	7.20	14.69
				WinterPrec : WinterTemp	-2.39	1.05	-4.45	-0.34
				Latitude : Longitude	-6.92	2.71	-12.23	-1.61
Retention	-21.28	4.18×10^4	4.1	*Latitude*	1.10	0.34	0.44	1.77
				SummerPrec : SummerTemp	1.53	0.56	0.42	2.63
				SummerPrec : WinterPrec	1.74	0.50	0.76	2.71
				SummerTemp : WinterPrec	-3.41	1.12	-5.61	-1.21
				SummerTemp : WinterTemp	1.04	0.48	0.10	1.98
				WinterPrec : WinterTemp	2.07	0.64	0.82	3.31

Table 2.5. Continued.

Trait	$Diff_{AICc}$	ER_1	ER_2	Predictor	Estimate	SE	Lower 95% CI	Upper 95% CI
DBH	-3.27	5.14	4.63	*SummerPrec : SummerTemp*	35.37	16.64	2.76	67.98
Ht96	-3.53	5.84	5.84	*Latitude*	-24.14	11.91	-47.48	-0.80
				SummerPrec : WinterTemp	22.44	10.36	2.13	42.76
				SummerTemp : WinterTemp	33.32	15.66	2.62	64.01
Form	-23.49	1.26×10^5	1.47	*SummerPrec : WinterPrec*	1.11	0.27	0.58	1.64
				SummerPrec : WinterTemp	0.81	0.32	0.19	1.43
				SummerTemp : WinterPrec	-1.38	0.56	-2.48	-0.27
				WinterPrec : WinterTemp	1.50	0.40	0.72	2.28
NoBranches	-8.85	83.5	4.57	*NS*				

2.4. Discussion

2.4.1. Evidence for local adaptation

The phenological traits of *Budburst* and *Retention* respectively describe the timing of growth initiation in spring and growth cessation in autumn. Local adaptation of such phenological traits has been observed for many temperate tree species and is considered to relate mainly to geographical variation in temperature, with selection for phenologies that minimise frost injury to active tissues while maximising the time available for growth (Aitken et al. 2008). The minimum estimates of Q_{st} reported here (0.405 and 0.781) are higher than the Q_{st} of analogous traits reported elsewhere for *Q. petraea* (e.g. leaf flushing Q_{st}=0.15 & 0.27, leaf yellowing Q_{st}=0.15 & 0.18, Jensen and Hansen 2008), but similar to those for several other species (e.g. needle flush in *Pinus albicaulis* - Q_{st}=0.47, Bower and Aitken (2008), and bud set in *Picea sitchensis* - Q_{st}=0.89, Mimura and Aitken (2007), and in *Pinus sylvestris* - Q_{st}=0.86, Savolainen et al. 2004). This likely reflects the greater environmental range of provenances considered here, relative to the study of Jensen and Hansen (2008) that featured only provenances from North-west Europe. The Q_{st}-F_{st} tests presented here provided strong evidence for local adaptation of both phenological traits. Significant clines along gradients in summer temperature, winter temperature, and winter precipitation were identified for both traits, suggesting that these environmental factors are involved in the local adaptation of the traits. However, inclusion of the geographical variables latitude and longitude substantially improved the modelling of variation in bud-burst phenology, and to a lesser extent for leaf-fall phenology, and altered which temperature and precipitation clines were considered to be of significance. When the geographical variables were included for modelling of bud-burst phenology, the cline with summer temperature was no longer significant, but additional significant clines with latitude and summer precipitation were apparent. For modelling of leaf-fall phenology, clines with summer temperature, winter temperature, and winter precipitation were no longer significant, but a significant cline with latitude was apparent.

Architectural traits are of considerable importance for silviculture as they directly influence the commercial value of timber crops, with straight un-forked stems usually being most desirable and considered to have the best form. While natural selection is not necessarily expected to directly improve tree form (Worrell 1992), forking, excessive branching, and otherwise poor form, may all result from environmentally induced damage to apical shoots, with selection expected for phenotypes that minimise such damage. Geographic variation in the environmental causes of damage (e.g. frost or wind) may therefore result in patterns of local adaptation that are manifested in tree architecture. Morphological traits have received less attention in the literature than phenological and growth traits, but reported values of Q_{st} are generally low (e.g. <0.3, Savolainen et al 2007). The minimum estimates of Q_{st} reported here are within this range for the number of forks, and the number of branches (0.015 and 0.088 respectively), but the value for tree form (0.385) is relatively high. Q_{st}-F_{st} tests offered no evidence of local adaptation in the number of forks, but strong evidence for both number of branches and tree form. Significant clines along gradients in summer temperature and winter temperature were identified for both the number of branches and tree form, with further clines with winter and summer precipitation also identified for the number of branches, suggesting that these environmental factors are involved in the local adaptation of the traits. However, inclusion of the geographical variables latitude and longitude slightly improved the modelling of variation in both the number of branches and tree form, and altered the inference with the result that no clear clines along temperature or precipitation gradients remained.

Tree size is expected to reflect the combined influence of many physiological factors such as photosynthetic rate, water use efficiency, growth rate and duration, and root-shoot biomass allocation, with selective pressure for large trees that compete well for light and other resources. Variation in tree size between provenances could be the result of local adaptation of any one or more of these factors, and as such, tree size is often used as an index of tree health and overall degree of adaptation (Aitken 2004, Savolainen et al. 2007). The minimum estimates of Q_{st} reported here for diameter, and height at 5 and 11 years of age (0.042, 0.066, and 0.028) are similar to Q_{st} values reported elsewhere for *Q. petraea* (e.g. height year 1 - Q_{st}=0.06, diameter year

1 - Q_{st}=0.01, Jensen and Hansen 2008), but are low compared to those for several other species (Savolainen et al. 2007). Q_{st}-F_{st} tests provided evidence of local adaptation for diameter at breast height and height at five years of age, but the minimum estimate of Q_{st} for height at 11 years of age was not quite large enough to be considered indicative of local adaptation. No clines along temperature and precipitation gradients were apparent for either diameter at breast height or height at 3 years of age, suggesting that these environmental influences are not strongly involved in the local adaptation of these traits. Inclusion of the geographical variables latitude and longitude slightly improved the modelling of variation in both traits, and a significant cline along a gradient in latitude was apparent for height in 1996.

In summary, several phenotypic traits relating to phenology, architecture, and size in *Q. petraea* showed signs of having become locally adapted to their environments. Geographical variation in temperature and precipitation was implicated in the local adaptation of phenological and architectural traits, but not size traits. However, the substantial effect of incorporating additional geographic variables suggested that geographical variation in temperature and precipitation alone may not offer the most appropriate explanation for patterns of local adaptation.

2.4.3. Predicting the performance of matched provenances

Due to the predicted direction of climate change in southern Britain and much of Western Europe (Giorgi and Coppola 2009, Jenkins et al. 2009), Climate Matching is expected to match planting sites in this region with tree provenances that have experienced higher recent winter precipitation, lower summer precipitation, and higher summer and winter temperatures. By assessing the identified clines with temperature and precipitation gradients for various locally adapted phenotypic traits, it is possible to assess the relative performance at Petite Charnie of trees from provenances with such climates. Although interactions between variables complicate their interpretation, trees from provenances with higher winter precipitation and lower summer precipitation than Petite Charnie appeared to exhibit early bud-burst phenology, early leaf-fall phenology, poor form, and less branching relative to tree of

local provenance. Similarly, trees from provenances with higher summer temperature than Petite Charnie also exhibit early bud-burst phenology, early leaf-fall phenology, poor form, and less branching relative to tree of local provenance. Interpretation of the influence of increased winter temperature is complicated as conditions at Petite Charnie are towards the upper limit of the observed range. However, the patterns appear to contradict those for precipitation and summer temperature, with trees from provenances with higher winter temperature than Petite Charnie exhibiting late bud-burst phenology, late leaf-fall phenology, improved form, and greater branching relative to trees of local provenance (Figure 2.4). Differences in tree size could not be predicted, due to an absence of clinal relationships between tree size and the temperature and precipitation gradients.

Tree form is a particularly relevant trait in the context of Climate Matching as it can be considered to reflect an overall degree of adaptation. Trees possessing traits that entail good growth rates and the avoidance of damage in their environment would be considered as well adapted, consequently growing tall and straight and thus having good form (details of how form is measured are provided in Appendix 2.1.3). Trees possessing traits that entail relatively poor growth rates or susceptibility to damage (e.g. from frost, wind, pests and pathogens) in their environment would be considered as poorly adapted, consequently growing slowly and without apical dominance and thus having poor form. As Climate Matching is intended to promote adaptation, it is concerning those trees from provenances with warmer summer temperature, higher winter precipitation, and lower summer precipitation, exhibited poorer form at the Petite Charnie trial relative to trees of local provenance. However, this may reflect an important characteristic of climate matching: that matched provenances are expected to be well adapted to the future climate of a planting site, but not necessarily to the present climate. Thus, the observed relatively poor adaptation (as characterised by low form scores) of supposedly matched provenances may be due to a temporal lag in adaptation, which will decrease as time passes and the climate changes. The extent of such a lag can be expected to depend on the difference in present climate between the planting site and the matched site, which is in turn determined by the severity of carbon emissions scenario and the timing of the projections (i.e. 2050's or 2080's) used in the climate matching. Finding a suitable

balance between present and future adaptation should be an important part of any climate matching decision process, and might be dependent on the planting objectives. In a commercial plantation for example, where timber quality is a primary concern, a large lag in adaptation could increase the risk of damage to young trees, resulting in poor form and low timber value. Climate matching with little or no lag in adaptation may therefore be most prudent (e.g. based on a low carbon emissions scenario or a short projection time). Alternatively, if the primary concern of a plantation is survival and longevity, perhaps for an amenity or conservation forest, then a larger lag in adaptation may be more appropriate, reducing the risk of serious effects of maladaptation or even mortality in mature trees.

2.4.2. Climate Matching of *Q. petraea*

A principle objective of this study was to explore whether provenances of *Q. petraea* are locally adapted to the temperature and precipitation variables that are used in Climate Matching analysis. The results on this issue are somewhat ambiguous. Several important phenotypic traits showed signs of being locally adapted to provenance environments, and for some of these traits, clines along gradients in summer and winter temperatures and precipitation suggested that these variables are involved in the adaptation. However, important traits - such as tree diameter and height - showed evidence of local adaptation but did not appear to be strongly influenced by gradients in these environmental variables. Furthermore, the consideration of additional environmental gradients (i.e. longitude and latitude), generally improved the modelling of variation in the phenotypic traits. When considered together, these results suggest that while gradients in temperature and precipitation are likely to be part of the selection regime that drives local adaptation of *Q. petraea*, they are not necessarily the only or even the strongest influences. Therefore, if Climate Matching is based solely on these variables, then there is a risk that matched provenances will be maladapted to alternate aspects of the environment at their planting sites.

The geographical variables latitude and longitude are often used in studies of genecology as surrogates for overall environmental variation (Aitken 2004). They are

expected to correlate with various gradients in temperature and moisture but also with other environmental gradients. Photoperiod for example is determined by latitude and time of year (Forsythe et al. 1995), and while a correlation with temperature is expected, photoperiod will not be altered by any changes in climate. Photoperiodic cues have been implicated in the local adaptation of bud set phenology in trees (Howe et al. 2003), and the matching of sites from different latitudes could therefore result in climate matched provenances being maladapted to the photoperiod of their planting sites, with consequent sub-optimal growth or increased risk of cold injury. The risk of maladaptation to overlooked abiotic factors such as photoperiod could potentially be minimised by including geographic variables such as latitude, longitude and altitude in the calculation of the climatic difference index used in Climate Matching. The effect of each of these could be weighted so that while sites would still be matched primarily in terms of temperature and precipitation, sites from similar latitude, longitude, and altitude to the planting site would be preferred over more distant sites. For example, under a high emissions scenario for climate change by the 2080s, Brechfa in Wales was matched with coastal areas of Brittany in France, and also with coastal areas of northern Spain (Broadmeadow et al. 2005). If latitude and longitude were included in the calculation, then the French sites would clearly be preferred as they have very similar longitude to Brechfa, and less than half the difference in latitude relative to the Spanish sites.

Whilst modification of Climate Matching analysis to incorporate more general environmental gradients might minimise the risk of maladaptation to abiotic factors, it does not address the risk of maladaptation to biotic factors such as soil conditions and pathogens. These are potentially important influences, and are widely implicated in the local adaptation of plants (Sork et al. 1993, Roy 1998, Wright 2007, Pregitzer et al. 2010). The risk of maladaptation to pathogens is spectacularly exemplified at the Petite Charnie provenance trials by a provenance of *Q. robur* from the Southwest of France that grows well in its native environment but is highly susceptible to oak mildew (*Microsphaera alphitoides*) at Petite Charnie, sustaining heavy damage with consequent poor form and high mortality (A. Duccouso, personal communication). Although this is perhaps an extreme case as no other provenance at Petite Charnie shows such a degree of susceptibility, there are various literature examples of non-

local tree and plant provenances being more susceptible to pests and pathogens (Sork et al. 1993, Roy 1998, Kaltz et al. 1999).

In conclusion, while this study suggests that geographical gradients in temperature and precipitation are involved in the local adaptation of *Q. petraea* and therefore provides some justification for Climate Matching, it also highlights the likely importance of other abiotic and biotic influences that are potentially overlooked by Climate Matching analysis. Further investigation of the patterns of local adaptation in focal tree species, and the relative importance of various influences would be valuable in guiding if, where, and how Climate Matching can be most successfully applied. Such information could come from further studies such as this, that analyse empirical data from established provenance trials.

Chapter 3 – Effects of tree provenance and phenotype on a community of herbivorous insects: implications for Climate Matching

3.1. Introduction

Considerable changes are expected in the climate of the UK within the current century, particularly in southern regions where temperatures and winter precipitation are predicted to increase, and summer precipitation is predicted to decrease (Jenkins et al. 2009). It is consequently expected that trees locally adapted to the current climate will become increasingly maladapted as their environment changes, with resulting declines in forest health and productivity (Broadmeadow et al. 2005). The adaptive forest management strategy of Climate Matching aims to preserve the productivity and ecosystem services of forests by using climate models to predict the future climate of a planting site, and then identifying provenances (geographic origins) where tree populations are locally adapted to such climates (Broadmeadow et al. 2005, Bolte et al. 2009). An initial study of Climate Matching considered four UK planting sites based on predicted changes in temperature, precipitation, and diurnal temperature range, under low and high carbon emission scenarios. These planting sites were generally matched with provenances from lower latitudes, with for example Kelty in eastern Scotland being matched with areas of Ireland and western Britain by the 2050's, and with southern Brittany by 2080's, and Alice Holt in the south of England being matched with Brittany by the 2050's, and areas of Italy, Sardinia, and Greece by the 2080's (Broadmeadow et al. 2005). Based on these predictions, trial plantations of several commercially important broadleaved tree species have recently been established in the UK (Forest Research et al. 2010).

Trees play a key role in forest ecosystems, serving as 'foundation species' that structure the biotic environment for diverse ecological communities of plants, microbes, and animals (Whitham et al. 2006). In recent years, research in the field of Community Genetics has widely demonstrated that variation within foundation tree species can influence the structure of associated communities (Dungey et al. 2000,

Underwood and Rausher 2000, Wimp et al. 2005, Bangert et al. 2006), with effects that potentially span several trophic levels (Dickson and Whitham 1996, Bailey and Whitham 2003, Johnson 2008, Jones et al. 2011b). Although a primary objective of Climate Matching and of adaptive forest management in general will be to optimise the commercial productivity of planted forests, the conservation and enhancement of forest biodiversity is also viewed as an essential element of sustainable forest management, and is reflected in UK and European policy (Forestry Commission 2004, EC 2005). When implementing Climate Matching, it will therefore be important to consider not only the potential for improvements in productivity, but also the effect that introduced tree provenances may have on associated ecological communities, a question that has yet to be specifically addressed for any European tree species. In this study, I make use of a large established provenance trial in France to investigate this issue for a commercially important European tree species and an associated guild of specialist insect herbivores – *Quercus petraea* and oak gallwasps (Hymenoptera: Cynipidae: Cynipini).

3.1.1. Study system

Sessile oak (*Quercus petraea*) is an abundant and economically important forest tree, with a current natural distribution from Spain to Russia and from Scotland to Turkey, between sea level and 1600m elevation (Kleinschmit 1993). Together with the closely related Pedunculate oak (*Q. robur*), it represents almost 25% of broadleaved high forest in Great Britain (Forestry Commission, 2003). Under recent climate change scenarios it was predicted that oak productivity will decrease in the south and east of the UK, with increased risk of drought mortality (Broadmeadow et al. 2005), and *Q. petraea* is therefore considered to be a candidate species for Climate Matching in these regions. In Chapter 2, analysis of provenance trial data for *Q. petraea* suggested the geographic variation in temperature and precipitation was involved in the local adaptation of particular traits, thus providing some justification for a Climate Matching approach.

In the UK, the two native species of *Quercus* are host to at least 423 species of phytophagous invertebrates, considerably more than most tree genera and second

only to willows (Kennedy and Southwood 1984). In continental Europe the associated diversity is even greater, with over 630 species of herbivorous insect recorded on *Quercus* in Hungary (Csóka 1998). These herbivorous communities include numerous species of oak gallwasp, whose larvae induce complex galls within which they feed upon nutritive tissue inside specialised chambers. Approximately 200 species are recognised from the Palaearctic region (Csóka et al. 2005), almost all of which have a complex cyclically parthenogenetic lifecycle with alternating sexual and asexual generations within each year (Stone et al. 2002, Stone et al. 2008). In Western Europe, females of the asexual generation typically emerge early in the year to oviposit eggs that induce galls of the sexual generation, developing once buds have burst in the early spring. Adults emerge from these galls between May and July, and mate before laying eggs of the asexual generation whose galls develop during summer and autumn (Askew 1962). Galls occur on various plant organs including buds, leaves, catkins, stems and roots, and may be single or multi-chambered (Stone et al. 2002, Csóka et al. 2005). Gall morphology and its location on the tree is generally diagnostic of a particular generation of a single species, and keys to Western Palaearctic species based on gall morphology are available (Buhr 1965, Ambrus 1974, Redfern and Shirley 2002). As the galls of the two generations are temporally and morphologically distinct, and often differ in their relative abundance by several orders of magnitude, they are usually treated as separate 'gall-types' in ecological studies of gallwasp communities (Schönrogge and Crawley 2000, Kaartinen and Roslin 2011).

Western Palaearctic oak galls are a popular model ecological system, and have recently been the focus of studies of biological invasions (Schönrogge et al. 1995, Schönrogge and Crawley 2000), comparative phylogeography (Hayward and Stone 2006), local adaptation (Tack and Roslin 2010), habitat fragmentation (Kaartinen and Roslin 2011), and community genetics (Tack et al. 2010). The distinctive morphologies and sessile nature of galls makes them easy to identify and establish densities during field studies, and they are usually present in appropriate abundance and diversity for comparative study at various spatial scales, i.e. between sites (Schönrogge et al. 1995, Schönrogge et al. 1998, Schönrogge and Crawley 2000), or between individual trees within sites (Kaartinen and Roslin 2011, Egan and Ott

2007). They are also relatively easy to rear, potentially allowing for the study of multi-trophic interactions involving inquiline gallwasps (Hymenoptera: Cynipidae: Synergini) and parasitoids (mainly Hymenoptera: Chalcidoidea) that also inhabit the galls (see Chapter 4). These communities are relatively closed, in that individual species of oak gallwasp are specialised parasites of a limited range of oaks, and individual species of inquiline and parasitoid are generally specialised inhabitants of a limited range of oak galls (Askew 1961a, Askew 1980, Stone et al. 2002, Csóka et al. 2005).

3.1.2. How might host tree variation influence gallwasp communities?

Understanding how herbivorous communities are influenced by 'bottom-up' effects from their host-plants is a long standing objective of ecological study (Hunter and Price 1992, Hunter et al. 1997), and observed correlations between plant phenotypic traits and herbivore abundance and diversity have led to the development of various hypotheses. Based on the observation that out-breaks of herbivorous pests on *Eucalyptus* in Australia coincided with weather conditions that were unfavourable for the trees, White (1969) proposed the plant stress hypothesis whereby herbivores are expected to prefer, or to perform better, on plants that are stressed by physical damage or unfavourable environmental conditions. This is primarily considered to be because such stress results in elevated concentrations of mobilised nitrogen within the plants vegetative tissues, presenting a more favourable food resource for herbivores (White 1969, 1974, 1984), but has been extended to also reflect that stressed plants may allocate fewer resources to the synthesis of defensive compounds (Rhoades 1985). Field-studies have provided support for the hypothesis in a variety of organisms including a dipteran gall former (Debruyn 1995). However, acceptance of its wider relevance has been mixed (Larsson 1989, Koricheva et al. 1998), leading White (2009) to clarify that it may apply specifically to senescence-feeders, i.e. those that feed primarily on mature plant organs that are in the process of senescence, as opposed to flush-feeders that feed on young organs that are still developing. In relation to the present study, under the plant stress hypothesis it would be expected that variation in gallwasp abundance and species richness would correlate with variation in stress, as could be determined by assessment of tree health or growth. It

may also be expected that this pattern would be stronger for asexual generation gallwasps that develop on mature oak tissues, rather than the sexual generation that develop on young tissue.

In contrast to the stress hypothesis, Price (1991) proposed the plant-vigour hypothesis, based on observations that particular herbivores responded positively towards plants or plant organs that exhibited high growth rates or achieved large ultimate size, relative to the population mean (i.e. those that were most vigorous). Various mechanisms for such patterns have been suggested, including the greater availability of oviposition sites, greater resource quality (i.e. nutrient content), and lower concentrations of defensive compounds exhibited by vigorously growing plants (Price 1991, Cornelissen et al. 2008). While the plant-stress hypothesis is expected to apply particularly to herbivores that feed on plants during their senescence phase (White 2009), the plant-vigour hypothesis is most applicable to those that are intimately involved in the processes of plant growth (Price 1991). Many gall-formers fall into this latter category as they must redirect normal plant development and stimulate plant tissues to form their galls (Price 1991, Harper et al. 2004), and field-studies have provided support for the hypothesis in a range of galling herbivores (Prado and Vieira 1999, Kopelke et al. 2003, and see Cornelissen et al. (2008) for meta-analysis across herbivore guilds) including two species of oak gallwasps (Ito and Hijii 2001). In relation to the present study, under the plant-vigour hypothesis it would be expected that variation in gallwasp abundance and species richness between oak provenances would correlate with variation in vigour, as could be determined by assessment of tree size. This may apply particularly to sexual generation gall-types that develop on young plant tissues following growth initiation in spring.

Further to the potential influences of stress and vigour, synchrony in the timing of development between herbivores and their hosts may be important aspect of plant-herbivore interactions (Yukawa 2000, van Asch and Visser 2007, Singer and Parmesan 2010). Experimental study of the oak winter moth *Operophtera brumata* indicates that herbivore fitness is greatest when the timing of egg-hatch coincides with the bursting of buds in its host plant, declining sharply if egg-hatch occurs

before or after bud-burst (van Asch et al. 2007). Given that the timing of egg-hatch in *O. brumata* is genetically determined, it is expected that populations will evolve so that the timing of egg-hatch will match the mean timing of bud-burst in their host plant population (van Asch et al. 2007). Where this pattern applies, it can be hypothesised that herbivore abundance will be greatest on trees whose phenology most closely matches the requirements of the herbivore population, with relatively lower abundance on trees that deviate in either direction (henceforth referred to as the synchronisation hypothesis). Results from studies that have considered the role of host plant phenology in oak-gallwasp interactions are mixed, with demonstrated effects in some cases (Askew 1962, Crawley and Akhteruzzaman 1988) but not in others (Ito and Hijii 2001). In relation to the present study, under the synchronisation hypothesis it is expected that the abundance of individual gallwasp species will correlate with host plant phenology, declining to either side of an optimum. If the optimum is shared between species, then a similar pattern would be expected for gall-type richness.

3.1.3. Objectives

The purpose of this study is to begin to explore how Climate Matching of forest tree species may influence associated ecological communities, by assessing the effects of host tree provenance and phenotype on a community of gallwasps. Specifically I ask: (i) does host-tree provenance influence gall-type abundances; (ii) which aspects of tree phenotype are implicated and is there support for the plant-stress, plant-vigour, and synchronisation hypotheses; (iii) how is the richness and structure of the gall community influenced; (iv) how consistent are these patterns? Based on the answers to these questions, the expected effects of Climate Matching are then discussed.

3.2. Materials and methods

3.2.1. Petite Charnie provenance trial

The oak provenance trial in the forest of La Petite Charnie in Sarthe, Northwest France, was established in the early 1990's by the French National Institute for

Agricultural Research (INRA) to aid genetic conservation and management of European white oaks (Ducousso et al. 1996, and see Chapters 1 and 2 for further details of the trial design). A set of 20 provenances of a single age cohort was selected to encompass the largest possible range of geographic distances from the trial site (Figure 3.1 and Table 3.1). This set included the 17 provenances previously assessed for evidence of local adaptation in various phenotypic traits (see Chapter 2).

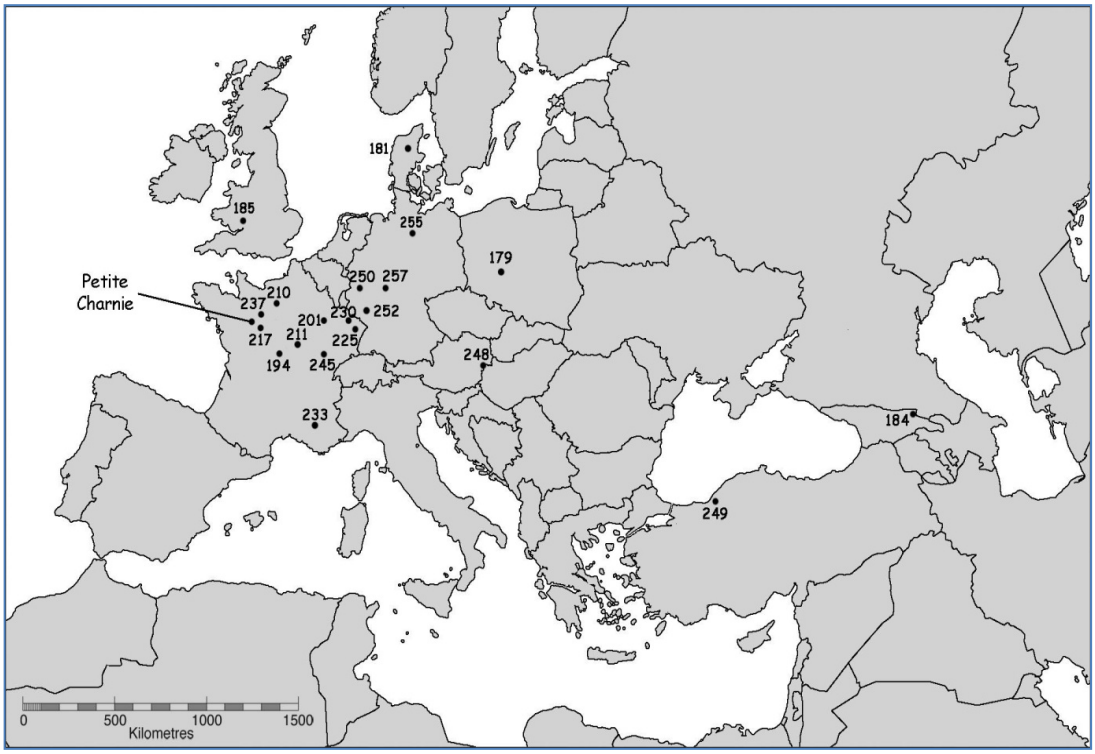


Figure 3.1. Location of La Petite Charnie provenance trials and the 20 selected study provenances. Provenance codes follow Table 3.1.

Measures of spring bud-burst phenology in 1995 (*Budburst*), autumn leaf-fall phenology in 2001 (*Retention*), tree diameter at a height of 1.3 meters in 2001 (*DBH*), and tree shape in 2001 (*Form*), were provided by INRA for all trees of the study provenances (full details of how these traits were measured are provided in Appendix 2.1). Patterns of differentiation in each of these traits for a subset of the study provenances were indicative of local adaptation to the environment at their sites of origin (see Chapter 2). Although not particularly recent, these measurements were assumed to provide a representation of relative differences between trees at the time of this study (2008 and 2009). Independent studies have reported that while the

timing of spring bud-burst and autumn leaf-fall in oaks can vary considerably between years, rankings of individual trees are relatively constant (Askew 1962, Crawley and Akhteruzzaman 1988). To allow investigation of this assumption, measurements of bud-burst phenology were repeated on April 15th 2009 (referred to as *Budburst*$_{2009}$), following the original scoring system for a subset of the trees from 8 of the study provenances (all living trees within 2 parcelles of each provenance in both soil zones 1 & 2, i.e. a maximum of 768 trees).

Table 3.1. Summary of the 20 studied provenances showing their 3 digit provenance codes, their geographic locality (Forest and Country), latitude (Lat), longitude (Lon), altitude (Alt), and the mean values across all trees of each provenance for the phenotypic traits *Budburst, Retention, DBH,* and *Form*.

Code	Forest	Country	Lon (DD)	Lat (DD)	Alt (m)	Mean *Budburst*	Mean *Retention*	Mean *DBH*	Mean *Form*
179	Sycow	Poland	17.93	51.18	210	1.42	2.58	104	4.76
181	Horbylunde	Denmark	9.41	56.13	80	0.88	3.04	89	3.99
184	Telavi	Georgia	45.47	41.88	700	3.79	0.28	77	3.06
185	Blakeney	UK	-2.5	51.78	76	1.08	2.20	120	4.24
194	Soudrain	France	2.38	46.95	178	1.24	2.64	110	4.33
201	La Haie Renaut	France	4.95	48.67	180	1.77	1.72	109	4.24
210	Saint Germain	France	2.08	48.9	60	1.71	2.68	113	4.13
211	Prémery	France	3.6	47.2	300	1.55	2.64	106	4.34
217	Bercé	France	0.39	47.81	155	1.59	2.42	98	4.20
225	Still	France	7.25	48.58	688	1.55	2.01	105	4.56
230	Romersberg	France	6.73	48.82	220	1.06	2.81	101	4.17
233	Vachères	France	5.63	43.98	650	3.69	1.50	94	4.15
237	Réno Valdieu	France	0.67	48.5	230	1.65	2.75	116	4.14
245	Etangs	France	4.96	46.93	200	2.40	2.21	113	4.19
248	Klostermarienberg	Austria	16.57	47.41	310	3.23	1.61	103	3.90
249	Bolu	Turkey	31.67	40.92	1200	1.58	0.48	94	4.39
250	Cochem	Germany	7.05	50.08	400	1.86	2.44	112	4.43
252	Johanneskreuz	Germany	7.83	49.4	460	1.03	2.34	116	4.42
255	Spakensehl	Germany	10.6	52.8	115	0.54	2.67	102	4.74
257	Wolfgang	Germany	9.05	50.15	160	1.61	2.46	96	3.92

3.2.2. Gall surveys

For each of the study provenances, 2 parcelles were selected from within each of the 5 soil zones in Tranch 4 of the provenance trial, and in Spring 2008, 12 living trees

were chosen from within each study parcelle (total=2400 trees). To minimise edge effects, trees were preferentially selected from the two internal columns of the parcelle, but the nearest alternative was used where an internal tree had died. All trees were surveyed once for sexual generation Cynipini galls during spring (May-June), and once for asexual generation galls during autumn (August-September), in both 2008 and 2009 (in autumn 2008 soil zone 5 was not surveyed due to time constraints). In the rare instances where a tree had died between surveys, the nearest alternative within the parcelle was used for subsequent surveys. At each visit to a particular tree, 10 twigs (defined as a module of woody growth over the two previous years) were haphazardly selected, and all parts of its most terminal shoot (defined as a module of woody growth from the previous year) were inspected for galls.

3.2.3. Predictor variables

The selection of appropriate predictor variables for modelling ecological data is a subjective process, dependent on the available data and the questions of interest. Recent reviews suggest that only predictors, transformations, and interactions with strong *a priori* biological justification should be included (Bolker et al. 2009, Grueber et al. 2011).

Given the plant vigour and stress hypotheses, and the likely importance of phenological synchronisation, there are clear *a priori* biological reasons for including predictors relating to tree productivity (i.e. *DBH*), health (i.e. *Form*), and phenology (i.e. *Budburst* and *Retention*) in models of gall abundances and community structure. Any effects of health or productivity are expected to be relatively linear (e.g. under the vigour hypothesis, herbivore abundance would positively correlate with rate of linear growth, Price 1991) and the predictors *DBH* and *Form* are therefore considered on their natural scale. The effects of the phenological variables however might not be constant, with the synchronisation hypothesis predicting a decline in herbivore abundance as phenology deviates in either direction from a particular optimal value. The *Budburst* and *Retention* predictor variables are therefore considered on their natural scale, and with a quadratic transformation (i.e. $Budburst^2$ and $Retention^2$).

Interaction terms within a model reflect the possibility that the effect of a particular predictor may depend upon the value of one or more additional predictors. While such interactions may be of practical interest, their higher-order effects can be difficult to interpret, and their inclusion can over-parameterise models resulting in poor parameter estimates (i.e. failed convergence). To prevent these issues, only pair-wise interactions between *Budburst, Retention, DBH,* and *Form* in their standard states were included here. Furthermore, it was considered that co-linearity between predictors would limit the relevance of their pair-wise interaction (i.e. because for a given value of one, the probable range of the other would be relatively small), and the four standard state predictors were therefore assessed for co-linearity (described in section 2.3.7), with subsequent exclusion of the interaction between co-linear pairs.

3.2.4. Mixed effect modelling

Given the hierarchical structure of the collected data (i.e. shoots nested within trees, within parcelles, within soil-zones), traits measured at the various scales were considered to be subject to the nested random effects of their higher order groupings, and the crossed random effect of provenance. To incorporate this complex random effects structure, data were analysed using mixed effects models with the *lmer* function of the *lme4* package (Bates et al. 2011) in R version 2.13.0 (R Development Core Team 2011). Models were implemented with maximum likelihood estimation to allow for comparison of models with differing fixed effect structures (Bolker et al. 2009).

Measurements taken at the parcelle level (i.e. for investigation of community structure) were modelled with random effects for soil-zone and provenance following Format-1 (Table 3.2). Measurements taken at the tree level (i.e. for investigation of co-linearity between phenotypic traits) were modelled with random effects for soil-zone, parcelle, and provenance, following Format-2 (Table 3.2). With the exception of gall-type richness, the response variables at these levels were either continuous or ordinal with a moderate number of levels (e.g. for *Budburst, Retention* and *Form*), and were therefore analysed using linear mixed effect models with a

Gaussian error family. Models were assessed for heteroscedasticity and normality of errors by plotting standardised residuals against fitted values, and ordered residuals against the quantiles of the normal distribution (Crawley 2007).

Data for gall-type richness per tree and galls per shoot were in the form of counts, and were therefore analysed using generalised linear mixed models with a Poisson error family (Bolker et al. 2009). In addition to random effects for provenance and the various groupings, these Poisson error models also contained a random effect with a number of unique factor levels equal to the number of observations (called '*observations*' in Table 3.2), to account for over-dispersion. Models for gall-type richness therefore followed Format-3, and models of galls per shoot followed Format-4 (Table 3.2).

Table 3.2. Description of the different random effect formats used in the modelling. *SoilZone, Parcelle, Tree, Provenance* and *Observation* are all categorical variables and are described in the text. Notation is provided to show how each format is described for the *lmer* function of the *lme4* package in R.

	Random effects	
	Description	**R notation**
Format-1	*SoilZone* crossed with *Provenance*	(1 | *SoilZone*) + (1 | *Provenance*)
Format-2	*Parcelle* nested within *SoilZone*, crossed with *Provenance*	(1 | *SoilZone / Parcelle*) + (1 | *Provenance*)
Format-3	*Parcelle* nested within *SoilZone*, crossed with *Provenance*, crossed with *Observation*	(1 | *SoilZone / Parcelle*) + (1 | *Provenance*) + (1 | *Observation*)
Format-4	*Tree* nested within *Parcelle*, nested within *SoilZone*, crossed with *Provenance*, crossed with *Observation*	(1 | *SoilZone / Parcelle / Tree*) + (1 | *Provenance*) + (1 | *Observation*)

3.2.5. Multi-model inference

In ecological studies, it is often of interest to examine the relationships between a response variable and several potential covariates to determine which, if any, are important predictors of variation in the response. Approaches to such questions have traditionally involved stepwise comparison of nested models to identify a single

model that contains only predictors deemed to explain a significant amount of deviance in the response, as determined by null hypothesis testing (Crawley 2007). However, such approaches have been criticised on several grounds, including the dependency of the identified model on the employed selection algorithm (Calcagno and de Mazancourt 2010), and issues of multiple hypothesis testing (Whittingham et al. 2006). An alternative that is increasing in popularity is to use information criteria (IC) such as the Akaike information criteria (AIC, Akaike 1974) to compare the performance of multiple competing models.

The AIC and related ICs use deviance as a measure of the fit of a particular model to a given dataset, with a penalty applied for the number of estimated parameters. The AIC is generally used in its corrected form (AICc), to account for potentially small samples. When multiple predictor variables are being considered, models containing all possible combinations can be ranked in order of performance by their IC score to identify the best approximating model. Various statistical software packages are available for automating the calculation of IC scores for potentially large model sets (e.g. Calcagno and de Mazancourt 2010, Barton 2011). Additional derived statistics such as model weight (the probability that a particular model is the best approximating model within a set) and evidence ratios (a measure of how much more likely one model is compared to another) can be used to assess model uncertainty (full details on the calculation of AIC and related statistics are provided by Symonds and Moussalli 2011). In situations where no single model is clearly superior to all others (i.e. the model weight of the best approximating model does not approach 1), model averaging can be employed to account for model uncertainty and obtain robust parameter and error estimates across multiple models, where the contribution of each model is weighted by its relative performance (Grueber et al. 2011, Symonds and Moussalli 2011). The model averaged parameter estimate for a particular predictor can either be based on all models, where it receives a value of zero for models that do not contain the predictor (termed ‘full-model averaging’), or can be based only on those models that do feature the predictor (termed ‘natural-averaging’, Symonds and Moussalli 2011).

In this study, the relationships between various response variables (i.e. counts of gall abundance, species richness, and community similarity scores) and predictor variables (i.e. tree provenance, or tree phenotypic traits) were investigated through an IC based approach. Differences in AICc scores and evidence ratios were used to compare the performance of models and make inferences about the importance of particular predictor variables. Where assessing the influence of multiple predictors, these were centred and standardized to allow for the interpretation of parameter estimates in models containing interaction terms (Gelman 2008, Schielzeth 2010). The effect and the importance of individual predictors were assessed from estimates of their slope parameters (with confidence intervals), obtained through model averaging. As there was potential for co-linearity of predictors, natural averaging rather than full-model averaging was applied, to avoid shrinkage towards zero.

The use of AICc and related ICs is subject to issues of boundary effects and uncertainty in the estimation of degrees of freedom in models that contain random effects (Bolker et al. 2009). Such estimation of degrees of freedom for random effects is not straightforward, and although there is no clear consensus approach, the default method adopted in the packages used here (i.e. using the minimum of 1 d.f for each random effect in the model) is potentially dubious (Bolker et al. 2009). However, as these issues also apply to alternative methods, such as likelihood ratio testing, the AICc approach was still considered to be the most appropriate option for the analysis presented here.

3.2.6. Non-independence of provenances

A common shortcoming of studies of intraspecies populations is that the populations are treated as statistically independent entities. This is unlikely to be true, as populations that are genetically similar through gene-flow or phylogeographic history, can be expected to co-vary in traits independently of any population-specific effects (Stone et al. 2011). Where the effects of non-independence are severe, failure to address this statistically will increase error rates in the inference of population-specific effects.

In Chapter 2, this issue was addressed for a sample of 17 provenances with a genetic autocorrelation approach. The primary axis of a principle coordinate analysis of data from microsatellite markers was used as a linear measure of genetic distance between provenances, and this axis (referred to as *PCoA*) was included as a fixed effect in models of phenotypic traits to remove the portion of trait variation explained by genealogical correlation. This approach would also be applicable to the present study, but microsatellite data was only available for 17 of the 20 study provenances. Therefore, all analysis was initially conducted for the subset of 17 provenances, both with and without *PCoA* as an additional fixed effect. While inclusion of *PCoA* had a small effect on parameter estimates, it resulted in little change in the direction or the significance of relationships. As the primary intention of the modelling was to reveal the nature of important influences, the effect of non-independence of provenances was considered to be negligible, and was not accounted for in the analysis of the full set of 20 provenances presented here.

3.2.7. Co-linearity between tree phenotypic traits

For all trees of the 20 study provenances, the relationship between all pairings of the phenotypic traits *Budburst, Retention, DBH* and *Form* were investigated by modelling individual tree measurements with the corresponding measurement of another trait included as a fixed effect, and random effects following format-2 (Table 3.2). The significance of the relationship was assessed by two means: firstly by examination of 95% confidence intervals for the slope parameter of the fixed effect, as derived from a posterior distribution of 50,000 samples generated using Markov Chain Monte Carlo methods with the *mcmcsamp* and *pvals.fnc* functions from the *lme4* and *LanguageR* R packages respectively (Baayen 2011, Bates et al. 2011); and secondly by comparison of corrected Akaike information criterion (AICc) scores between models with and without the second trait as a fixed effect.

For trees where measures of spring bud-burst phenology were available from both 1995 and 2009, the relationship between these measures was investigated by modelling $Budburst_{2009}$ with the corresponding measurements of *Budburst* included as a fixed effect, and random effects following Format-2 (Table 3.2). Again, the

significance of the relationship was assessed through the estimation of the fixed effect slope parameter and its confidence intervals, and through comparison of AICc between models with and without the fixed effect.

3.2.8. Variation in abundance of individual gall-types

To assess variation in the abundance of individual gall-types between tree provenances, measurements of galls per shoot were modelled with provenance included as a fixed effect, and random effects following Format-4 (Table 3.2). This was conducted separately for each gall-type for each of the two survey years. Models with and without the fixed effect were compared by examination of AICc scores and the evidence ratio between them. Where estimation of an additional parameter for each provenance (i.e. when provenance was included as a fixed effect) resulted in an improved model despite the penalty for additional complexity, provenance was considered to explain a substantial amount of the variation in gall abundance.

For those gall-types that showed a substantial provenance effect in at least one of the survey years, the relationship with tree phenotypic traits was investigated through multi-model inference by modelling of gall counts with various predictor variables included as fixed effects (these predictors included *Budburst, Retention, DBH, Form,* the interactions between non co-linear pairings of these, $Budburst^2$ and $Retention^2$) with random effects following Format-4 (Table 3.2). Global models containing all the predictors were standardised to allow for model averaging and comparison of their relative effects using the *stdz.model* function of the *arm* R package (Gelman 2008, Gelman et al. 2011, Grueber et al. 2011), and all possible sub-models were tested and ranked by their AICc scores using the *dredge* function of the *MuMIn* package in R (Barton 2011). Model averaged estimates of standardized fixed effect slope parameters with their standard errors and confidence intervals were obtained by natural averaging using the *model.avg* function of the *MuMIn* R package. Predictors were considered to be of significance where the 95% confidence intervals of their slope parameter estimate did not include zero. Significant relationships were visualised and effect ranges estimated using the *plotLMER.fnc* function of the

LanguageR R package (Baayen 2011), as applied to un-standardized models containing all significant predictors.

3.2.9. Community level variation

To assess variation in community richness, a measure was obtained for each tree in each survey season as the number of gall-types encountered across all 10 shoots. As for the analysis of individual gall-types, variation in richness between provenances was assessed by comparing models with and without provenance as a fixed effect, with random effects following Format-3 (Table 3.2). Relationships with the selected phenotypic predictor variables were also investigated through multi-model inference, as described for gall-types but with random effects following Format-3 (Table 3.2).

Variation in community structure (i.e. species identity and relative abundance) was assessed using non-metric multidimensional scaling (NMDS), an ordination technique that allows variation in community data to be summarised along one or more axes (Whitham et al. 2006). To avoid the difficulty of applying NMDS to units (i.e. shoots or trees) where all species were absent, or the bias of completely excluding such units, the gall count data were summed for each parcelle in each survey season to give values for the total number of galls of each type from 120 shoots (10 from each of 12 trees). Analysis was performed using the *metaMDS* function of the *Vegan* R package (Oksanen et al. 2011), with Wisconsin double standardization of large values, a Bray-Curtis dissimilarity index, and 50 random starts (a method for reducing risk of convergence on local optima), to obtain a single axis of variation. This axis is thus a linear representation of community similarity, and parcelles that have similar values support similar communities. Variation in community structure between provenances was assessed as for species richness and gall abundances, by comparing models with and without provenance as a fixed effect, with random effects following Format-1 (Table 3.2). The relationship with phenotypic predictor variables was investigated by averaging the relevant phenotypic traits across study trees within each parcelle, and modelling NMDS scores with parcelle averages of the predictor variables included as fixed effects. Multi-model inference was conducted as for gall counts and species richness data.

3.3. Results

3.3.1. Co-linearity of phenotypic traits

Modelling of tree phenotypic traits against one another indicated negative co-linearity between *Budburst* and *Retention,* and positive co-linearity between *Retention* and *DBH,* and *DBH* and *Form.* These relationships were supported by both methods of inference (i.e. assessment of confidence intervals for fixed effect slope parameters, and comparison of AICc for models with and without one trait as a fixed effect) and in both directions of pairing (i.e. when either of a pair was the response variable, see Table 3.3). When *Form* was modelled with *Budburst* as a fixed effect, 95% confidence intervals for the slope parameter of *Budburst* were both less than zero, suggesting some negative co-linearity between these traits. Similarly, when *DBH* was modelled with *Budburst* as a fixed effect, the model had a slightly lower AICc score than the model without *DBH*, indicating some co-linearity. However, as these relationships were not supported by the alternate means of inference, or by either means when the pairings were reversed (Table 3.3), the co-linearity between these pairs of traits was considered to be negligible. There was no indication of co-linearity between *Retention* and *Form*. As co-linearity was considered to limit the relevance of pair-wise interaction terms as predictor variables, only interaction terms for *Budburst* and *Form, Budburst* and *DBH*, and *Retention* and *Form* were included as predictors in the modelling of gall abundance and community data.

For a sample of 582 trees, measurements of spring bud-burst phenology could be compared between 2009 and 1995. Modelling of the measurements from 2009 ($Budburst_{2009}$) against those from 1995 (*Budburst*) indicated strong positive co-linearity, supported by both means of inference (Table 3.3 and Figure 3.2). This result indicates a general consistency in relative phenology between 1995 and 2009.

Table 3.3. Summary results for modelling of relationships between phenotypic traits, showing the pair-wise combinations of response and predictor variables, estimates of the slope parameter for predictor variables with 95% confidence intervals, and the difference in AICc score between models with and without the predictor variables ($Diff_{AICc}$). Estimates of Slope or $Diff_{AICc}$ that suggested co-linearity are highlighted in bold font.

Response	Predictor	Slope	lower 95% CI	Upper 95% CI	$Diff_{AICc}$
Budburst	*Retention*	**-0.0793**	-0.1295	-0.042	**-4.935**
Retention	*Budburst*	**-0.0766**	-0.1173	-0.0401	**-6.846**
Budburst	*DBH*	0.0008	-0.0002	0.00021	13.043
DBH	*Budburst*	0.9627	1.113	-0.4231	**-0.78**
Budburst	*Form*	-0.0443	-0.0952	0.0017	4.298
Form	*Budburst*	**-0.0416**	-0.0762	-0.008	2.746
Retention	*DBH*	**0.0029**	0.0019	0.0039	**-15.976**
DBH	*Retention*	**5.023**	3.429	6.545	**-38.89**
Retention	*Form*	0.0224	-0.0206	0.0693	6.772
Form	*Retention*	0.0226	-0.0114	0.0614	6.672
DBH	*Form*	**3.943**	2.259	5.851	**-18.3**
Form	*DBH*	**0.0022**	0.0013	0.0031	**-5.93**
Budburst$_{2009}$	*Budburst*	**0.1756**	0.122	0.2539	**-25.331**

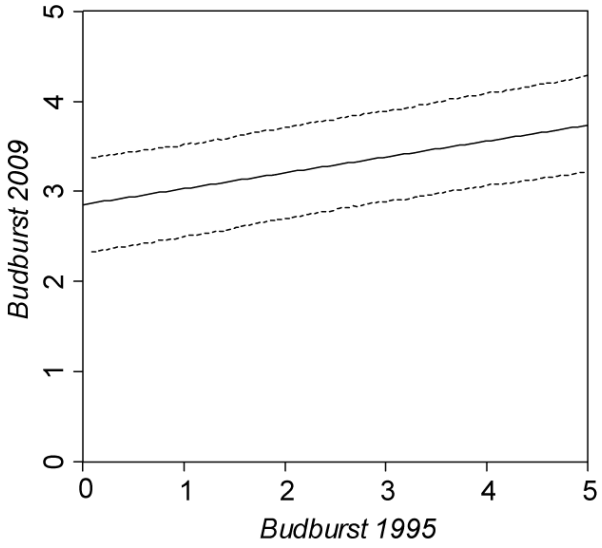


Figure 3.2. Illustration of the estimated relationship (solid line) with 95% confidence intervals (dashed lines) between spring bud-burst phenology for the same 582 trees in 2009 and 1995. Estimates of the slope parameter and its confidence intervals are provided in Table 3.3. Axis units are on an ordinal scale from closed buds (0) to fully developed leaves (5 - scoring methodology is described in detail in Appendix 2.1).

3.3.2. Influence of tree provenance and phenotype on gall abundances

During the four survey seasons approximately 725,000 galls were recorded, including seven sexual generation and 13 asexual generation gall-types, and representing 15 different gall-former species (Table 3.4). The total counts of individual gall-types varied dramatically, ranging from just 4 galls of the rarest type (the asexual generation of *Cynips longiventris*) to over 330,000 of the most common (the asexual generation of *Neuroterus anthracinus*). The abundance of individual gall-types also varied substantially between survey years, with seven types exhibiting a greater than 10 fold difference. More than 1000 galls were recorded in at least one year for 11 of the gall-types.

Modelling of the abundance of individual gall-types indicated that provenance explained a substantial amount variation in at least one year for 10 of the 20 types (a total of 15 cases), as inferred where the inclusion of provenance as a fixed effect resulted in an improved model (i.e. with a lower AICc score, see Tables 3.4). The range of differences in provenance mean galls per shoot varied considerably between these gall-types, from 0.22 for several types, to 20.8 for asexual generation *N. anthracinus* in autumn 2008 (Table 3.5). The most local provenance to the trial site (provenance 217) had the highest mean gall abundance for only one of the 15 cases. All cases of a substantial provenance effect occurred where more than 1000 galls of a particular type were found in a season. The degree of model improvement through inclusion of provenance as a fixed effect ranged from being very slight for asexual generation *N. numismalis* in 2009 (i.e. model with fixed effect was 1.2 times more likely to be better than the model without), to being very large for sexual generation *N. quercusbaccarum* in 2008 (i.e. model with fixed effect was 1.2×10^{39} times more likely to be better than the model without, Table 3.4).

For the gall-types that showed a substantial provenance effect in at least one year, further modelling with combinations of the tree phenotype predictor variables (i.e. *Budburst, Retention, DBH, Form, Budburst*^{2}*, Retention*^{2}*, Budburst:DBH, Budburst:Form, Retention:Form*) revealed best approximating models that

performed better than those without any fixed effects in all cases (see Table 3.6). The degree of improvement in model performance ranged from being very slight for asexual generation *Cynips divisa* in autumn 2009 (i.e. best approximating model was 1.12 times more likely to be better than model without), to being very large for sexual generation *N. quercusbaccarum* in 2008 (i.e. best approximating model was 1.13×10^{22} times more likely to be better than the model without). Various significant relationships between tree phenotypes and the abundance of particular gall types were apparent, as inferred where the 95% confidence intervals for the slope parameter of a predictor did not include zero. Details of the parameter estimates for these significant predictor variables are provided in Table 3.6, and plots of the relationships are shown in Figure 3.3.

Earlier spring bud-burst (i.e. higher *Budburst* scores) corresponded with increasing abundance of the sexual generation gall-types of *N. anthracinus* in 2008, *N. numismalis* in 2009, and *N. quercusbaccarum* in both 2008 and 2009, with trees of the earliest bud-burst phenology supporting between 15 and 115% more galls per shoot than those with the latest phenology (Figure 3.3). An opposite relationship was observed for asexual generation gall-types *N. anthracinus* in 2009 and *N. quercusbaccarum* in 2008 and 2009, with trees of the earliest bud-burst phenology supporting between 40 and 65% less galls per shoot than those with the latest phenology (Figure 3.3). For the sexual generation galls of *N. numismalis* in 2009, the relationship appeared to involve a positive interaction with *Form*, with the magnitude of the positive correlation between gall abundance and *Budburst* being greatest where the values of *Form* were high (Figure 3.3b). For the asexual generation galls of *N. anthracinus*, a significant positive parameter estimate for the squared transformation of *Budburst* indicated a curved relationship, with gall abundance decreasing towards a minimum at a *Budburst* score between 3 and 4 (i.e. towards the higher end of its range), before increasing again slightly (Figure 3.3d).

Later autumn leaf-fall (i.e. higher *Retention* score) corresponded with increasing abundance of the asexual generation gall-types of *N. albipes* in 2008, and *N. quercusbaccarum* in both 2008 and 2009, with trees of the latest leaf-fall phenology supporting between 60 and 145% more galls per shoot than those with the earliest

phenology (Figure 3.3). A curved relationship with sexual generation *N. quercusbaccarum* in 2008 was also indicated a significant positive parameter estimate for the squared transformation of *Retention*, with gall abundance declining slightly to either side of a maximum between a *Retention* score of between 1 and 2.

Greater tree diameter (*DBH*) corresponded with decreasing abundance of the sexual generation gall-types of *N. anthracinus* and *N. quercusbaccarum* in 2008, and the asexual generation gall-type of *N. quercusbaccarum* in 2008, with trees of the greatest diameter supporting between 20 and 65% less galls than those with the smallest diameter. The opposite relationship was observed for the sexual generation gall-types of *N. anthracinus* and *N. quercusbaccarum* in 2008, and the asexual generation of *N. albipes* in 2009 with trees of the greatest diameter supporting between 45 and 70% more galls than those of the smallest diameter (Figure 3.3).

Table 3.4. Summary results for modelling of gall-type abundances in relation to provenance, showing the difference in AICc scores ($Diff_{AICc}$) and the evidence ratios (ER) between models with and without provenance as a fixed effect, the number of galls (# G) and the mean number of galls per shoot (μG_{ps}) recorded in each year, and the ratios of gall numbers between 2008 and 2009. Values of $Diff_{AICc}$ that indicated a substantial provenances effect are highlighted in bold font.

Gall-type	2008				2009				Ratio $G_{2008} : G_{2009}$
	$Diff_{AICc}$	**ER**	**# G**	**μG_{ps}**	**$Diff_{AICc}$**	**ER**	**# G**	**μG_{ps}**	
Andricus inflator (Sex)	35.17	4.3×10^{7}	52	2.2×10^{-3}	38.05	1.8×10^{8}	0	0.00	52 : 0
Andricus testacipes (Sex)	21.10	3.8×10^{4}	973	0.04	9.98	147	2291	0.10	1 : 2.4
Biorhiza pallida (Sex)	36.57	8.7×10^{7}	13	5.4×10^{-4}	35.91	6.3×10^{7}	10	4.2×10^{-4}	1.3 : 1
Neuroterus albipes (Sex)	21.64	5.0×10^{4}	822	0.03	**-6.67**	28	5828	0.24	1 : 7.1
Neuroterus anthracinus (Sex)	**-20.49**	2.8×10^{4}	14278	0.59	**-14.55**	1400	7411	0.31	1.9 : 1
Neuroterus numismalis (Sex)	26.42	5.5×10^{5}	1154	0.05	**-55.90**	1.4×10^{12}	17486	0.73	1 : 15.1
Neuroterus quercusbaccarum (Sex)	**-179.98**	1.2×10^{39}	5808	0.24	**-126.70**	3.3×10^{27}	7834	0.33	1 : 1.3
Andricus callidoma (Asex)	36.99	1.1×10^{8}	4	2.1×10^{-4}	34.59	3.2×10^{7}	45	1.9×10^{-3}	1 : 11.3
Andricus fecundatrix (Asex)	25.45	3.3×10^{5}	334	0.02	**-39.75**	4.3×10^{8}	1103	0.05	1 : 3.3
Andricus glandulae (Asex)	35.45	4.9×10^{7}	83	4.3×10^{-3}	29.01	2.0×10^{6}	335	0.01	1 : 4
Andricus inflator (Asex)	33.90	2.3×10^{7}	44	2.3×10^{-3}	36.91	1.0×10^{8}	3	1.2×10^{-4}	14.7 : 1
Andricus kollari (Asex)	37.32	1.2×10^{8}	2	1.0×10^{-4}	33.84	2.2×10^{7}	43	1.8×10^{-3}	1 : 21.5
Andricus solatarius (Asex)	20.83	3.3×10^{4}	528	0.03	15.19	2000	726	0.03	1 : 1.4
Cynips divisa (Asex)	**-10.74**	215	1901	0.10	28.27	1.4×10^{6}	184	7.7×10^{-3}	10.3 : 1
Cynips longiventris (Asex)	37.60	1.5×10^{8}	1	5.2×10^{-5}	37.14	1.2×10^{8}	3	1.3×10^{-4}	1 : 3
Cynips quercusfolii (Asex)	32.70	1.3×10^{7}	197	0.01	20.98	3.6×10^{4}	456	0.02	1 : 2.3
Neuroterus albipes (Asex)	**-97.41**	1.4×10^{21}	45771	2.38	**-138.98**	1.5×10^{30}	27531	1.15	1.7 : 1
Neuroterus anthracinus (Asex)	**-49.31**	5.1×10^{10}	313507	16.33	**-10.23**	170	23820	0.99	13.2 : 1
Neuroterus numismalis (Asex)	9.99	148	12194	0.64	**-0.34**	1.2	12124	0.51	1 : 1
Neuroterus quercusbaccarum (Asex)	**-114.26**	6.5×10^{24}	123708	6.44	**-134.74**	1.8×10^{29}	94830	3.95	1.3 : 1

Table 3.5. The minimum, median, and maximum values of mean number of galls per shoot (**μ** G_{ps}) by provenance for gall-types where a substantial provenance effect was observed in a least one survey year (i.e. where inclusion of provenance as a fixed effect resulted in a lower model AICc score, see Table 3.4, highlighted here in bold font). Provenance codes follow Table 3.1.

Gall-type	2008							2009						
	Min		Median		Max		Range	Min		Median		Max		Range
	Prov	μ G_{ps}	Prov	μ G_{ps}	Prov	μ G_{ps}		Prov	μ G_{ps}	Prov	μ G_{ps}	Prov	μ G_{ps}	
Neuroterus albipes (Sex)	233	0.01	194	0.04	217	0.07	0.06	**255**	**0.12**	**248**	**0.23**	**217**	**0.34**	**0.22**
Neuroterus anthracinus (Sex)	**249**	**0.26**	**257**	**0.62**	**245**	**0.77**	**0.51**	**249**	**0.18**	**225**	**0.29**	**233**	**0.46**	**0.28**
Neuroterus numismalis (Sex)	255	0.03	230	0.05	201	0.08	0.05	**255**	**0.31**	**252**	**0.62**	**248**	**1.57**	**1.21**
Neuroterus quercusbaccarum (Sex)	**185**	**0.12**	**250**	**0.22**	**245**	**0.41**	**0.29**	**255**	**0.21**	**201**	**0.31**	**248**	**0.57**	**0.36**
Andricus fecundatrix (Asex)	184, 248	0.00	211	0.01	185	0.04	0.04	**184, 245 248, 249**	**0.00**	**237**	**0.03**	**255**	**0.22**	**0.22**
Cynips divisa (Asex)	**249**	**0.02**	**217**	**0.10**	**185**	**0.24**	**0.22**	184,233 248	0.00	245	0.01	185	0.02	0.02
Neuroterus albipes (Asex)	**248**	**0.23**	**252**	**2.41**	**185**	**4.04**	**3.81**	**249**	**0.22**	**210**	**0.99**	**185**	**3.40**	**3.18**
Neuroterus anthracinus (Asex)	**249**	**3.03**	**194**	**17.60**	**225**	**23.83**	**20.8**	**249**	**0.10**	**179**	**0.99**	**185**	**2.18**	**2.08**
Neuroterus numismalis (Asex)	233	0.01	217	0.40	185	2.01	2.0	**233, 248**	**0.00**	**211**	**0.22**	**252**	**1.75**	**1.75**
Neuroterus quercusbaccarum (Asex)	**233**	**0.91**	**211**	**6.79**	**255**	**10.55**	**9.64**	**248**	**0.21**	**225**	**4.22**	**255**	**8.36**	**8.15**

Table 3.6. Summary results for modelling of gall-type abundances in relation to phenotypic predictor variables, showing the difference in AICc scores ($Diff_{AICc}$) and the evidence ratios (ER) between the best approximating models and those without any fixed effects. Estimates of slope parameters, their standard errors (SE), and confidence intervals, are presented for significant relationships (i.e. where 95% confidence intervals did not include zero). 'NS' indicates that no significant relationships were identified. The table is continued on the following page.

Gall-type	Year	$Diff_{AICc}$	ER	Predictor	Estimate	SE	Lower 95% CI	Upper 95% CI
Neuroterus albipes (sex)	2008	-3.97	7.27	NS				
	2009	-1.19	1.81	NS				
Neuroterus anthracinus (sex)	2008	-5.48	15.48	*Budburst*	0.090	0.045	0.001	0.178
				DBH	-0.086	0.036	-0.158	-0.015
	2009	-12.28	463.35	*DBH*	0.110	0.037	0.038	0.182
Neuroterus numismalis (sex)	2008	-2.78	4.01	NS				
	2009	-53.21	3.58×10^{11}	*Budburst*	0.398	0.059	0.283	0.513
				DBH	0.125	0.049	0.028	0.221
				Budburst : Form	0.196	0.093	0.015	0.378
Neuroterus quercusbaccarum (sex)	2008	-101.55	1.13×10^{22}	*Retention*^{2}	-0.284	0.130	-0.538	-0.029
				Budburst	0.481	0.080	0.325	0.637
				DBH	-0.159	0.064	-0.284	-0.034
				Retention	-0.149	0.072	-0.290	-0.007
	2009	-40.81	7.26×10^{8}	*Budburst*	0.249	0.053	0.145	0.353

Table 3.6. Continued.

Gall-type	Year	$Diff_{AICc}$	ER	Predictor	Estimate	SE	Lower 95% CI	Upper 95% CI
Andricus fecundatrix (asex)	2008	-2.20	3.00	NS				
	2009	-24.17	1.77×10^5	NS				
Cynips divisa (asex)	2008	-18.42	1×10^3	NS				
	2009	-0.22	1.12	NS				
Neuroterus albipes (asex)	2008	-12.59	541.86	*Retention*	0.228	0.071	0.089	0.367
	2009	-11.12	259.82	*DBH*	0.152	0.062	0.030	0.274
Neuroterus anthracinus (asex)	2008	-10.50	190.57	NS				
	2009	-18.73	1.17×10^4	$Budburst^2$	0.331	0.130	0.077	0.585
				Budburst	-0.413	0.088	-0.586	-0.240
Neuroterus numismalis (asex)	2008	-14.62	1.50×10^3	NS				
	2009	-14.15	1.18×10^3	NS				
Neuroterus quercusbaccarum (asex)	2008	-68.85	8.92×10^{14}	*Budburst*	-0.500	0.109	-0.713	-0.287
				DBH	-0.355	0.080	-0.510	-0.199
				Retention	0.492	0.088	0.319	0.664
	2009	-69.37	1.16×10^{15}	*Budburst*	-0.551	0.081	-0.709	-0.392
				Retention	0.290	0.070	0.154	0.427

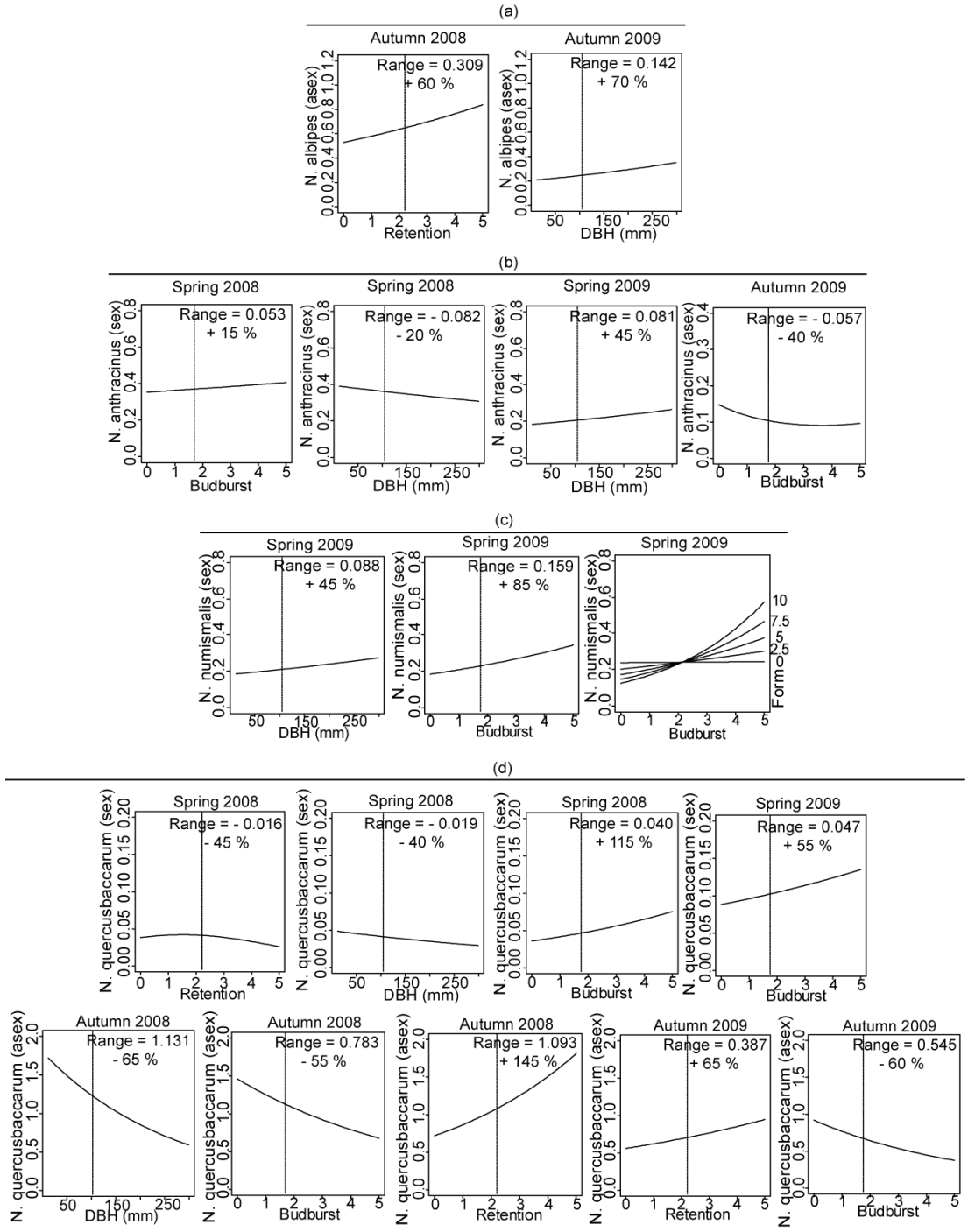


Figure 3.3. Graphical illustration of significant relationships between gall abundance per shoot and phenotypic predictor variables for (a) *Neuroterus albipes,* (b) *N. anthracinus,* (c) *N. numismalis,* (d) *N. quercusbaccarum*. The estimated range of the response variable (y-axis) across the observed range of a predictor variable (x-axis) is provided as a value, and as a percentage of y at the minimum observed value of *x*. Dashed vertical lines indicate the mean predictor value for the provenance most local to the trial site (i.e. provenance 217, Bercé). Values on the right axis of a plot involving an interaction indicate the values of the second interaction term at which the relationship between the first term and the response are plotted.

3.3.3. Influence of tree provenance and phenotype on gall community structure

Modelling of the gall-type richness data indicated that provenance explained a substantial amount of variation in spring 2008, autumn 2008, and autumn 2009, but not in spring 2009, as inferred where the inclusion of provenance as a fixed effect resulted in an improved model (i.e. with a lower AICc score). The degree of model improvement was slight for spring 2008 (i.e. the model with fixed effect was 2.74 times more likely to be better than the model without), but was large for both autumn 2008 and 2009 (i.e. models with the fixed effects were respectively 1.5×10^7 and 2×10^{10} times more likely to be better than the models without, see Table 3.6). The range of differences in provenance mean gall-types per tree were greater in the autumn seasons (2.11 and 2.87) than in spring (0.97 and 0.89). The identity of the provenance with the lowest and highest mean gall-types per tree was the same in both autumn seasons (provenance 248 from Klostermarienberg in Austria, and provenance 185 from Blakeney in the UK, Table 3.6).

Further modelling with combinations of the tree phenotype predictor variables revealed models that performed moderately better than those without any fixed effects for all four seasons (see Table 3.7). Various significant relationships were apparent, as inferred where the 95% confidence intervals for the slope parameter of a predictor did not include zero. Details of the parameter estimates for these significant predictor variables are provided in Table 3.7, and plots of the relationships are shown in Figure 3.4.

Earlier spring bud-burst (i.e. high *Budburst* score) corresponded with higher gall-type richness in spring 2008, with trees of the earliest bud-burst phenology supporting 15% more gall-types than trees with the latest phenology (Figure 3.4a). This relationship was reversed in autumn 2009, with trees of the earliest bud-burst phenology supporting 15% less gall-types than those with the latest phenology (3.4d). Later autumn leaf-fall (i.e. high *Retention* scores) corresponded with increasing gall-type richness in autumn 2008, with trees of the latest leaf-fall phenology supporting 25% more gall-types than those with the earliest phenology (Figure 3.4c). Greater tree diameter (i.e. higher *DBH)* corresponded with increasing

gall-type richness in spring 2009, with trees of the greatest diameter supporting 20% more gall-types than those with the smallest diameter (Figure 3.4b).

Table 3.6. Summary results for modelling of gall-type richness in relation to provenance for each of the four survey seasons, showing the difference in AICc scores ($Diff_{AICc}$) and the evidence ratios (ER) between models with and without provenance as a fixed effect, the number of gall-types recorded (# GT), and the mean number of gall-types per tree (**μ** GT) across all trees and within the minimum, median, and maximum provenances. Provenance codes follow Table 3.1.

Season	$Diff_{AICc}$	ER	# GT	μ GT	Min		Median		Max		Range μ GT
					Prov	μ GT	Prov	μ GT	Prov	μ GT	
Spring 2008	-2.01	2.74	7	2.28	184	1.88	194	2.27	245	2.85	0.97
Spring 2009	11.6	330	6	3.63	249	3.10	211	3.64	210	3.99	0.89
Autumn 2008	-33.05	1.5×10^7	13	3.66	248	2.25	225 257	3.79	185	4.36	2.11
Autumn 2009	-47.43	2.0×10^{10}	13	3.09	248	1.27	181	3.30	185	4.14	2.87

Table 3.7. Summary results for modelling of gall community response variables in relation to phenotypic predictor variables, showing the difference in AICc scores ($Diff_{AICc}$) and the evidence ratios (ER) between the best approximating models and those without any fixed effects. Estimates of slope parameters, their standard errors (SE), and confidence intervals, are presented for significant relationships. 'NS' indicates that no significant relationships with predictor variables were identified.

Response variable	Season	$Diff_{AICc}$	ER	Predictor	Estimate	SE	Lower 95% CI	Upper 95% CI
Gall-type richness	Spring 2008	-7.87	51.2	*Budburst*	0.128	0.040	0.049	0.207
	Spring 2009	-3.76	6.57	*DBH*	0.054	0.024	0.007	0.101
	Autumn 2008	-14.6	1450	*Retention*	0.107	0.030	0.048	0.165
	Autumn 2009	-11.6	323	*Budburst*	-0.097	0.032	-0.160	-0.035
NMDS score	Spring 2008	-8.89	85.3	*DBH*	0.077	0.039	0.000	0.155
				Retention	0.088	0.037	0.016	0.161
	Spring 2009	-0.45	1.25	*NS*				
	Autumn 2008	-15.5	2310	*Budburst*	0.115	0.051	0.016	0.215
				Form	-0.091	0.045	-0.179	-0.003
	Autumn 2009	-24.0	1.59 x 10^5	*Budburst*	0.239	0.065	0.112	0.367
				DBH	0.139	0.054	0.033	0.245

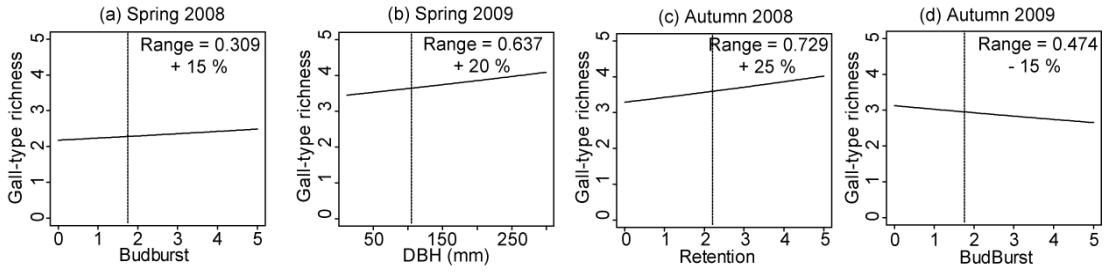


Figure 3.4. Graphical illustration of significant relationships between gall-type richness per tree and phenotypic predictor variables in each of the four survey seasons (a-d). The estimated range of the response variable (*y-axis*) across the observed range of a predictor variable (*x-axis*) is provided as a value, and as a percentage of *y* at the minimum observed value of *x*. Dashed vertical lines indicate the mean predictor value for the provenance most local to the trial site (i.e. provenance 217, Bercé).

The parcelle NMDS scores are a measure of how similar – in terms of species composition and relative abundance – a particular parcelle is to others within the sample. As such, it is the difference in scores between parcelles rather than the scores themselves that are of interest. Significant trends in relation to predictors such as provenance or mean tree phenotype indicate that parcelles with similar predictor values support gall communities that are similar in their species composition and relative abundance.

Modelling of parcelle NMDS scores indicated that provenance explained a substantial amount of variation in autumn 2008, as inferred where the inclusion of provenance as a fixed effect resulted in an improved model (i.e. with a lower AICc score, Figure 3.5d). The inclusion of provenance as a fixed effect did not improve models for spring 2008, spring 2009, or autumn 2008 (Figure 3.5a-c). Further modelling with combinations of the tree phenotype predictor variables did however reveal models that performed better than those without any fixed effects for all four seasons (see Table 3.7). Significant relationships with NMDS scores were apparent for: leaf-fall phenology and tree diameter in spring 2008; bud-burst phenology and tree shape in autumn 2008; and bud-burst phenology and tree diameter in autumn 2009; as inferred where the 95% confidence intervals for the slope parameter of a predictor did not include zero. Details of the parameter estimates for these significant predictor variables are provided in Table 3.7, and plots of the relationships are shown in Figure 3.5.

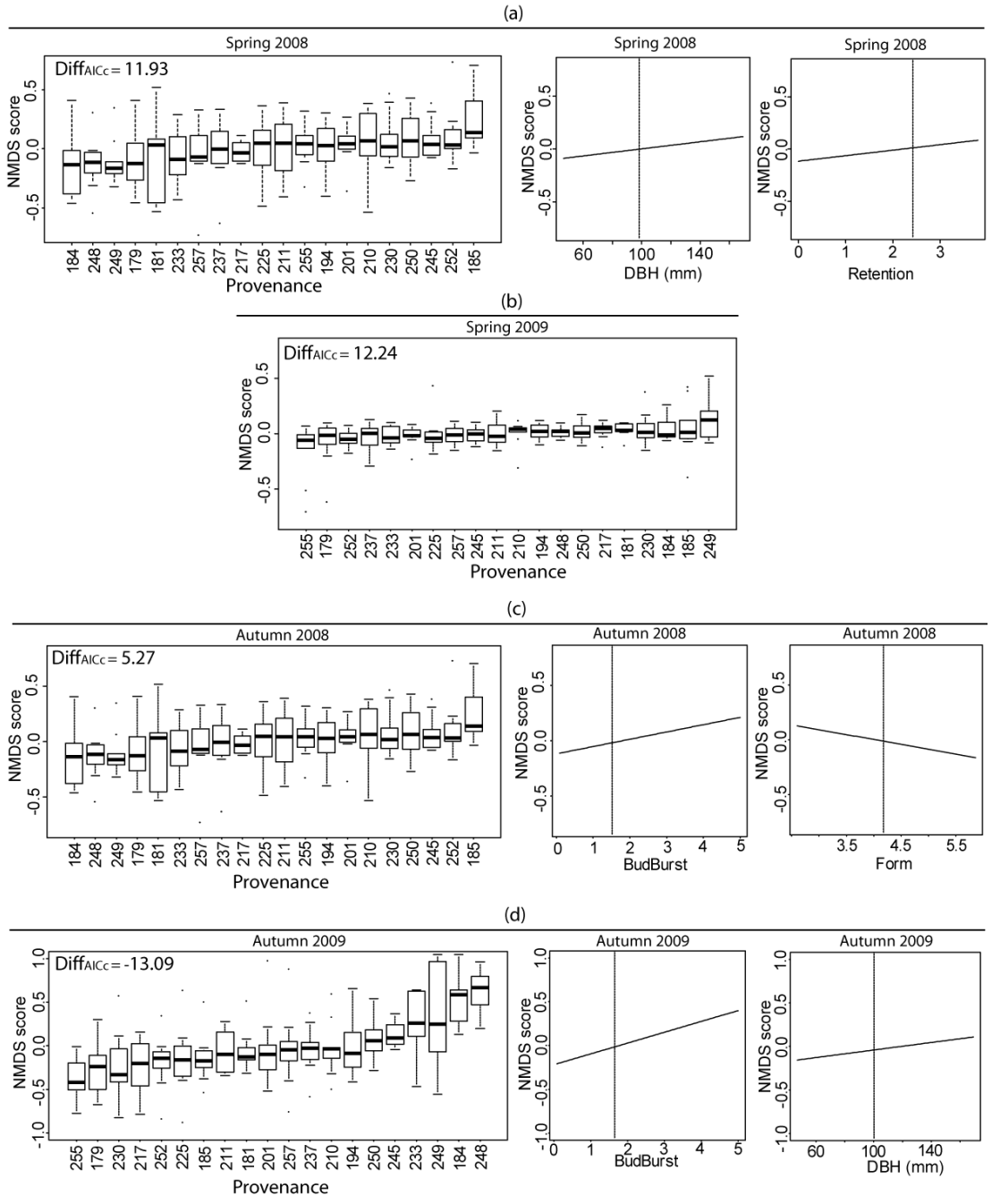


Figure 3.5. Graphical illustration of the relationships between parcelle NMDS score and tree provenance (box and whisker plots), and significant relationships with phenotypic predictor variables in each of the four survey seasons (a-d). The difference in AICc score ($Diff_{AICc}$) between models of NMDS scores with and without provenance as a fixed effect are shown in the box and whisker plots, and provenances are ranked in ascending order of mean NMDS score. Provenance codes follow Table 3.1.

3.4. Discussion

Does host-tree provenance influence gall-type abundance?

The abundance of twenty gall-types was assessed in each of two years, and provenance explained a substantial amount of the variation in abundance in 15 of these 40 cases. If only the more abundant cases are considered (i.e. where more than 1000 galls of a particular type were recorded within a season), then provenance explained a substantial amount of the variation in 15 out of 18 cases. These results suggest that host-tree provenance has an important influence on gall-type abundance.

Which aspects of tree phenotype are involved and is there support for the plant-stress, plant-vigour, and synchronisation hypotheses?

Various significant relationships with tree phenotypic traits were apparent, potentially indicating the mechanisms that may underlay the effect of provenance. Spring bud-burst phenology appeared to be particularly influential, correlating with gall abundance in 7 of the 15 cases where provenance had an effect. Tree diameter also appeared to be influential, correlating with abundance in 6 cases. Autumn leaf-fall phenology correlated with abundance in 4 cases. The *Form* trait that was considered to reflect the degree of tree damage did not appear to be influential, correlating with gall-type abundance in only one case as part of an interaction with spring bud-burst phenology.

These significant relationships with various tree phenotypic traits provide mixed support for the hypotheses outlined in section 3.1.2. The positive relationships between tree diameter and the abundance of sexual generation *N. anthracinus* in 2009, sexual *N. numismalis* in 2009, and asexual *N. albipes* in 2009, are consistent with the expectation from the plant-vigour hypothesis that gallwasps should prefer or perform better on trees that are growing most vigorously. However, the negative relationships between tree diameter and the abundance of sexual generation *N. anthracinus* in 2008, and with both sexual and asexual generations of *N. quercusbaccarum* in 2008 are inconsistent with the vigour hypothesis. Furthermore, the observation of both positive and negative relationships for sexual and asexual generation gall-types is inconsistent with the expectation that the vigour hypothesis

should apply in particular to flush-feeding sexual generation gall-types. Perhaps unsurprisingly given their contrasting expectations, support for the plant-stress hypothesis was also mixed. The three cases of negative relationships with tree diameter were consistent with expectation that gallwasps should prefer or perform better on trees that are most stressed, while the three cases of positive relationships were inconsistent with this expectation. Again, the mixed relationships within both sexual and asexual gall-types are inconsistent with the further expectation that the stress hypothesis should apply in particular to senescence-feeding asexual generation gallwasps.

Despite the numerous relationships involving spring bud-burst and autumn leaf-fall phenology, there was poor support for the synchronisation hypothesis and its expectation that gallwasps should prefer or perform better on trees with a particular optimum phenology. Only for the sexual generation of *N. quercusbaccarum* in 2008 was an optimal relationship with autumn leaf-fall phenology suggested, and even then a non-optimal relationship with spring bud-burst phenology was also apparent. The remaining significant relationships were generally constant, with no apparent optimum within the studied range of the phenological trait. Later leaf-fall phenology corresponded with increasing abundance of asexual generation *N. albipes* in 2008, and asexual *N. quercusbaccarum* in both 2008 and 2009. Relationships with bud-burst phenology contrasted between sexual and asexual generations, with earlier bud-burst corresponding with higher abundance of sexual generation *N. anthracinus* in 2008, sexual *N. numismalis* in 2009, and sexual *N. quercusbaccarum* in both 2008 and 2009, but with lower abundance of asexual *N. anthracinus* in 2009 and asexual *N. quercusbaccarum* in both 2008 and 2009.

How are the richness and structure of gall communities also influenced?

Provenance explained a substantial amount of variation in gall-type richness in three of the four survey seasons, and community similarity in one of the four seasons, indicating that provenance can have an important influence on these community parameters. Spring bud-burst phenology again appeared to be the most influential phenotypic trait, having significant relationships with gall-type richness in two of the three cases of a substantial provenance effect, and with community similarity in the

one case. Tree diameter was involved in a significant relationship with gall-type richness in one of the three cases, and with community similarity in the one case. Autumn leaf-fall phenology was involved in a significant relationship with gall-type richness in one of the three cases.

The significant relationships between gall-type richness and the phenological traits again appeared to be linear rather than curved, with no indication of an optimum within the observed phenological range, and therefore did not support the synchronisation hypothesis. The direction of relationships were the same as for gall abundances, with earlier spring bud-burst phenology corresponding with higher richness of sexual generation gall-types but lower richness of asexual generation gall-types, and later leaf-fall phenology corresponding with higher richness of asexual generation gall-types. Although not influential in the seasons when an effect of provenance was detected, a significant positive relationship between sexual generation richness and tree diameter was identified in one season. The direction of this relationship is consistent with the expectation of the plant vigour hypothesis, that flush-feeding sexual generation gallwasps will be most abundant on vigorous trees.

How consistent are these patterns?

Although numerous relationships with tree provenance and phenotype were detected at certain times, these relationships were not particularly consistent between years. The abundance of only 5 gall-types – sexual generation *N. anthracinus* and *N. quercusbaccarum*, and asexual generation *N. albipes, N. anthracinus,* and *N. quercusbaccarum* - varied substantially between provenances in both survey years, and only for 3 of these – sexual generation *N. anthracinus*, and both sexual and asexual *N. quercusbaccarum* - were significant relationships with the same phenotypic trait detected in both years. Similarly, variation in gall gall-type richness and community similarity between provenances was considered to be substantial in spring 2008 and autumn 2009 respectively, but not in spring 2009 and autumn 2008. There are several possible explanations for this lack of consistency: (1) the observed relationships are false positives (i.e. type I errors) and the corresponding null observations represent the true pattern; (2) the relationships are consistent between years but are not always detected due to a lack of statistical power (i.e. type II error);

or (3) that the relationships are temporally variable, applying within certain years but not others. Although it is possible that some of the less well supported relationships were type I errors, many were so strongly supported that this is an unlikely explanation (i.e. evidence ratios indicated that certain approximating models were several orders of magnitude more likely to perform better than null models). The second scenario is perhaps more likely, as although sampling effort was consistent, the number of galls sampled varied dramatically, with 7 of the 20 gall-types showing more than a ten-fold difference in overall abundance between years. Given the well know relationship between sample size and type II error (Crawley 2007), this would lead to disparity in the size of effect that would be considered as significant in different years. Where relationships were only significant in one year, parameter estimates for the same relationships in the alternate year were generally of the same sign, as would be expected under scenario (2). However, there was a notable exception to this where abundance of sexual generation *N. anthracinus* showed a significant positive relationship with tree diameter in spring 2008, but a significant negative relationship in spring 2009. If neither of these results is erroneous, then this represents an extreme case of scenario (3), although it is difficult to conceive the mechanism that would explain such a pattern.

3.4.1. How might Climate Matching affect gallwasp communities?

Recent projections of climate change within the 21st century indicate that much of the UK and Western Europe will experience an increase in temperature and winter precipitation, and a decrease in summer precipitation (Giorgi and Coppola 2009, Jenkins et al. 2009). Consequently, Climate Matching within this region can be expected to involve the introduction of trees from provenances that have experienced warmer temperatures, higher winter rainfall and lower summer rainfall, relative to trees of local provenance. Given the patterns of local adaptation in relation to temperature identified for *Q. petraea* in Chapter 2, such introduced provenances will likely exhibit inherently earlier bud-burst and leaf-fall phenologies. Furthermore, as the introduced provenances are expected to be relatively better adapted to the forthcoming climate than trees of local provenance, they are likely to exhibit relatively higher growth rates and achieve a relatively greater ultimate size (i.e. will

be more vigorous). As Climate Matching is a recently developed strategy, the scale at which it will be implemented is still to be decided, but it is likely to constitute only part of any forest management plan, with introduced provenances being planted along side trees of local provenance (Broadmeadow and Ray 2005, Hubert and Cottrell 2007). In this context, how might the expected differences in phenology and growth between local and introduced provenances affect gallwasp communities?

From the relationships with bud-burst and leaf-fall phenology observed in this study, sexual generation gallwasps generally appear to respond positively to trees with the earliest phenology, and would therefore be predicted to favour trees of introduced provenances. Conversely, the asexual generations appear to respond negatively towards trees with earlier phenology, and would therefore be predicted to favour trees of local provenance. It is less straightforward to make predictions with regards to tree growth as the relationships with tree diameter identified in this study were mixed. Under the plant-stress and plant-vigour hypothesis, a pattern similar to that for phenology might be expected, with the flush-feeding sexual generation gallwasps favouring the more vigorous trees of introduced provenances, and the senescence-feeding asexual generations favouring the more stressed trees of local provenance. However, support for both of these hypotheses was ambiguous, and while variation in growth is considered to be a potentially important influence, the effects of differences in growth between introduced and local tree provenances remain unclear.

If adult female gallwasps are free to disperse between trees of different provenance, then Climate Matching may improve the phenological conditions for sexual generation gallwasps while maintaining typical conditions for asexual generations, potentially resulting in greater overall species abundances and community richness. Alternatively, if dispersal is limited (i.e. towards the centre of large single provenance stands, and / or where provenances are spatial isolated from one another), then gallwasps colonising the introduced provenances may be subject to improved phenological conditions for sexual generations but poorer phenological conditions for asexual generations, in which case their persistence may depend on whether the asexual generations can tolerate these less favourable conditions. Current recommendations for the application of Climate Matching advocate mixed planting

of matched and local provenances (Broadmeadow and Ray 2005, Hubert and Cottrell 2007), and if practiced in this way, it is considered unlikely to have a severely negative impact on gallwasp communities.

Further to any effects on gallwasp abundance and diversity, there is the possibility that differences in phenotype between local and introduced provenances may influence the genetic structure of gallwasp populations. Differentiation within species of insect herbivores between patchy resources has been demonstrated at various scales, including between host-plant species (Via 1991), between neighbouring patches of a single host species (McCauley and Eanes 1987), and between individual host trees (reviewed in Mopper 2005). Gallwasps possess several traits expected to favour fine-scale differentiation, including sedentary feeding mode, and a partially parthenogenetic lifecycle (Mopper 2005), and two recent studies have found evidence for local adaptation of gallwasps to individual trees (Egan and Ott 2007, Tack and Roslin 2010). Climate Matching might therefore be expected to increase the genetic structuring of gallwasp populations, with the formation of differentiated demes on local and introduced provenances. As migration between demes can limit or prevent differentiation (Tack and Roslin 2010), the effect of Climate Matching on population structure may depend on the spatial arrangement of the trees and might be greatest where introduced provenances are spatially isolated from trees of local provenances, i.e. where they are planted in monocultures, rather than being intermixed

3.4.2. Wider relevance

This study represents the first exploration of how Climate Matching of a forest tree species may influence an associated ecological community – a potentially important consideration if the strategy, or any other that significantly alters the genetic, physical, or spatial structure of tree populations, is to be widely implemented. The results presented here, and in the wider literature, lead to the following inferences: (i) host-tree provenance can affect the abundance, richness, and community structure of gallwasps; (ii) variation in phenology and growth traits are of importance, particularly the timing of spring bud-burst; (iii) Climate Matching may differentially

effect sexual and asexual gallwasp generations but the overall impact on abundance and diversity is unlikely to be severely negative; (iv) phenotypic differences between trees of introduced and matched provenances may lead to increased genetic structuring of gallwasp populations. What are the wider implications and relevance of these results?

Insect herbivores are often embedded within complex ecological networks, serving as hosts to a diversity of parasitoids, predators, and pathogens. Bottom-up influences may pass through these networks, with the potential for both direct and indirect effects on abundance and diversity of species at higher trophic levels (Price et al. 1980, Dickson and Whitham 1996, Ode 2006, Bukovinszky et al. 2008, Johnson 2008, Bukovinszky et al. 2009, Jones et al. 2011b). Oak galls are host to a variety of inquiline gallwasps and hymenopteran parasitoids that attack the gall inhabitants, and were selected for this study in part because they represent a convenient system for investigating multi-trophic interactions (Schönrogge and Crawley 2000, Stone et al. 2002, and see Chapter 4). Although the effects of Climate Matching on gallwasp communities are predicted to be subtle (i.e. not severely negative), they could potentially be more severe at higher trophic levels, especially if ecological cues or requirements differ. For example, gallwasp species in this study occurred across trees with a range of phenologies, and were often most abundant on trees that differed substantially from those which they would usually encounter (i.e. the trees of local provenance). If inquilines and parasitoids have more specific phenological niches than gallwasps, they may be unable to target the galls on introduced provenances, and hence would be more negatively influenced by Climate Matching. Data that will allow the investigation of such multi-trophic effects of tree provenance on the oak gall community have been collected (see Chapter 4), and will be analysed in the near future.

The broad phenological tolerance demonstrated here for gallwasps is in sharp contrast to patterns observed for particular free-feeders whose fitness is strongly linked to phenological synchrony with host-plants (van Asch et al. 2007, van Asch and Visser 2007). This is possibly due to the particularly intricate relationship between gall-forming herbivores and their host plants, with gallwasp larvae having

the ability regulate their resources and microclimate (Schönrogge et al. 2000a, Stone and Schönrogge 2003, Harper et al. 2004), which may increase their tolerance of host-plant variation. If so, then the extrapolation of results from gall communities could be misleading. Extension of the provenance trial survey approach to include additional herbivorous guilds would allow for more thorough evaluation of the community level consequences of Climate Matching, and would also further understanding of how such communities may vary in their resilience to environmental change.

.

Chapter 4 – Assessing the taxonomy of an oak gall community with DNA barcodes

4.1. Introduction

Interaction networks involving insect parasitoids and their hosts are popular systems for investigating the ecological and evolutionary forces that structure complex communities. In recent years their use has led to fresh insight into the structuring properties of indirect interactions (Morris et al. 2004, Tack et al. 2011), the community level consequences of biological invasions (Schönrogge and Crawley 2000, Henneman and Memmott 2001) and the hidden functional effects of habitat modification (Tylianakis et al. 2007). Species level taxa are usually the fundamental unit of these networks, and their quality is thus dependent on accurate species level identification. Unfortunately, obtaining species level accuracy may be a far from trivial task as insect parasitoids suffer particularly from the taxonomic impediment in terms of both the quantity and quality of existing species descriptions. As few as 1% of global parasitoid species have so far been described (Godfray 1994), and the inaccuracy of parasitoid species taxa based on morphology have been estimated at approximately 25% (Smith et al. 2008, Smith et al. 2011). If unresolved, taxonomic error could be a major hindrance to the study of host-parasitoid networks.

The vast majority of recognised species level taxa were described on the basis of differences in morphological characters (henceforth termed 'morpho-species'), but there is a growing appreciation that such taxa may be discordant with modern species concepts that view 'existence as a separately evolving meta-population lineage' as the principal property of species (De Queiroz 2005, 2007). Integrated taxonomic approaches, principally involving molecular tools, continue to reveal examples of species level taxonomic error including instances of over-splitting, in which variants of a single species are classified as two or more, and under-splitting (or 'lumping'), in which two or more distinct species are classified together. Over-splitting may result from phenotypic plasticity where intraspecific morphological variation, potentially relating to differences between generations or between individuals from

different hosts, is incorrectly interpreted as being characteristic of distinct species (Stone et al. 2008, Ács et al. 2010, Nicholls et al. 2010). Conversely, lumping may result from a lack of distinguishing morphological characters between distinct species, or the difficulty in confidently defining such characters, resulting in complexes of morphologically cryptic species (see Bickford et al. 2007, and Pfenninger and Schwenk 2007 for reviews).

Taxonomic error may pose various problems for the interpretation of host-parasitoid networks. Analysis and comparison of networks often involves the estimation of one or more metrics that summarise diversity (e.g. the number and proportions of taxa at various trophic levels, Martinez and Lawton 1995), and the distribution of links between taxa (e.g. the proportion of realised links (known as connectance), the mean number of links per taxon (linkage-density), and the mean number of taxa linked to each host taxon (vulnerability) or parasitoid taxon (generality). Many of these metrics are available in both qualitative and quantitative forms, although the quantitative versions that account for interaction frequencies are generally preferred as being less sensitive to sampling effort, and capable of demonstrating structural differences that would otherwise be missed (Bersier et al. 2002, Tylianakis et al. 2007). Taxonomic error within a network would obviously bias the estimation of these metrics, with the direction and extent of such bias being dependent on the type of error (i.e. splitting or lumping), and on how the trophic links of 'true' species differs from those of morpho-species. If for example a single true parasitoid species is over-split into distinct taxa that differ from one another in their host use, then estimates of the number of taxa and the proportion or parasitoid taxa would be positively biased, while generality would be negatively biased. Alternatively, if two or more host-specialised but morphologically cryptic species are lumped together (e.g. Smith et al. 2006), then the opposite bias would apply. Instances where cryptic or over-split species are trophically redundant (i.e. not ecologically different from their morpho-species) may have less of an effect on network properties, but could still impede interpretation of important factors such as population sizes and migration rates. Further issues arise when species or networks are compared between sites, as erroneous classifications may result in miss-interpretation of species turn-over and the ecological role of particular taxa.

Molecular techniques have great potential for identifying separately evolving lineages (i.e. true species) and are increasingly being used to augment morphological-based taxonomy (Sites and Marshall 2003, Tautz et al. 2003, Vogler and Monaghan 2007). In particular, DNA barcoding (Hebert et al. 2003) has been highly influential in the identification and resolution of species level taxonomic error in a range of taxa including insect parasitoids (Smith et al. 2006, Smith et al. 2007, Smith et al. 2008, Kaartinen et al. 2010). DNA barcodes are short sequences from a standardised region of DNA, such as the approximately 650 base pair (bp) Folmer region of the mitochondrial cytochrome c oxidase subunit I gene commonly used for Metazoa. Based on the assumption that species level taxa are monophyletic at the barcode locus and that variation within species is less than variation between species (Hebert et al. 2003), the primary application of DNA barcoding is as a means of assigning query specimens to existing taxa if they differ from voucher sequences by less than a specified threshold (Hebert et al. 2003, Ratnasingham and Hebert 2007). However, under these same assumptions, DNA barcodes can also offer a basis for assessing the accuracy of established morpho-species boundaries. If a sample of barcode sequences from multiple species is grouped into molecular operational taxonomic units (MOTUs) across a range of sequence similarity thresholds (Blaxter et al. 2005), then the assumption that variation within species is less than and discrete from variation between species will be characterised by a barcoding gap where the number of MOTUs defined from the data is constant across a range of threshold values (Meyer and Paulay 2005, Ács et al. 2010, see Figure 4.1). If such a gap is apparent for a particular dataset then it is likely that the MOTUs defined at thresholds within the gap represent meaningful independent lineages, and these MOTUs can be compared with morpho-species classifications to identify potential taxonomic error. MOTUs that contain all sequences from two or more distinct morpho-species are indicative of over-splitting, whereas the presence of a single morpho-species in multiple MOTUs can be indicative of lumping (Ács et al. 2010).

While MOTUs based on barcodes are a useful unit for developing and assessing taxonomic hypotheses, and 'flagging' potential instances of taxonomic error, there are important limitations to their use. Species monophyly of the barcode locus arises through the loss of shared ancestral polymorphism between lineages through genetic

drift, in a process known as lineage sorting (Funk and Omland 2003). In young species with moderate effective population size, complete lineage sorting of any particular locus may take many thousands of generations, during which time the assumption of species monophyly will be violated and inferences based on barcodes are likely to be discordant with 'true' patterns of speciation (Hudson and Coyne 2002, Meyer and Paulay 2005, Hickerson et al. 2006, Wiemers and Fiedler 2007). Even where lineage sorting is complete, variable levels of intraspecific diversity relating to demographic history may preclude any clear barcoding gap and make it difficult to determine species boundaries (Meyer and Paulay 2005, Bazin et al. 2006, Lohse 2009, Lukhtanov et al. 2009). Furthermore, several studies have revealed discordance between relationships based on barcodes and those supported by nuclear sequence markers, attributable to introgression (Rokas et al. 2003b, Hurst and Jiggins 2005), that could also lead to error if identification were based solely on single locus barcode data. Consequently, barcode MOTUs are considered to represent a taxonomic 'stepping stone' whose link with 'true species' cannot confidently be made without congruent support from additional taxonomic characters (Vieites et al. 2009, Padial et al. 2010, Goldstein and DeSalle 2011).

Despite their widely advocated and demonstrated potential, DNA barcoding and alternative molecular markers have only rarely been applied to assess the taxonomy of host-parasitoid networks. In a pioneering study, van Veen et al. (2003) used sequence data from a region of ribosomal DNA to confirm the taxonomic distinctiveness of four parasitoid species within their aphid-parasitoid study system, adding confidence to earlier interpretation of the potential for apparent competition (Müller et al. 1999). More recently, Kaartinen et al (2010) barcoded a sub-sample of adult parasitoids and inquilines reared from gall-inducing and leaf-mining hosts on *Quercus robur* over two years of study in southern Finland, and Smith et al (2011) barcoded a sub-sample of primary and secondary parasitoids reared from caterpillars of the spruce budworm *Choristoneura fumiferana* and related Lepidoptera from 10 years of study in eastern Canada. Both of these studies revealed cryptic species that were supported by additional molecular markers within 11% (Kaartinen et al. 2010) and 24% (Smith et al. 2011) of established parasitoid morpho-species, having a small but appreciable effect on the interpretation of network properties. In the present

study, a further contribution to this small body of literature is made by presenting the first 'fully' barcoded web (in which every adult individual has been barcoded) for a host-parasitoid community where ecological study continues to be hindered by taxonomic uncertainty – the hymenopteran inquilines and parasitoids reared from gall-inducing hosts on *Quercus petraea.*

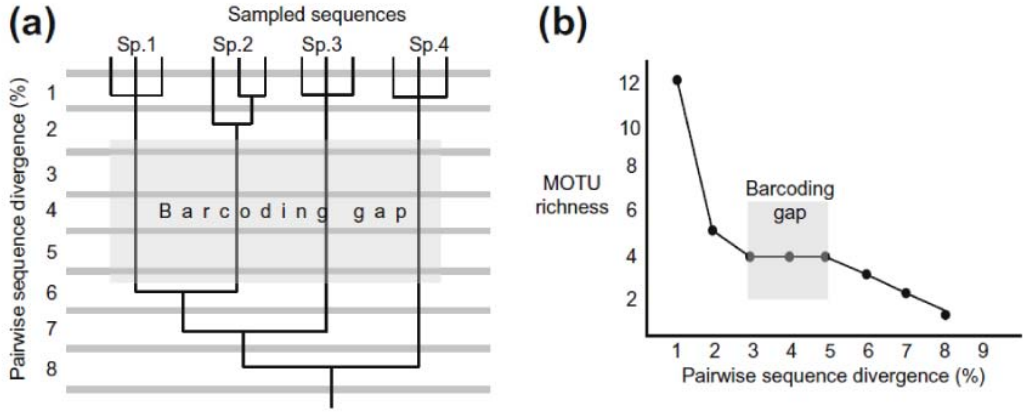


Figure 4.1. An idealised illustration of the barcoding gap, reproduced with permission from Ács et al. (2010). (a) A barcode gene-tree where each of four species is monophyletic and the variation between species considerably exceeds variation within species. (b) The relationship between the number of MOTUs and the defining percentage sequence divergence for the same sample, where the barcoding gap is characterised by a plateau of MOTU richness between 3 and 5% divergence.

4.1.1. Study system

This study focuses on the various inhabitants of cynipid galls on Sessile oak (*Quercus petraea*), an abundant and economically important forest tree that occurs naturally throughout the Western Palaearctic. These galls are induced by species belonging to the Cynipini tribe (Hymenoptera: Cynipidae), whose phytophagous larvae develop and pupate within specialised chambers inside the gall. In the Western Palaearctic, the gall-inducers are typically bivoltine with a sexual generation that develops during spring and an asexual generation that develops in late summer and autumn. Galls occur on various plant organs including buds, leaves, catkins, stems and roots, and may be single or multi-chambered (Stone et al. 2002, Csóka et al. 2005). The morphology of Cynipini galls is often complex, representing an extended phenotype of the gallwasp larva (Stone and Cook 1998, Schönrogge et al. 2000a, Bailey et al. 2009). Gall morphology and its location on the tree are generally diagnostic of a particular generation of a single species, and keys to Western

Palaearctic species based on gall morphology are available (Buhr 1965, Ambrus 1974, Redfern and Shirley 2002). DNA based studies have largely complemented the established morphological taxonomy, facilitating synonymization where alternate generations had been described as separate species (Rokas et al. 2003b, Stone et al. 2008), and confirming the species level status of moderately distinct gall morpho-types (Challis et al. 2007). Notable exceptions include the paraphyly of mitochondrial genes in several species when considered against morphological characters, attributable to hybridization and subsequent introgression within glacial refugia (Rokas et al. 2003b), and the case of *Andricus burgundus*, a small aggregated sexual generation gall on the catkins of *Quercus cerris,* that through DNA based methods was revealed to be a complex of the sexual generations of at least 6 other species (Stone et al. 2008). Based on yet unpublished barcode data from across the geographic range of many Western Palaearctic species, further taxonomic error within Cynipini from Western Europe is considered unlikely (G. Stone, personal communication).

Cynipini galls are frequently colonised by inquilines belonging to the closely related Synergini tribe (Hymenoptera: Cynipidae), whose larvae are able to modify the tissue of existing galls but are unable to induce their own. These inquiline larvae are completely phytophagous but some species cause the death of the gall-inducers by entering their chambers and smothering young larvae. Inquiline species may be univoltine or bivoltine, and multiple individuals and species may coexist within individual galls, even those that initially contained a single gall inducing larva (Csóka et al. 2005). Inquiline taxonomy is problematic and a recent study of a comprehensive sample of Western Palaearctic species using multiple molecular markers (including DNA barcodes) revealed severe flaws in the established morphological taxonomy (Ács et al. 2010). At the species level, several morphological species shared very similar barcode sequences and were considered to have been over-split, while others were revealed to contain cryptic species. At a higher level, two long established sections within the widespread and species rich genus *Synergus* were found not to represent natural groupings, and monophyly of the genus *Saphonecrus* was not supported. A comprehensive revision of the tribe is clearly required, but in the interim, the MOTUs established by Ács et al (2010)

provide useful proximal species level taxa for further study. This was illustrated in the barcoded network study of Kaartinen et al (2010), who found poor concordance between morphological and molecular based inquiline taxa, but could relate several MOTUs to those identified by Ács et al (2010).

Cynipid galls are also host to a diversity of parasitoids that target the gall-inducer, inquilines, or other parasitoids present within the gall. These include members of the hymenopteran families Eulophidae, Eupelmidae, Eurytomidae, Ormyridae, Pteromalidae, and Torymidae. They are generally solitary ectoparasitoids of larvae or pupae, although species of both solitary (e.g. *Sycophila biguttata* and *Pediobius lysis*) and gregarious (e.g. *Baryscapus berhidanus*) endoparasitoids are known (Schönrogge et al. 1995, Csóka et al. 2005). In the western Palaearctic, Cynipini gall parasitoids have been studied in some detail (Askew 1961c, Askew 1961a, Askew 1961b, Nieves-Aldrey and Askew 1988, Schönrogge et al. 1995, Stone et al. 1995, Schönrogge et al. 1996, Schönrogge and Crawley 2000), and their taxonomy is accessible through a comprehensive morphological key (Askew and Thúroczy, unpublished). While DNA based studies have generally supported the monophyly of established parasitoid taxa, they have in several instances revealed the presence of morphologically cryptic species (Kaartinen et al. 2010, Nicholls et al. 2010, see Chapter 5).

Western Palaearctic Cynipini gall communities are a popular model system for community ecology, and have recently been the focus of studies of biological invasions (Schönrogge et al. 1995, Schönrogge and Crawley 2000), comparative phylogeography (Hayward and Stone 2006), Stone et al. In prep.), local adaptation (Tack and Roslin 2010), habitat fragmentation (Kaartinen and Roslin 2011), and community genetics (Tack et al. 2010, see Chapter 3). The sessile nature of galls and their often conspicuous morphology means it is usually straightforward to establish densities in the field, and if collected at an appropriately advanced phase of development then many gall inhabitants suffer from low rearing mortality relative to other herbivorous guilds such as free feeders. The morphological taxonomy of gall inhabitants is also relatively well developed (although see earlier discussion of inquilines), with taxonomic keys that are accessible to non-specialists. Furthermore,

the communities are relatively closed in that gall formers use a limited range of oak host plants, and associated inquilines and parasitoids occur almost exclusively within Cynipini galls (Stone et al. 2002, Bailey et al. 2009). Therefore, populations at the higher trophic levels can be studied by sampling only the oaks present at a site, without the risk of bias due to unsampled hosts. To achieve this for communities associated with other host guilds, such as leaf-miners, it may be necessary to sample a much wider range of plant taxa (van Veen et al. 2006). There are however important limitations to the use of Cynipini gall communities for studying community ecology. The most convenient means of study is to collect and rear mature galls, and to interpret the abundance and identity of emerging adult insects, but as galls are often microcosms of trophic activity with dissection studies revealing frequent secondary parasitism and cannibalism (Askew 1961a, Schönrogge et al. 1995), the identity of the host of emerging parasitoids cannot be confidently established. Networks based on emerging adults are therefore association rather than trophic networks, and network properties such as connectivity (a measure of the proportion of realised links) and the potential for indirect interactions should be treated with caution as they are likely to be under and over-estimated respectively (Schonrogge and Crawley 2000). This lack of trophic resolution also limits the ways in which gall based networks can be compared or combined with more fully resolved networks, such as those based on leaf-mining or free-feeding insects.

A further complication involves the role of the Synergini inquilines within the community. Strictly, the term inquiline is reserved for organisms that do not have a deleterious effect on the host species, and while this may be true for some Synergini, such as members of the *Synergus* genus within the asexual generation galls of *Andricus quercuscalicis* (Schönrogge et al. 1995), it is clearly violated in other instances where inquilines can be a major cause of gall-inducer mortality (Csóka et al. 2005). In previous studies of gall communities based on dissections, a clear distinction was made between the role of inquilines and the role of parasitoids (Askew 1961a, Schönrogge et al. 1995). In studies based upon rearings however, inquilines and parasitoids were considered as equivalents by Schönrogge and Crawley (2000) in the estimation of indirect interactions between gall species, and by Kaartinen et al (2010 & 2011) in the estimation of a range of network properties. In

this present study, inquilines and parasitoids were not considered as equivalents, in reflection of their fundamental life-history differences, and also the difference in taxonomic resolution (i.e. meaningful morphological identifications can be obtained for the parasitoids but not for the inquilines).

4.1.2. Objectives

The primary objective of this study is to investigate the taxonomy within a community of Cynipini gallwasps and associated inquilines and parasitoids at the study site, and to accurately establish specimen identities that could be used in further ecological analysis. In approaching this objective, consideration is also given to how reliable ecological data for individual species may be compiled, and how not revealing taxonomic error may bias the estimation of network properties. Specifically I ask: (i) within my sample, do gall inquilines and parasitoids show signs of a barcoding gap that would support the use of MOTUs for detecting meaningful taxonomic units; (ii) can the parasitoid and inquiline MOTUs identified here be linked with those from other studies to add confidence to their status and compile reliable ecological data; (iii) in a guild where morphological taxonomy is well established (i.e. gall parasitoids), is there discordance between taxa based on morphology and MOTUs; (iv) if so how does this influence the estimation of various network properties?

4.2. Methods

4.2.1. Gall collection and rearing

During 2009, quantitative surveys of gall abundance were conducted at a provenance trial of *Quercus petraea* in the forest of Petite Charnie, Sarthe, France (see Chapter 3). Sexual generation galls were surveyed in spring (May-June) and asexual generation galls in autumn (August-October). During these surveys, galls were collected from within 32 ‘parcelles’ (each parcelle being a block of 24 trees of a single provenance, although provenance is not considered further here). As the galls of alternate generations of individual gall-inducing species often differ dramatically

in their morphology, abundance, and associated parasitoid communities (Bailey et al. 2009), they are treated here as distinct 'gall-types'. The number of collected galls of each type approximated its relative abundance in the gall surveys. Most galls were collected directly from the trees, transferred to small ventilated containers, and reared in an outside insectary at the Centre for Ecology and Hydrology in Wallingford, Oxfordshire, UK. However, for asexual generation galls of members of the genus *Neuroterus* that are prone to desiccation, galls were collected from the ground at the centre of parcelles in mid-October 2009, and were refrigerated at 4 ˚C in sealed bags of moist sphagnum moss until February 2010 when they were transferred to ventilated containers in the insectary. Containers were checked regularly until November 2010, and emerging adult insects were preserved in 99% ethanol. Adult parasitoids were identified following a morphological key to Western Palaearctic oak gall parasitoid species (Askew, R. and Thúroczy, C. unpublished).

4.2.2. DNA extraction and sequencing

DNA was extracted for all adult parasitoids and inquilines from a single leg (abdomens were used for some small male specimens) following a chelex and proteinase K protocol. For all parasitoid individuals a fragment of the mitochondrial *coxI* gene was amplified using the forward primer COI_pf1; 5' AGG RGY YCC WGA TAT AGC WTT YCC 3' (designed by J. Nicholls), and the reverse primer COI_2437d: 5' –GCT ART CAT CTA AAW AYT TTA ATW CCW-3' (modified from (Simon et al. 1994)'s C1-J-2441 primer by J. Nicholls). This fragment, henceforth referred to as COI, overlaps with the standard 'Folmer' DNA-barcode region by approximately 370 bp. Although this is less than the 500 bp required for formal barcode status (Ratnasingham and Hebert 2007), the COI fragment was used in preference as a poly-T series within the Folmer region has been found to reduce sequence quality in several families of chalcid parasitoids (see Chapter 5). For all inquiline individuals, a fragment of the Folmer region was amplified using the forward primer LCO1490: 5' GGT CAA CAA ATC ATA AAG ATA TTG G 3' (Folmer et al. 1994), and the reverse primer HCOd: 5' TAW ACY TCD GGR TGI CCA AAA AAY CA-3' (modified from Folmer *et al's* (1994) HCO2198 by J. Nicholls). Each 20 µl PCR mix consisted of 1 µl of DNA template, 0.1 µl of Taq

polymerase (5 U/µl, Bioline), 0.25 µl of each primer (20 µM), 1 µl of dNTP's (25 mM each), 0.8 µl of $MgCl_2$ (50 mM), 2 µl of 10 x Bioline PCR buffer, 2 µl of bovine serum albumin (10 mg/mL), and 12.6 µl of milipure H_2O. Cycling conditions were 94 °C for 2 minutes, followed by 4 cycles of 94 °C for 30 seconds, 45 °C for 1 minute, and 72 °C for 1 minute, then 34 cycles of 94 °C for 30 seconds, 50 °C for 1 minute, 72 °C for 1 minute, with a final step of 72 °C for 5 minutes.

Excess dNTPs were removed from PCR products by adding a solution of shrimp alkaline phosphatase (SAP) and exonuclease, incubating at 37°C on a PCR block for 40 minutes, then heating to 94°C for 15 minutes. Clean PCR product for each individual was sequenced in the forward direction using ABI BigDye chemistry (Perkin Elmer Biosystems Waltham MA) on ABI 3700 and 3730 sequencers at the GenePool, Edinburgh. Chromatograms were checked by eye, and trimmed to a standardised length of 610 bp for COI fragments and 620 bp for Folmer fragments using Sequencher version 4.9 (Gene Codes Corporation 2009). All sequences were checked for an open reading frame, a lack of which would indicate base calling error or amplification of a nuclear pseudogene (Rokas et al. 2003a). Alignments of all unique haplotypes were compiled separately for COI and Folmer fragments (i.e. for parasitoids and inquilines).

4.2.3. MOTUs and their discordance with morphological taxa

Alignments of parasitoid and inquiline haplotypes were analysed with jMOTU version 1.0.8 (Jones et al. 2011a), which clusters sequences into MOTUs based on threshold base-pair differences using a combination of BLAST and the Needleman-Wunch exact global alignment algorithm. To assess for a barcoding gap, the number of defined MOTUs was calculated across a threshold range from 0 to 90 base pairs (bp) for each alignment (corresponding to 0 and approximately 15% of the fragment length). The existence of a barcoding gap is characterised by a plateau of MOTU richness across a range of threshold values, bounded at either end by a relatively steep decline (see Figure 4.1). Although there is no formal means of defining a barcode gap, its 'divisive' and 'inclusive' limits can be estimated from a MOTU richness plot (Ács et al. 2010). Based on the assumption of species monophyly and

greater variation between than within species, it is expected that MOTUs defined at thresholds within the barcoding gap represent meaningful independently evolving lineages, but there is likely to be a gradient from increased risk of over-splitting meaningful taxa at the divisive limit, towards lumping of taxa at the inclusive limit.

To facilitate comparison of taxa between this and other studies, all COI and Folmer fragment sequences published by Kaartinen et al. (2010) and Ács et al. (2010) were obtained from GenBank, trimmed to matching length, and added to the parasitoid or inquiline alignments. Clustering into MOTUs was repeated and MOTU composition examined at the divisive threshold base pair difference (selected from assessment of barcoding gap for sequences generated in this study). Where individual MOTUs contained sequences from this and one or both of the alternate studies, host associations and geographic locations were extracted.

The relationships between barcode haplotypes were visualised in neighbour-joining trees of each parasitoid family, and of the inquilines. Trees were constructed with MEGA version 4 (Tamura et al. 2007) using p-distances. For each parasitoid tree, groupings based on morphology and the inclusive and divisive limits of the barcoding gap were illustrated and assessed for discordance.

4.2.4. Network structure

The influence of discordance between MOTUs and morpho-species on network properties was investigated by constructing a series of networks using the *Bipartite* package version 1.16 (Dormann et al. 2008) in R version 2.13.0 (R Development Core Team 2011). Each emerging adult insect constituted 1 link with its host gall type, and the sexual and asexual generation galls of single gall-inducing species were treated as distinct taxa. Individual bipartite networks were constructed where taxon membership in the higher level was based on parasitoid morphological ID's, and parasitoid MOTUs at thresholds from 0 to 90 bp (hence a total of 1 + 91 = 92 networks). For each network, the following metrics were calculated using the *networklevel* function of *Bipartite*: quantitative generality (G_q, the weighted mean number of host species per parasitoid species, Tylianakis et al. 2007), quantitative vulnerability (V_q, the weighted mean number of parasitoid species per host species,

Tylianakis et al. 2007), quantitative linkage density ($L.D_q$, the mean number of links per species assuming equal biomass of all species and weighted by relative inflows and outflows, Bersier et al. 2002), and weighted connectance (C_w, a weighted proportion of realised links, Tylianakis et al. 2007). The calculation of these metrics for bipartite host-parasitoid networks is described in the supplementary methods of Tylianakis et al. (2007).

4.3. Results

4.3.1. Specimens and sequences

A total of 16,644 galls of 17 types were collected for rearing. From these, 2556 adult insects emerged, of which 98 were gall-inducers, 1053 were inquilines, and 1405 were parasitoids. The parasitoids were identified into 23 morpho-species. Sequence data were generated for all parasitoids and inquilines with the exception of 19 and 13 individuals respectively. In these cases, either the PCR failed to amplify a fragment or the sequence trace file was indecipherable, despite repeated attempts on fresh extractions. This is presumably due to degradation of specimen DNA, or mutations within the priming sites. These individuals (1.3% of the total) were excluded from all further analysis. Details of gall-inducers are shown in Table 4.1, and inquilines and parasitoids in Table 4.2.

Table 4.1. List of the 17 gall-types encountered in this study, with details of the number of galls reared, and the numbers of emerging adult gall-inducers (GI), inquilines (Inq), and parasitoids (Para). Codes shown here are used to label Figures 4.3, 4.4, and 4.6.

Species	Generation	Code	No. Reared	No. GI	No. Inq	No. Para
***Andricus callidoma* (Hartig, 1841)**	Asexual	AcallAsex	11	0	3	1
***Andricus curvator* (Hartig, 1840)**	Sexual	AcurvSex	43	6	5	7
***Andricus fecundator* (Hartig, 1840)**	Asexual	AfecAsex	338	0	49	10
***Andricus glandulae* (Hartig, 1840)**	Asexual	AglanAsex	161	0	2	5
***Andricus kollari* (Hartig, 1843)**	Asexual	AkollAsex	66	0	138	13
***Andricus solitarius* (Fonscolombe, 1832)**	Asexual	AsolAsex	294	5	3	98
***Biorhiza pallida* (Olivier, 1791).**	Sexual	BpalSex	2	23	0	10
***Cynips divisa* (Hartig, 1840)**	Asexual	CdivAsex	117	8	5	15
***Cynips quercusfolii* (Hartig, 1840)**	Asexual	CqfAsex	124	18	12	28
***Neuroterus albipes* (Schenck, 1863)**	Asexual	NalbAsex	4187	5	7	352
***Neuroterus albipes* (Schenck, 1863)**	Sexual	NalbSex	759	1	6	79
***Neuroterus anthracinus* (Curtis, 1838)**	Asexual	NantAsex	2048	0	34	3
***Neuroterus anthracinus* (Curtis, 1838)**	Sexual	NantSex	1111	7	4	266
***Neuroterus numismalis* (Geoffroy 1785)**	Asexual	NnAsex	1115	9	0	89
***Neuroterus numismalis* (Geoffroy 1785)**	Sexual	NnSex	1492	3	207	123
***Neuroterus quercusbaccarum* (Linnaeus, 1758)**	Asexual	NqbAsex	3554	2	0	2
***Neuroterus quercusbaccarum* (Linnaeus, 1758)**	Sexual	NqbSex	1222	11	565	285

Table 4.2. List of parasitoid morpho-species, and inquiline and parasitoid MOTUs as defined at the divisive thresholds (11 and 13 bp respectively), with details of the numbers of individuals sequenced (No. S) and not sequenced (No. NS, provided for morphological taxa only). Codes shown here are used to label Figures 4.3, 4.4, and 4.6. Where MOTUs were matched with other studies, the recorded locations are indicated by two letter country codes following those provided by the International Organization for Standardization. Records labelled as [1] are from Kaartinen et al. (2010), and [2] are from Ács et al (2010). Table is continued on the following page.

Taxa	Family	Code	No. S	No. NS	Matching records
***Aprostocetus aethiops* (Zetterstedt, 1838)**	Eulophidae	Aaet	18	0	
Aprostocetus aethiops MOTU 1	"	Aaet_1	16		
Aprostocetus aethiops MOTU 2	"	Aaet_2	2		FI[1]
***Aprostocetus cerricola* (Erdös, 1954)**	Eulophidae	Acer	1	0	HU[1]
***Aulogymnus arsames* (Walker, 1838)**	Eulophidae	Aars	24	0	FI[1], HU[1]
***Aulogymnus skianeuros* (Ratzeburg, 1844)**	Eulophidae	Aski	4	0	
***Cirrospilus diallus* (Walker, 1838)**	Eulophidae	Cdia	1	0	FI[1]
***Pediobius lysis* (Walker, 1839)**	Eulophidae	Plys	171	1	
Pediobius lysis MOTU 1	"	Plys_1	25		
Pediobius lysis MOTU 2	"	Plys_2	146		
***Eupelmus annulatus* (Nees, 1834)**	Eupelmidae	Eann	1	0	ES[1]
***Eupelmus urozonus* (Dalman, 1820)**	Eupelmidae	Euro	58	0	
Eupelmus urozonus MOTU 1	"	Euro_1	57		PT[1], ES[1], FI[1], HU[1]
Eupelmus urozonus MOTU 2	"	Euro_2	1		
***Eupelmus splendens* (Giraud, 1872)**	Eupelmidae	Espl	3	0	
***Eurytoma brunniventris* (Ratzeburg, 1852)**	Eurytomidae	Ebru	111	4	
Eurytoma brunniventris MOTU 1	"	Ebru_1	100		
Eurytoma brunniventris MOTU 2	"	Ebru_2	10		
Eurytoma brunniventris MOTU 3	"	Ebru_3	1		
***Sycophila variegata* (Curtis, 1831)**	Eurytomidae	Svar	6	0	
***Ormyrus nitidulus* (Fabricius, 1804)**	Ormyridae	Onit	6	0	ES[1], FR[1]
***Ormyrus pomaceus* (Geoffroy, 1785)**	Ormyridae	Opom	157	4	
Ormyrus pomaceus MOTU 1	"	Opom_1	41		
Ormyrus pomaceus MOTU 2	"	Opom_2	96		
Ormyrus pomaceus MOTU 3	"	Opom_3	18		FI[1], HU[1]
Ormyrus pomaceus MOTU 4	"	Opom_4	1		
Ormyrus pomaceus MOTU 5	"	Opom_5	1		

Table 4.2. Continued

Taxa	Family	Code	No. S	No. NS	Matching records
***Mesopolobus dubius* (Walker, 1834)**	Pteromalidae	Mdub	3	0	GB[1]
***Mesopolobus fasciiventris* (Westwood, 1833)**	Pteromalidae	Mfas	171	4	FI[1], HU[1], ES[1]
***Mesopolobus fuscipes* (Walker, 1834)**	Pteromalidae	Mfus	4	0	GB[1], HU[1]
***Mesopolobus mediterraneus* (Mayr, 1903)**	Pteromalidae	Mmed	4	0	GB[1]
***Mesopolobus tibialis* (Westwood, 1833)**	Pteromalidae	Mtib	346	2	FI[1], HU[1]
***Ormocerus vernalis* (Walker, 1834)**	Pteromalidae	Over	56	3	
Ormocerus vernalis MOTU 1	"	Over_1	3		
Ormocerus vernalis MOTU 2	"	Over_2	53		
***Megastigmus dorsalis* (Fabricius, 1798)**	Torymidae	Mdor	23	0	
Megastigmus dorsalis MOTU 1	"	Mdor_1	6		
Megastigmus dorsalis MOTU 2	"	Mdor_2	17		
***Torymus auratus* (Müller, 1764)**	Torymidae	Taur	38	1	ES[1], HU[1]
***Torymus flavipes* (Walker, 1833)**	Torymidae	Tfla	163	0	
Torymus flavipes MOTU 1	"	Tfla_1	136		FI[1], HU[1]
Torymus flavipes MOTU 2	"	Tfla_2	27		FI[1]
***Torymus geranii* (Walker, 1833)**	Torymidae	Tger	17	0	FI[1], HU[1]
Inquilines	Cynipidae		1040	13	
Inquiline MOTU 1	"	Inq_1	93		FI[1]
Inquiline MOTU 2	"	Inq_2	10		FI[1]
Inquiline MOTU 3	"	Inq_3	9		
Inquiline MOTU 4	"	Inq_4	194		FI[1], ES[2]
Inquiline MOTU 5	"	Inq_5	238		
Inquiline MOTU 6	"	Inq_6	312		FI[1], HU[2]
Inquiline MOTU 7	"	Inq_7	138		
Inquiline MOTU 8	"	Inq_8	42		
Inquiline MOTU 9	"	Inq_9	2		
Inquiline MOTU 10	"	Inq_10	2		HU[2]

4.3.2. The barcoding gap

The relationship between the number of MOTUs and the defining threshold base pair difference for the parasitoid and inquiline alignments is shown in Figure 4.2. A plateau of MOTU richness was apparent in both cases, characteristic of a barcoding gap. From the plotted relationship, the divisive and inclusive limits of the barcoding gap were considered to be 11 and 55 bp for the inquilines (corresponding to 1.8 – 8.9% of the fragment length), and 13 and 55 bp for the parasitoids (corresponding to 2.1 - 9.0% of the fragment length). At the divisive thresholds, 35 parasitoid and 10 inquiline MOTUs were defined, reducing to 26 and 6 by the inclusive threshold (a decrease of 25.7 and 40% respectively).

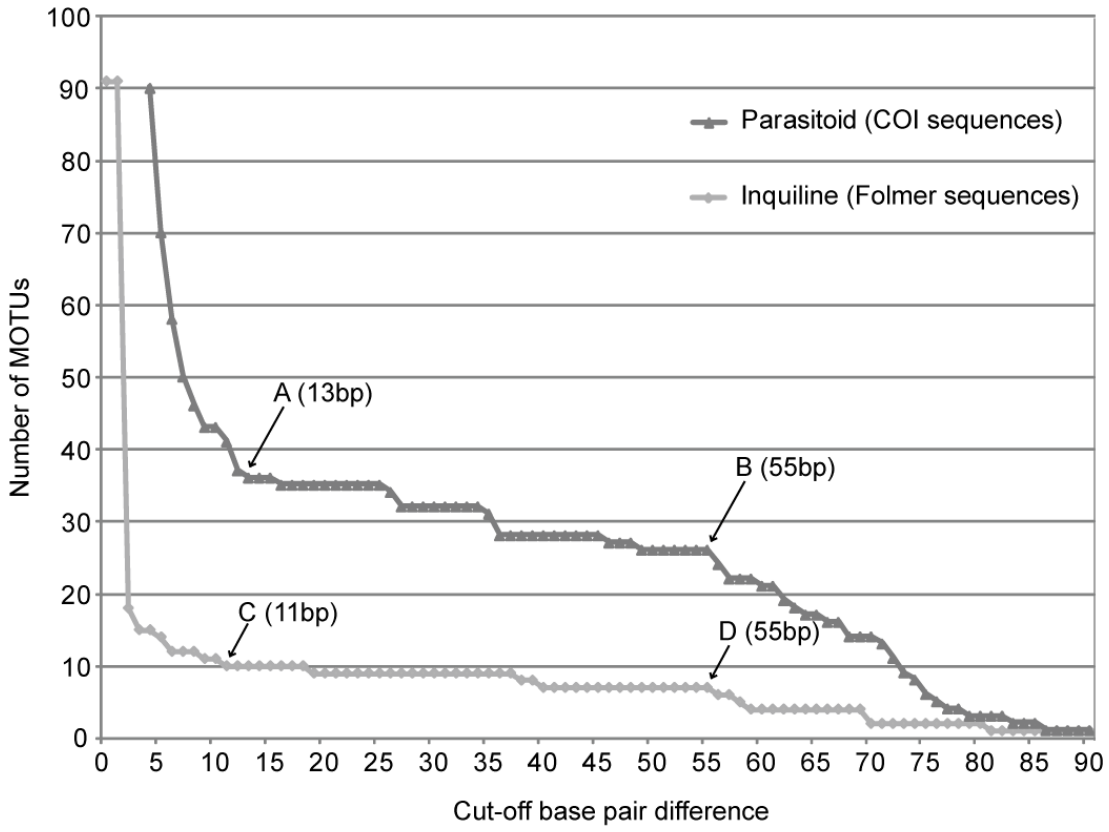


Figure 4.2. Relationship between number of MOTUs and defining threshold base pair difference for parasitoid and inquiline sequences. Parasitoid MOTU richness was 539 at the 0 bp threshold and for ease of visualisation is only shown between 4 and 90 bp. Arrows indicate the selected boundaries of the barcoding gap for parasitoids (A-B) and inquilines (C-D).

4.3.3. MOTUs

Gene trees of COI sequences for each of the six parasitoid families and the inquilines are shown in Figure 4.3. All parasitoid morpho-species were monophyletic except for *Aprostocetus aethiops* that was paraphyletic with respect to *A. cerricola* (fig 4.3a). Of the 35 parasitoid MOTUs defined at the divisive threshold, 15 corresponded exactly with individual morpho-species. There was no indication that any morpho-species had been over-split, but eight of the 23 contained multiple MOTUs, suggesting the presence of morphologically cryptic species. At the inclusive threshold, the morpho-species *A. aethiops* and *A. cerricola* formed a single MOTU, and four morpho-species still contained multiple MOTUs. This grouping together of the two morpho-species is potentially indicative of taxonomic over-splitting, although it did not occur until the relatively large threshold base pair difference of 36 bp (corresponding to 5.9% of fragment length), approximately in the middle of the barcoding gap.

Of the total 45 MOTUs defined at the divisive thresholds, 17 of the parasitoids and 5 of the inquilines were matched with those from other studies (Table 4.2). For the parasitoids, these included 12 of the 15 MOTUs that corresponded exactly with a single morpho-species, and 5 of the 20 MOTUs that did not. Both of the MOTUs from within the morpho-species *Torymus flavipes* could be matched, but at least one MOTU from within the other 7 morpho-species were not.

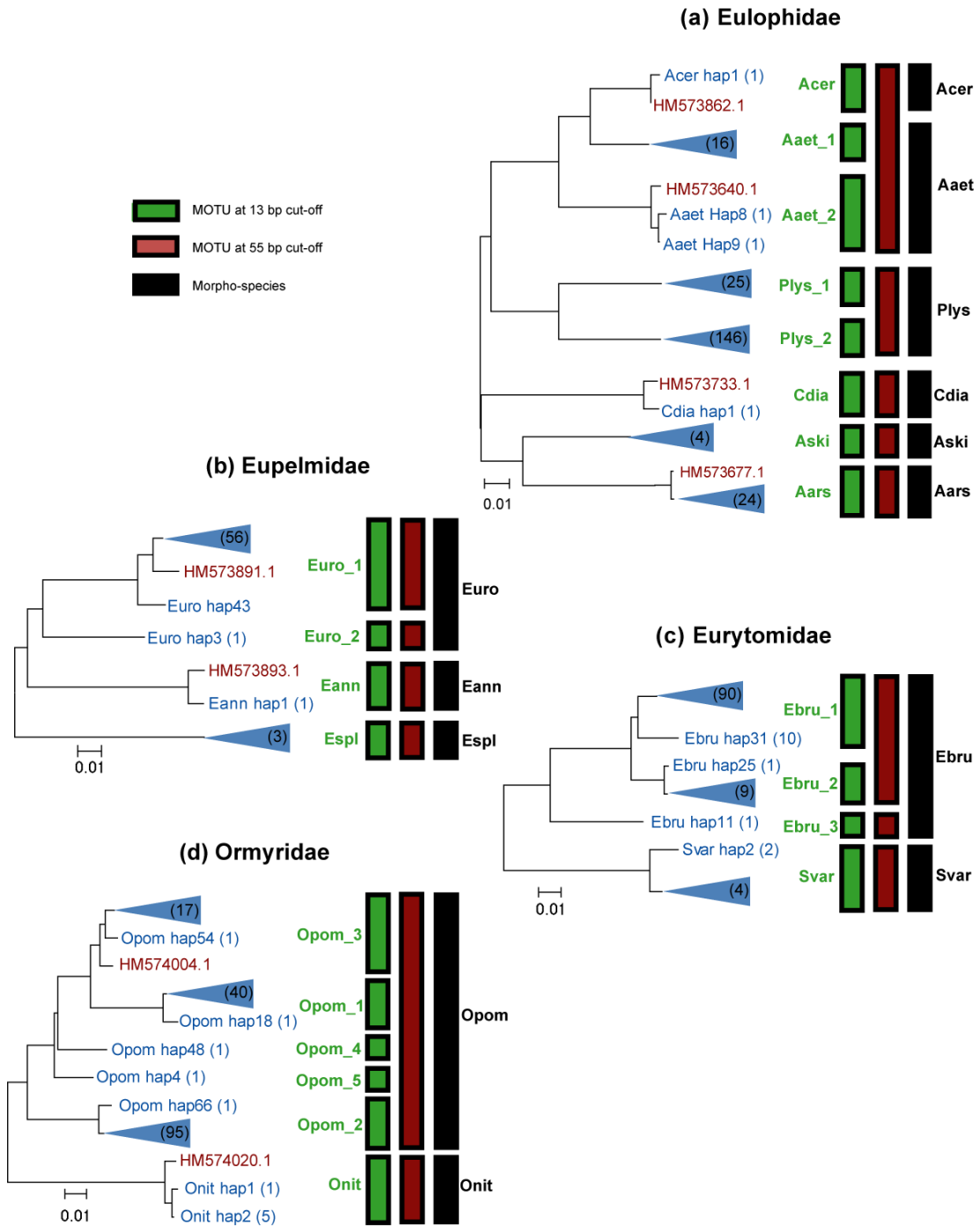


Figure 4.3. Neighbour-joining trees based on p-distances for COI haplotypes of six parasitoid families (a-f), and Folmer haplotypes of Synergini inquilines (g). Taxa labels follow the codes in Tables 4.1 and 4.2. Haplotypes generated in this study are shown in blue, and for ease of visualisation multiple similar haplotypes are in places represented by a blue filled triangle. Numbers within brackets indicate the individuals of a particular haplotype or group of haplotypes. Sequences shown in red or maroon are from the studies of Ács et al. (2010) and Kaartinen et al. (2010) respectively, and are coded by their GenBank accession number. Vertical bars to the right indicate membership of MOTUs at the divisive threshold (13bp for parasitoids, 11 bp for inquilines, shown in green), MOTUs at the inclusive threshold (55bp, in maroon), and morpho-species (in black). Taxon codes in black refer to morpho-species and codes in green refer to MOTUs at the inclusive threshold. Scale bars indicate a p-distance of 0.01. The Figure is continued on the following page.

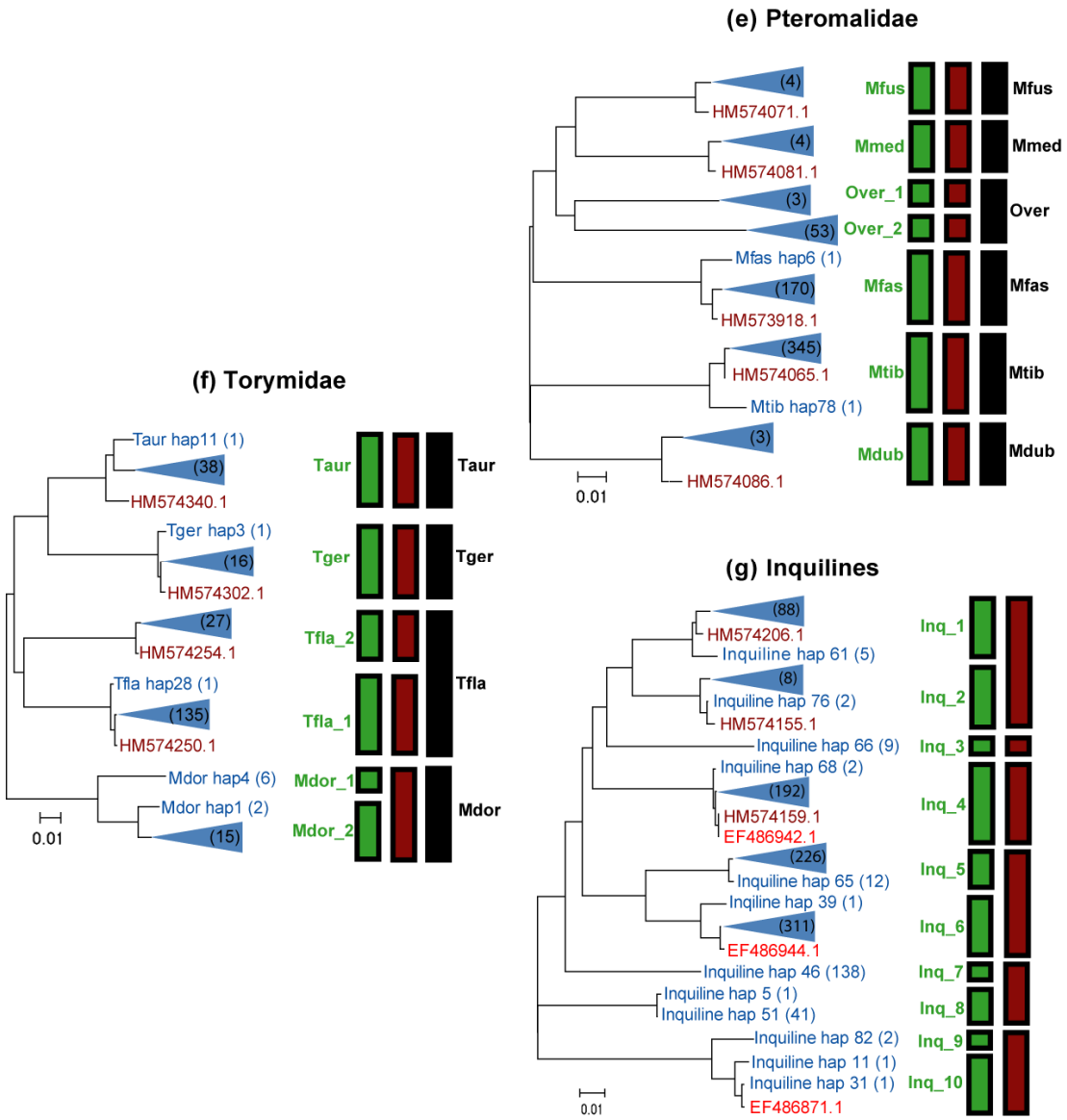


Figure 4.3. Continued

4.3.4. Networks

An illustration of bipartite host-parasitoid and host-inquiline networks based on taxa at the devisive limit of the barcoding gaps is shown in Figure 4.4, and the source host association data for these networks are provided in Appendix 4.1. At this devisive threshold, the values of G_q and C_q for the parasitoid network were lower then if parasitoid identity is based on morpho-species classifications, and the values of V_q, and $L.D_q$ were higher (see Figure 4.5). If considered as a percentage of the morpho-species based value, these differences are slight for G_q (- 3.0%), and moderate for C_w (- 12.8%), $L.D_q$ (+ 13.3%), and V_q (+ 25.2%). Differences decrease as the threshold increases within the barcoding gap, and all are within 5% of the morpho-species based values at the inclusive threshold of 55 bp (see Figure 4.5).

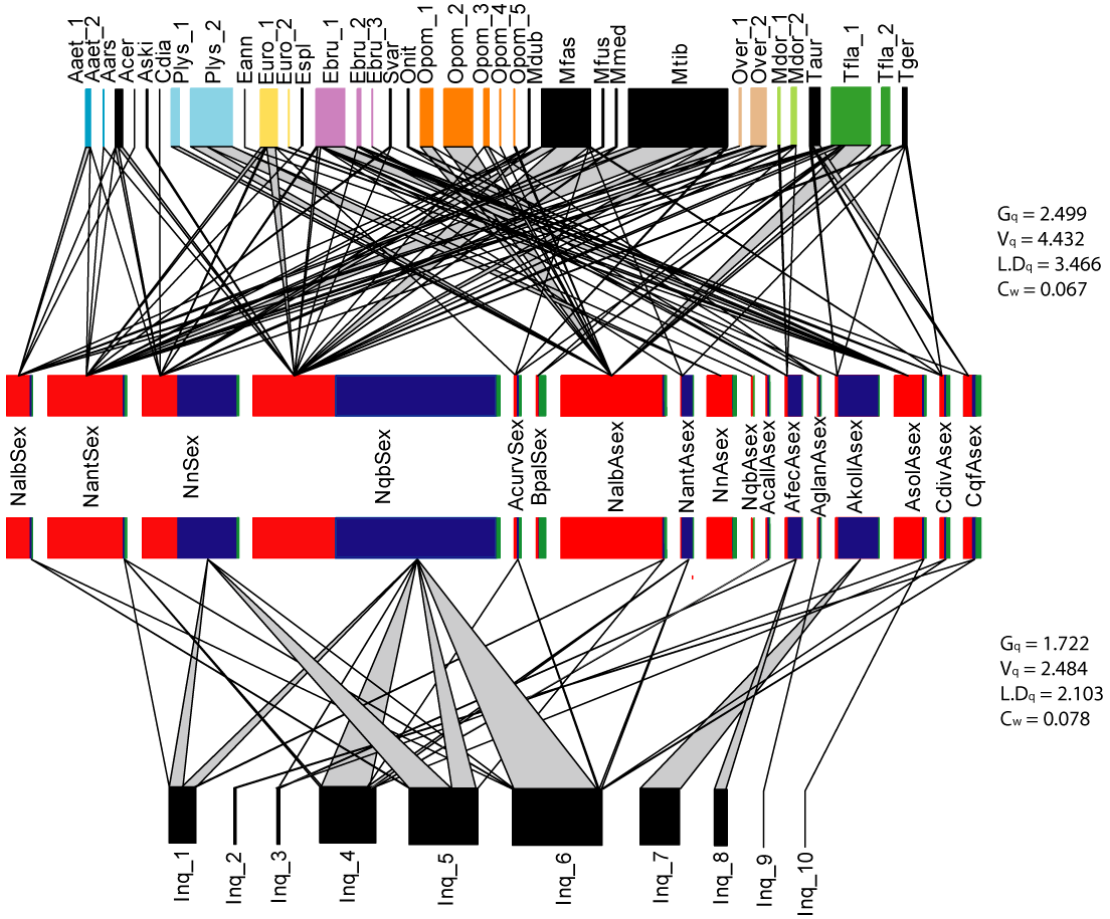


Figure 4.4. Bipartite networks of quantitative host associations for parasitoid (top) and inquiline (bottom) MOTUs at the divisive thresholds. Taxon labels follow the codes in Tables 4.1 and 4.2. Bar-widths for parasitoids and inquilines represent the number of individuals belonging to each taxon, and the basal width of grey triangles represents the number of individuals emerging from a particular host. Parasitoid MOTUs that were discordant with morpho-species are coloured. Bar-widths for the hosts represent the number of emerging adult insects, with the proportion of parasitoids highlighted in red, inquilines in blue, and gall-inducers in green. The values of various metrics are provided for each network (definitions given in text, section 4.2.4).

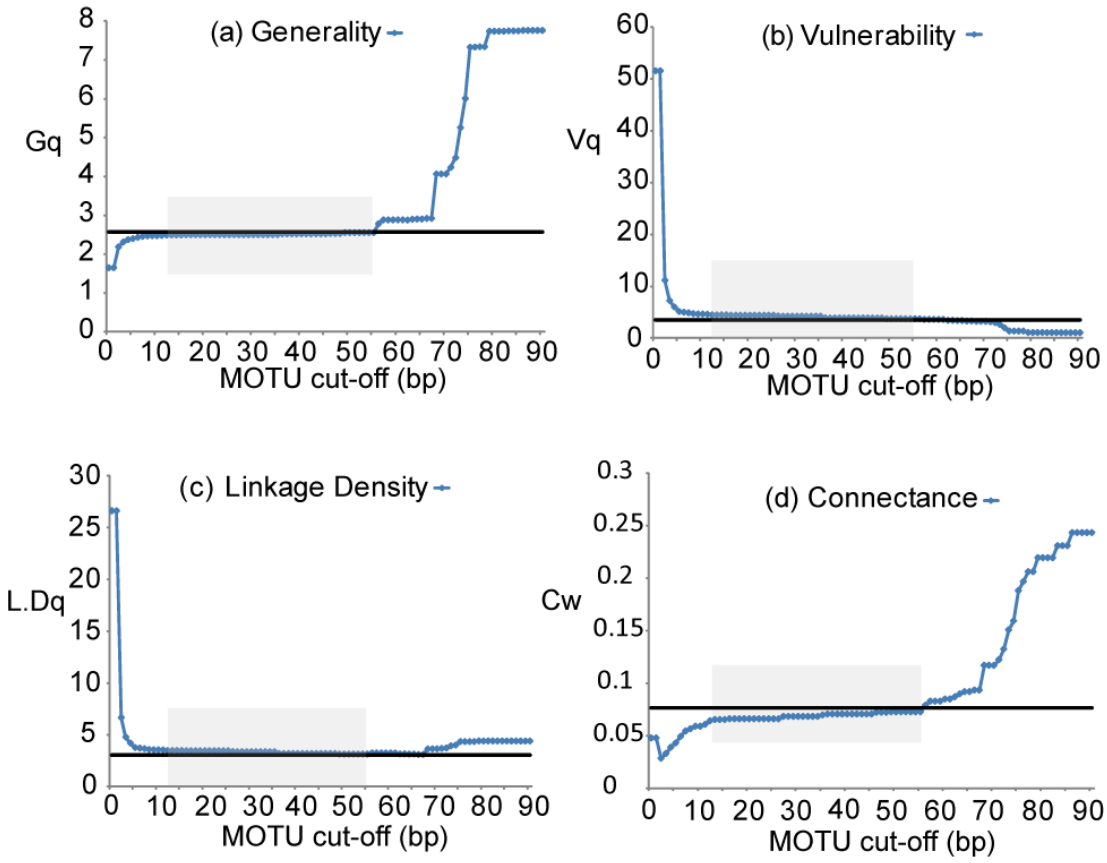


Figure 4.4. Variation in 4 network metrics (a-d) when parasitoid taxon membership is defined across a range of MOTU thresholds. Horizontal black lines indicate the value of metrics when parasitoids were identified from morphology. Grey shaded area indicates the location of the barcoding gap on the x-axis.

4.4. Discussion

4.4.1. The utility of DNA barcodes for studying oak gall communities

The utility of DNA barcoding as a means of both associating query specimens with established taxa and assessing species level taxonomy is dependent on the extent to which the assumptions of species monophyly and large relative distances between species are met for the taxa under consideration. A previous study of Western Palaearctic oak gall inquilines supported these assumptions, identifying a barcoding gap within which MOTUs based on barcodes and a nuclear gene were highly concordant. Similarly, for oak gall parasitoids, a barcoding gap has been observed for

some species within the genus *Cecidostiba* (see Chapter 5), and two studies have reported that divergence at mitochondrial loci was concordant with 'true' species limits, as inferred from additional nuclear markers (Kaartinen et al. 2010, Nicholls et al. 2010). Within the inquilines and parasitoids of oak galls, there have to date been no reported examples of paraphyly at mitochondrial genes in relation to 'trusted' species level classifications, as might arise through introgression or incomplete lineage sorting ('trusted' species are those based on multiple molecular markers, potentially in combination with morphological characters, rather than on morphology alone). In the present study, the assumptions of barcoding were further supported by the observation of a barcoding gap for both inquilines and parasitoids. Given this accumulated support, it is considered that DNA barcoding is a valid and valuable tool for investigating the morphological taxonomy and obtaining accurate identifications of oak gall inquilines and parasitoids.

4.4.2. Taxonomy of the studied community

DNA barcoding supported the taxonomic distinctiveness of 15 out of the 23 parasitoid morpho-species encountered in this study. There was little suggestion of taxonomic over-splitting, as the grouping of multiple morpho-species into a single MOTU only occurred above what was considered to be an unusually high level of intraspecific barcode variation. However, 8 parasitoid morpho-species each contained two or more MOTUs at the proposed divisive limit of the barcoding gap, which was considered to be indicative of taxonomic lumping. For *Torymus flavipes*, both of the MOTUs encountered here could be matched with those identified by Kaartinen et al. (2010), who proposed them to be distinct species based on concordant patterns of divergence for a nuclear marker (i.e. following the terminology of Vieites et al. 2009, they each represent a confirmed candidate species). For the other 7 cases, at least one of the MOTUs could not be matched with published data, and following the terminology of Vieites et al. (2009) I propose that these should presently be considered as unconfirmed candidate species (i.e. deep genealogical lineages of unknown status).

As previously described, the morphological taxonomy of Western Palaearctic Synergini inquilines is highly discordant with patterns of differentiation at mitochondrial and nuclear genes, and the validity of many named species is questionable, particularly within the large genus *Synergus* (Ács et al. 2010). In light of this, it was not considered worthwhile to attempt morphological identification of inquilines, but to instead rely on DNA barcoding for establishing identities. Of the 10 barcode MOTUs identified here, 5 could be matched with those from the studies of Ács et al. (2010), or Kaartinen et al (2010). One of these matches was with *Ceroptres clavicornis,* a species where morphological and molecular data have so far been concordant (Ács et al. 2010), suggesting that its existing classification is valid. The other 4 matches were with MOTUs from within the genus *Synergus*, and as these have been shown to be concordant for both mitochondrial and nuclear markers, but do not correspond to established Linnaean species names, I propose that they be considered as confirmed candidate species (i.e. again following Vieites et al. 2009). The remaining 5 inquiline MOTUs identified here did not match published records, and I propose that these represent unconfirmed candidate species.

In summary, the 2426 parasitoid and inquiline individuals that were barcoded in this study are currently considered to represent 45 taxa, of which 16 are recognised Linnaean species, 6 are confirmed candidate species, and 23 are unconfirmed candidate species. While the Linnaean and confirmed candidate species represent appropriate species level taxa for ecological analysis of this community, further taxonomic investigation is required to establish whether the unconfirmed candidate species actually represent independently evolving lineages. It is therefore intended that a sample of individuals from each unconfirmed candidate species be sequenced for several additional nuclear loci, to allow for quantitative assessment of their taxonomic status (see Chapter 5 for illustration of taxonomic analysis for multi-locus data).

4.4.3. Compiling accurate ecological data

Western Palaearctic oak gall communities have a long history of detailed study (e.g. (Askew 1961a, Schonrogge et al. 1995, Pujade-Villar et al. 2003), resulting in

extensive collections of specimens and catalogues of geographic distributions and host associations for established taxa (Schönrogge et al. in prep, Stone et al. in prep). These have been a valuable resource for evaluating the phylogeography of individual species (Hayward and Stone 2006, Stone et al. 2009), comparing ecological roles across sites (Schönrogge et al. 2007), and investigating character state evolution (Cook et al. 2002, Stone et al. 2009). Unfortunately, when species level classifications are modified, the value of existing ecological data for the taxa concerned is greatly reduced as records can no longer be confidently linked to a single species. In such situations, the compilation of species level data must either begin again from scratch, or specimens must be reassessed to establish taxon membership under the new classification. Providing the assumptions of DNA barcoding are satisfied, barcodes can offer a valuable tool for confidently establishing taxon membership and thus for re-compiling accurate ecological data. This may be particularly useful if newly defined species are morphologically cryptic, and morphological reassessment of specimens is therefore not an option. Additionally, even if classifications are subsequently revised further, the linking of ecological data to a barcode makes it relatively straightforward to re-associate ecological data with the appropriate species level taxa, without the need to revisit specimens.

The present study identifies potential taxonomic error within 8 parasitoid morpho-species, and if confirmed, existing ecological data for these species should be interpreted with caution. For *Torymus flavipes* where the two MOTUs defined here could be matched with those previously identified by Kaartinen et al. (2010), the host association data from these studies can be compiled to gain a better understanding of the ecology of these morphologically cryptic species. From their sample, Kaartinen et al. (2010) suggested that the two species may have specialised phenologies, with one species (species B, that matched with *T. flavipes* MOTU 2 in this study) targeting only asexual generation galls, and the other (species A, that matched with *T. flavipes* MOTU 1) predominantly targeting sexual generation galls. While this specialisation of species B is further supported by the data presented here, with all 27 individuals emerging from asexual generation galls of *Neuroterus albipes,* more than half of the 136 individuals of species A also emerged from *N. albipes* asexual galls,

indicating that these species are less discrete in their host associations than was initially believed.

For the Synergini inquilines where the validity of many morpho-species is currently dubious, perhaps the only reliable existing ecological data is that linked to the MOTUs identified by Ács et al. (2010) and Kaartinen et al (2010). Five of the 10 inquiline MOTUs identified here could be matched with one or both of these studies, and thus ecological data can begin to be re-compiled. For example, inquiline MOTU 6 that had 8 host gall types in this study is also associated with one of these types in Hungary (Ács et al. 2010), and with one of these and one further type in Finland (Kaartinen et al. 2010), giving it a total of 9 host gall associations. Full details of compiled geographic and host association data are provided in Appendix 4.2.

4.4.4. The extent of bias due to taxonomic error

A principal reason for assessing the taxonomy of this oak gall community is that undetected taxonomic error could bias the investigation of ecological communities, potentially resulting in incorrect interpretation of ecological processes. The analysis of MOTUs presented here indicates taxonomic lumping within the parasitoid community that would be undetected if identification were based on current morphological classifications. If the inferences based on MOTUs are assumed to be correct then 35% of encountered parasitoid morpho-species contained cryptic species. The number of parasitoid species increased by 65%, affecting the identity of 54% of all parasitoid individuals. This high encounter rate for cryptic species is comparable to the approximately 25% reported by Smith et al. (2008, 2011), and is consistent with their assertion that cryptic diversity of parasitoid insects is a real phenomenon that is not restricted to tropical communities. However, these percentages are much greater than those encountered in a similar study system by Kaartinen et al. (2010), illustrating that it may not be possible to make *a priori* generalisations about the frequency and impact of taxonomic error within a system.

In addition to the obvious influence on the estimation of species diversity within a community, undetected taxonomic error may also bias the estimation of network metrics that are commonly used to investigate and compare community structure

(Kaartinen et al. 2010). Again if it assumed that the taxonomic inferences based on MOTUs are correct for this study, then estimates of quantitative generality (G_q) and weighted connectance (C_w) for the bipartite parasitoid network would have been positively biased if based on morpho-species, while quantitative vulnerability (V_q) and quantitative linkage density ($L.D_q$) would have been negatively biased (Figure 4.5). The magnitude of bias varied between metrics, but was greatest for V_q where the estimate based on MOTU species was 25% greater than that based on morpho-species.

These results suggest that if undetected, taxonomic error would severely bias the estimation of various ecological parameters of this parasitoid community, and the assessment of morpho-species classifications using molecular based taxonomic techniques is therefore considered to be an important pre-requisite for accurate ecological investigation. Given the widespread reports of species level error in parasitoid morpho-species classifications, predominantly involving the lumping together of morphologically cryptic species (Smith et al. 2008, Smith et al. 2011), it is likely that this statement applies to many other host-parasitoid systems centred on various guilds of insect herbivores. However, the direction and magnitude of bias is likely to vary between communities depending on the type of taxonomic error (i.e. lumping or over-splitting) and on the particular trophic association patterns of the species involved, precluding generalisation beyond the few communities that have been studied in appropriate detail. Further study comparing bias across communities is required if the true extent of this issue is to be understood and addressed.

Chapter 5 – Cryptic species of oak gall parasitoid revealed through DNA barcoding and observations of monophyly across ten intron loci

5.1. Introduction

The number of recognised eukaryotic species level taxa is thought to be approximately 1.9 million (Hamilton et al. 2010), the vast majority of which have been described solely on the basis of morphological characters (henceforth referred to as morpho-species). While this system of delimitation has been of undoubted value in the epic task of classifying biological diversity, there is a growing appreciation that such taxa may be discordant with modern species concepts that view 'existence as a separately evolving metapopulation lineage' as the principal property of species (De Queiroz 2005, 2007). Integrated taxonomic approaches, usually involving molecular tools, continue to reveal examples of species level taxonomic error including instances of over-splitting, in which variants of a single species are classified as two or more, and more frequently under-splitting (or 'lumping'), in which two or more distinct species are classified together (Bickford et al. 2007).

If undetected, such taxonomic error can be a major hindrance to ecological research, potentially biasing estimates and comparisons of diversity and community structure (discussed in Chapter 4). The nature and extent of such bias is likely to be variable and difficult to predict, being dependent on the ecology and distributions of the taxa involved. Cases where cryptic or over-split taxa are sympatric may bias estimates of alpha diversity and the structure of interacting communities (Kaartinen et al. 2010, Smith et al. 2011), whereas those involving allopatric species could affect interpretation of beta diversity, and potentially the conservation status of the species involved (Roca et al. 2001). Cases where trophic interactions differ substantially between cryptic or over-split taxa may be of particular concern, potentially biasing

metrics such as generality that are used to characterise species or networks (Hassell and May 1986). Where correct taxonomy is integral to the objectives of ecological research, such as in systems of conservation priority or those that are models for investigating community structure and dynamics, there is a pressing need for reassessment of taxa that were established solely on morphological characters.

As described in Chapter 4, DNA barcodes – short sequences from a standardised region of DNA - can be a useful tool for investigating morpho-species accuracy. Where the assumption that barcode variation within species is less than and discrete from variation between species is supported by the observation of a barcoding gap (Meyer and Paulay 2005, Ács et al. 2010), molecular taxonomic units (MOTUs, Blaxter et al. 2005) defined at a sequence similarity thresholds within the barcoding gap can be compared with morpho-species classifications to identify points of discordance. However, while MOTUs based on barcodes are a useful unit for highlighting potential taxonomic error and developing alternate taxonomic hypotheses, they are potentially inconsistent with true patterns of speciation due to stochastic coalescent variation, incomplete lineage sorting, or introgression (Hudson and Coyne 2002, Rokas et al. 2003b, Meyer and Paulay 2005, Hickerson et al. 2006, Lohse 2009). Therefore, taxonomic hypotheses based on barcodes, or any single locus, can only be considered as unconfirmed candidate species (i.e. deep genealogical lineages of unknown status, Vieites et al. 2009), unless their taxonomic distinctiveness is supported by further independent taxonomic characters (Padial et al. 2010).

When a population of organisms diverges into two separate species, it is expected that genetic variation will initially be paraphyletic, i.e. an allele sampled from either species will as likely coalescence first with homologous alleles from the other species, as than within its own species (Baum and Shaw 1995). With the passage of generations, lineages will be lost within each species through drift, and both will eventually become monophyletic i.e. alleles will coalescence within species. This process, known as lineage sorting, is central to DNA barcoding as it is through lineage sorting that barcode sequence variation within species becomes discrete from variation among species (Hebert et al. 2003). As lineage sorting of unlinked

molecular markers is independent, patterns of monophyly at loci unlinked to the barcode locus can provide further taxonomic information.

The potential of using observed monophyly for taxonomic inference has long been recognised under the genealogical species concept (Baum and Shaw 1995), and has frequently been applied to qualitatively support taxonomic inferences (Monaghan et al. 2005, Smith et al. 2007, Kaartinen et al. 2010, Nicholls et al. 2010). Tests have been developed for assessing the probability of taxonomic distinctiveness from observations of monophyly against a null hypothesis of a single panmictic population under widely applied phylogenetic and population genetic models (Rosenberg 2006, 2007, Zhu et al. 2011). Where data can be obtained for an appropriate sample of additional molecular markers, these tests offer a means of explicitly assessing taxonomic hypotheses drawn from barcode data. In the present study, a DNA barcoding approach is supplemented with analysis of monophyly at additional markers to assess the validity of morpho-species taxa in a group where accurate taxonomy is of importance for ecological study – oak gall parasitoids in the genus *Cecidostiba*.

5.1.1. Study system

The Cynipini gallwasps (Hymenoptera; Cynipidae) are an intriguing group of insects whose larva induce complex galls on trees of the Fagacaea family. In the Western Palaearctic, these galls support multi-trophic communities that may include inquiline gallwasps (Hymenoptera; Cynipidae; Synergini) and hymenopteran parasitoids (of various families in the Chalcidoidea superfamily). These communities are a popular model system for community ecology, and have recently been the focus of studies of biological invasions (Schonrogge et al. 1995, Schönrogge and Crawley 2000), comparative phylogeography (Hayward and Stone 2006), local adaptation (Tack and Roslin 2010), habitat fragmentation (Kaartinen and Roslin 2011), and community genetics (Tack et al. 2010, see Chapter 3). Species within the community have for the most part been described and identified from morphology, but recent molecular approaches have revealed widespread error in the species level taxonomy of the inquilines and parasitoids (Ács et al. 2010, Kaartinen et al. 2010, Nicholls et al.

2010, see Chapter 4). This is a serious concern, potentially impeding further ecological study (see Chapter 4), and raising doubts about the validity of existing ecological records and inferences. Further assessment of established morpho-species and the development of an accurate taxonomic framework for the various components of the community will be of value if it is to continue as a model ecological system.

Members of the genus *Cecidostiba* (Hymenoptera: Pteromalidae) are ectoparasitoids of hymenopteran larvae and pupae within cynipid galls in the Palaearctic region. Eight morpho-species species are currently recognised, and the most commonly encountered are *C. fungosa* and *C. semifascia,* having been recorded from 67 and 13 species of host gall respectively, across much of the western Palaearctic (R.R. Askew, J.-L. Nieves Aldrey, G. Stone, K. Schönrogge, unpublished data). These are distinguishable from each other, and from other members of the genus, by a combination of morphological characters (see Figure 5.1). The remaining six formally recognised species have smaller geographic and host ranges and are rarely encountered in ecological studies of Western Palaearctic oak galls (see Table 5.1). Recently, an undescribed morphological candidate species of *Cecidostiba* has been reared from galls of several species in Iran. This candidate species (hereafter referred to as *Cecidostiba* species A) closely resembles *C. fungosa*, but is distinguishable by its paler legs, antenna, and wing venation (R. R. Askew, personal communication).

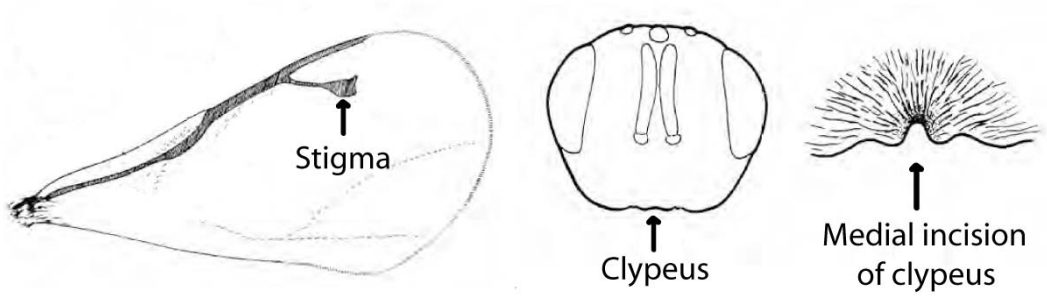


Figure 5.1. Illustration of morphological characters for distinguishing between *Cecidostiba* species. *C. fungosa* has an enlarged forewing stigma and a median incision on the anterior margin of the clypeus, while *C. semifascia* lacks a medial incision and has a narrow stigma, often with a dark band beneath it. Images modified and reproduced with permission from unpublished key by R,R. Askew & C. Thúroczy.

Table 5.1. Details of the eight formal - and one candidate - *Cecidostiba* morpho-species, including summarised host records and geographic range. Two letter country codes follow those provided by the International Organization for Standardization, and are presented in approximate geographic order from west to east. Records are largely taken from an unpublished catalogue of Western Palaearctic Cynipini gall parasitoids compiled by R.R. Askew, J.L. Nieves-Aldrey, J. Pujade-Villar, S.E. Sadeghi, K. Schönrogge, G. Melika and C.Thuróczy, with additional records from outside the Western Palaearctic and from alternate host gall tribes taken from Kamijo (1981) and Askew et al. (2006).

Species	Recorded host gall tribe	Number of recorded host gall species	Geographic distribution
Cecidostiba atra	Cynipini	3	ES
Cecidostiba docimus	Pediaspini	1	ES, FR, DE, IT, HU
Cecidostiba fungosa	Cynipini	67	ES, AD, GB, FR, BE, NL, CH, DK, DE, IT, SE, AT, HR, SK, CZ, HU, GR, RO, BG, HA, UA, IL, JO, TR, IR
Cecidostiba fushica	Cynipini	2	JP
Cecidostiba geganius	Cynipini, Diplolepini	3	ES, FR, NL, DE
Cecidostiba iliciana	Cynipini	4	ES, AD
Cecidostiba saportai	Cynipini	3	ES, AN, GB, FR,
Cecidostiba semifascia	Cynipini	13	ES, AD, GB, FR, CH, DE, SE, AT, HR, SK, HU, HA, IR
***Cecidostiba* species A**	Cynipini	4	IR

5.1.2. Objectives

The primary objective of this study is to assess the validity of morpho-species classifications for *Cecidostiba fungosa*, *C. semifascia*, and *C.* species A, proposing corrections where necessary and facilitating accurate species level identifications in future studies. The approach utilises DNA barcoding of individuals from across the geographic range of each morpho-species to develop taxonomic hypotheses that are

then tested by analysing observations of monophyly across 10 additional loci. Specifically I ask: (i) beginning with barcode MOTUs defined at a divisive sequence similarity threshold, do observations of monophyly across additional loci support the taxonomic distinctiveness of barcode MOTUs? (ii) If novel taxa are inferred, how do they differ from the original morpho-species in their geographic and host ranges? The implications of the resulting taxonomic inferences, and the utility of this approach are then discussed.

5.2. Methods

5.2.1. Sample selection

A set of 171 parasitoid individuals were selected from the collections of the Stone laboratory at the University of Edinburgh. Parasitoids in the collections are stored in 99% ethanol and have been identified into nominal species using morphological keys in collaboration with taxonomic experts including Dr Richard Askew, Dr George Melika, and Dr Csaba Thuróczy. The set included 134 individuals of *Cecidostiba fungosa* (the most common of the *Cecidostiba* species), 26 individuals of *Cecidostiba semifascia*, and 10 individuals of the undescribed candidate species (*Cecidostiba* species A). To maximise intra-species diversity, individuals of each species were sampled across the available range of host gall species and geographic localities (see Appendix 5.1 for collection and rearing details). One individual of *Caenacis lauta* (Pteromalidae) - a gall parasitoid from a closely related genus - was included as an out-group for phylogenetic analyses.

Following the definition of barcode MOTUs (described in section 5.2.4), a subset of individuals was selected for further sequencing. Based on the power analysis of Rosenberg (2007), a sample size of three individuals per MOTU was considered to offer an appropriate compromise between effort and resolution. This set therefore included three individuals from each of the eight barcode MOTUs that had at least three members, the one individual from both MOTUs that contained only a single individual, and the one individual of *Caenacis lauta* for use as an out-group. Individuals that had already been sequenced by Lohse et al. (2010) were

preferentially selected. Obtaining accurate DNA sequence is potentially problematic in diploid or polyploid organisms as allelic heterozygosity can result in dual peaks within sequencing chromatograms, and a time consuming cloning process may be required (Lohse et al. 2011). It was attempted to avoid these issues by preferentially selecting male specimens, which are haploid in hymenoptera, but as several MOTUs contained less than three males it was necessary to include five females.

5.2.2. Molecular methods

DNA was extracted for all individuals from a single leg (abdomens were used for some small male specimens) using a solution of chelex and proteinase K (Nicholls et al. 2010). For the full set of 176 individuals, a 652bp 'barcode' fragment of the mitochondrial *coxI* gene was amplified using the forward primer LCO1490 (Folmer et al. 1994), and the reverse primer HCOd: 5'-TAW ACY TCD GGR TGI CCA AAA AAY CA-3' (modified from Folmer *et al's* (1994) HCO2198 by J. Nicholls). Each 20 µl PCR mix consisted of 1 µl of DNA template, 0.1 µl of Taq polymerase (5 U/µl, Bioline), 0.25 µl of each primer (20 µM), 1 µl of dNTP's (25 mM each), 0.8 µl of $MgCl_2$ (50 mM), 2 µl of 10 x Bioline PCR buffer, 2 µl of bovine serum albumin (10 mg/mL), and 12.6 µl of milipure H_2O. Cycling conditions were 94 °C for 2 minutes, followed by 4 cycles of 94 °C for 30 seconds, 45 °C for 1 minute, and 72 °C for 1 minute, then 34 cycles of 94 °C for 30 seconds, 50 °C for 1 minute, 72 °C for 1 minute, with a final step of 72 °C for 5 minutes.

For the subset of 27 individuals, sequences of 10 nuclear exon-primed intron crossing (EPIC) loci were obtained from GenBank for 5 individuals of *C. fungosa* and one individual of *C. lauta* that were previously analysed by Lohse et al. (2010). These loci are known to amplify well in *C. fungosa* and are expected to contain sufficient sequence polymorphism to differentiate between closely related species (Lohse et al. 2011). They include 7 ribosomal protein genes (*RpS4, RpS8, RpS18, RpS23, RpL15, RpL37, and RpL37a*) and 3 regulatory genes (*AntSesB, Ran, Sansfille*). For the remaining 21 individuals, these 10 loci were amplified using established primers and PCR conditions (Lohse et al. 2011, full details provided in Appendix 5.2).

Excess dNTPs were removed from PCR products by adding a solution of shrimp alkaline phosphatase (SAP) and exonuclease, incubating at 37°C on a PCR block for 40 minutes, then heating to 94°C for 15 minutes. Clean PCR product from all loci was sequenced in both directions using ABI BigDye chemistry (Perkin Elmer Biosystems Waltham MA) on ABI 3700 and 3730 sequencers at the GenePool, Edinburgh.

5.2.3. Sequence alignment

For each locus, complementary forward and reverse sequences were aligned, checked by eye, and trimmed to exclude primer sequence using Sequence Navigator version 1.0.1 (for barcodes), and Sequencher version 4.9 (for EPIC loci). Additional trimming from the ends was occasionally required for EPIC loci where chromatograms could not be reliably interpreted. Alignment of sequences was carried out by eye for barcodes, and with MAFFT online version 6 (Katoh and Toh 2008) using the default settings for each EPIC locus. Exonic regions of the EPIC loci were identified by comparison with annotated sequences provided by Lohse et al. (2010), and all exons were checked for on open reading frame, a lack of which would indicate base calling error or amplification of a pseudogene. Indels were retained in the alignments and were treated as missing data in the analyses. Within the final alignments, the number of polymorphic sites was assessed using the program PAUP version 4.0 BETA (Swofford 2003), across all taxa and within each nominal *Cecidostiba* species.

The modelling of sequence evolution and the estimation of gene trees for the EPIC loci described below assume that allelic lineages have evolved without recombination. Each multiple sequence alignment was therefore checked for recombination with the program RDP3 version 3.44 (Martin et al. 2010), that collectively utilizes the following 'scanning window' methods to identify individual recombination events (Salminen and Martin 2009): RDP (Memmott et al. 2000), GENECONV (Padidam et al. 1999), MaxChi (Maynard Smith 1992), BootScan (Martin et al. 2005), and SiScan (Gibbs et al. 2000). One locus (*RpL37*) showed evidence of recombination within a 111 bp region (P-value = 0.026 with RDP

method, 0.032 with GENECONV method), and this region was consequently removed from the alignment prior to analysis.

5.2.4. MOTUs

The objective here was to define a series of molecular operational taxonomic units (MOTUs) based on barcode sequence similarity, whose taxonomic distinctiveness could then be assessed by analysing patterns of monophyly at the EPIC loci. Barcodes were analysed with jMOTU version 1.0.8 (Jones et al. 2011a), which clusters sequences into MOTUs based on a specified threshold of base-pair differences using a combination of BLAST and the Needleman-Wunch exact global alignment algorithm. To assess for a barcoding gap, the number of defined MOTUs was calculated across a threshold range from 0 to 60 base pairs (corresponding to 0 and ~ 9.2% of the fragment length). The existence of a barcoding gap is characterised by a plateau of MOTU richness across a range of threshold values, bounded at either end by a relatively steep decline (see Chapter 4, Figure 4.1). Although there is no formal means of defining a barcode gap, its 'divisive' and 'inclusive' limits can be estimated from a MOTU richness plot (Ács et al. 2010). Based on the assumption of species monophyly and greater variation between than within species, it is expected that MOTUs defined at thresholds within a barcoding gap represent meaningful independently evolving lineages, but there is likely to be a gradient from increased risk of over-splitting meaningful taxa at the divisive limit, towards lumping of taxa at the inclusive limit. To minimise the risk of lumping meaningful taxa within a single MOTU, a threshold was selected from immediately prior to the barcoding gap (i.e. before the plateau of MOTU richness). Each MOTU defined at the selected threshold was considered to represent a hypothesised species level taxa, and a sub-sample of individuals from each was selected for further sequencing of EPIC loci (see section 5.2.1)

The relationships between barcode haplotypes for all *Cecidostiba* individuals were visualised in a neighbour-joining tree, constructed with MEGA version 4 (Tamura et al. 2007) using p-distances. Groupings based on morphology and MOTUs at the selected threshold were illustrated in the tree and assessed for discordance.

5.2.5. Model selection and EPIC gene tree estimation

To allow for interpretation of patterns of monophyly, phylogenetic relationships between the sequences for each EPIC loci were estimated using Bayesian MC^3 sampling, implemented in the program MrBayes version 3.1.2 (Ronquist and Huelsenbeck 2003). All analyses initially used 2 runs of 2 million generations, with 4 chains per run and a hot chain temperature factor of 0.01. In most cases this allowed for adequate mixing between chains and convergence of the tree topology parameter between runs, as assessed using Tracer version 1.5 (Rambaut and Drummond 2007). Runs where convergence was not considered to be adequate (i.e. where the average standard deviation of split frequencies remained above 0.01) were repeated with a run length of 3 million generations. Parameters and trees were sampled every 1000 generations with a burn-in of 1000 samples.

As rates of sequence evolution are expected to vary between coding and non-coding DNA, all alignments were partitioned by intron and exon, with variable substitution rates allowed between partitions. A generalised time reversible (GTR) substitution model was applied where a partition contained all base substitution types, and a Hasegawa, Kishino and Yano (HKY) model where some types were absent. For evolutionary models with these substitution settings, models containing various additional parameters were compared using log Bayes factors (lnBFs), estimated as the natural log of twice the difference in harmonic mean likelihood between pairs of models (Kass and Raftery 1995, Ronquist et al. 2009). Such lnBFs represent a summary of the evidence for one model as opposed to another, and following Kass and Rafferty (1995), lnBFs of greater than 10 were considered to indicate a decisive difference in support between models. Beginning with models that included parameters for a proportion of invariable sites (I) and gamma distributed rate variation amongst sites (G) for each partition, models with unconstrained and clock constrained branch lengths were compared. For parsimony, the simpler unconstrained branch length models were accepted where the comparison of lnBFs did not indicate a decisive reduction in model performance. For the most appropriate branch length model, the posterior distributions of the estimated rate parameters were visualised using the program Tracer version 1.5 (Rambaut and Drummond 2007) and

simplifications (i.e. the removal of I or G) were attempted where posterior distributions failed to converge towards a single parameter value. Again, simplifications were accepted where comparison of lnBFs did not indicate a decisive reduction in model performance. For the resulting 'best' model for each locus (i.e. the most parsimonius yet adequate model), samples from the final 1 million generations were used as posterior distributions for tree topology and node support, to generate majority-rule consensus trees (i.e. containing all nodes that were present in more than 50% of sampled tree topologies).

5.2.6. Assessment of monophyly

In each of the estimated consensus gene trees, a particular grouping of individuals was considered monophyletic if all members descended from a node that did not also descend to members of any other group. Under the null hypothesis that a group of *c* individuals represents a single taxonomic entity, the probability of monophyly (P_A(a, *b*)) for a particular group *A*, consisting of *a* individuals, can be given by the equation:

$$P_A\,(a,b) = \frac{2}{\binom{a+b}{a}} \frac{a+b}{a\,(a+1)} \qquad Equation\ 5.1$$

where *b* is the number of individuals from outside of group *A*, and $a + b = c$ (the term $\binom{a+b}{a}$ is a binomial coefficient, Rosenberg 2007 eq.1). The corresponding probability of non-monophyly for a particular group *A* can be calculated as $1 - P_A(a, b)$.

Where the status of *A* is observed across multiple loci (*L*), the compound probability for a particular set of observations ($P_{(obs)}$) is given by the equation:

$$P_{(obs)} = \binom{L}{k} \left[\prod_{i=1}^{L} P_A(obs)_i\right] \qquad Equation\ 5.2$$

where *A* is monophyletic at *k* loci, and $P_A(obs)_i$ is the probability of the observed status of *A* at locus *i* (adapted from Rosenberg 2007 eq.6 to allow for *a* and *b* to differ between loci). Following equation 5.1, if *A* is monophyletic for a particular

locus then $P_A(obs) = P_A(a, b)$, and if A is not monophyletic then $P_A(obs) = 1 - P_A(a, b)$.

When applying Equations 5.1 and 5.2 to assess the taxonomic distinctiveness of a particular grouping A, the value of a is simply the number of individuals considered to belong to A, but the value of b is more subjective and requires some consideration. One option is to set the value of b to include all other individuals in the sample, whereby the alternative to the null hypothesis would effectively be that group A is not part of a single taxonomic entity that includes all sampled individuals (e.g. as the sample in this study included 27 individuals, if $a = 3$, then $b = 24$). However, rejection of the null hypothesis in this case would imply that group A is taxonomically distinct from some, but not necessarily from all taxa within the sample. To avoid this issue, I considered that the most appropriate approach was to always set $b = 1$, whereby the alternative to the null hypothesis is that group A is a distinct taxon that does not also contain any other single individual from within the sample. Therefore, the status (monophyletic or not monophyletic) of each barcode MOTU at each locus was observed, and the probability of each observation was calculated following Equation 5.1 with b set to 1, and a set to the number of sequenced individuals within the MOTU. At loci where a particular MOTU was represented by less than two individuals, its monophyletic status could not be interpreted and the locus was excluded from analysis of that grouping. The probability of the observed status of each barcode MOTU across all informative loci was calculated following equation 5.2. Where the observed patterns were not sufficiently unlikely (i.e. $P(obs) > 0.05$) to reject the null hypothesis for two or more 'sister' barcode MOTUs (i.e. those that descended from an otherwise exclusive node of the barcode tree), these were combined and patterns were reassessed.

The majority-rule consensus trees produced in Mr Bayes contain all nodes that were present in more than 50% of sampled tree topologies, and offer the proportional occurrence of each of these nodes as a measure of node support. Where a particular grouping is monophyletic in a consensus tree, the proportion of sampled tree topologies where that grouping was not monophyletic can be inferred as $1 - Ns_A$, where Ns_A is the support (expressed as a proportion) for the node that determines the

monophyly of group *A*. In the analysis described above it is assumed that the topology of the estimated consensus tree is accurate, but the measures of node support for monophyletic groups offer a means of incorporating gene tree uncertainty into the assessment of monophyly, thus minimising the effect of type I error (i.e. false positive inferences of monophyly). A corrected probability of monophyly for a group *A* ($CP_A(a, b)$) was calculated using the equation:

$$CP_A(a,b) = P_A(a,b) + (1 - Ns_A)(1 - P_A(a,b)) \qquad Equation\ 5.3$$

Where a grouping is not monophyletic in the consensus tree, it is not straightforward to infer the proportion of sampled tree topologies in which it was monophyletic without recovering the excluded tree topologies. The corresponding probability for a lack of monophyly of *A* therefore remains as 1 - ($P_A(a, b)$). As this approach considers false-positive observations of monophyly, but not false negatives, it offers a more conservative estimation of the probabilities associated with particular patterns of monophyly. The analysis was repeated using equation 5.3 to calculate the probability for the observed pattern of monophyly for each grouping at each locus, and these were combined across loci following equation 5.2 to give a corrected compound probability for each set of observations ($CP_{(obs)}$).

5.3. Results

5.3.1. Barcodes and MOTUs

A 652 bp barcode fragment was obtained from all 171 individuals. The alignment of these sequences contained 223 polymorphic sites (see Table 5.1), with an open reading frame in the forward direction from the 2nd base, and no evidence of insertions, deletions, or recombination. When read in the forward direction, all sequences contained a series of 11 consecutive thymine bases beginning from the 150th base pair. This 'poly-T' region appeared to cause slippage of the taq polymerase during the extension phase of the PCR reaction, resulting in sequence chromatograms that contained dual peaks downstream of the poly-t region. However, as the fragment was sequenced in both directions for all individuals, at least one sequence was always interpretable, and by carefully editing by eye it was possible to

obtain accurate sequence data for the full fragment. This issue has also been encountered in parasitoids from various other chalcid families (J. Nicholls, personal communication), but could be avoided in future by using alternate primers that do not span the poly-t region (e.g. the pF1 and 2437d primers described in Chapter 4).

The relationship between the number of barcode MOTUs and the threshold base pair difference is shown in Figure 5.1. A plateau of MOTU richness - characteristic of a barcoding gap - was apparent between thresholds of approximately 8 and 47 bp (corresponding to 1.2 – 7.2% of the fragment length). To minimise the risk of lumping multiple independent lineages, a threshold of 6 bp (0.9%) was selected for defining a series of MOTUs for further investigation. At this threshold, 10 *Cecidostiba* barcode MOTUs were defined containing between 1 and 103 individuals.

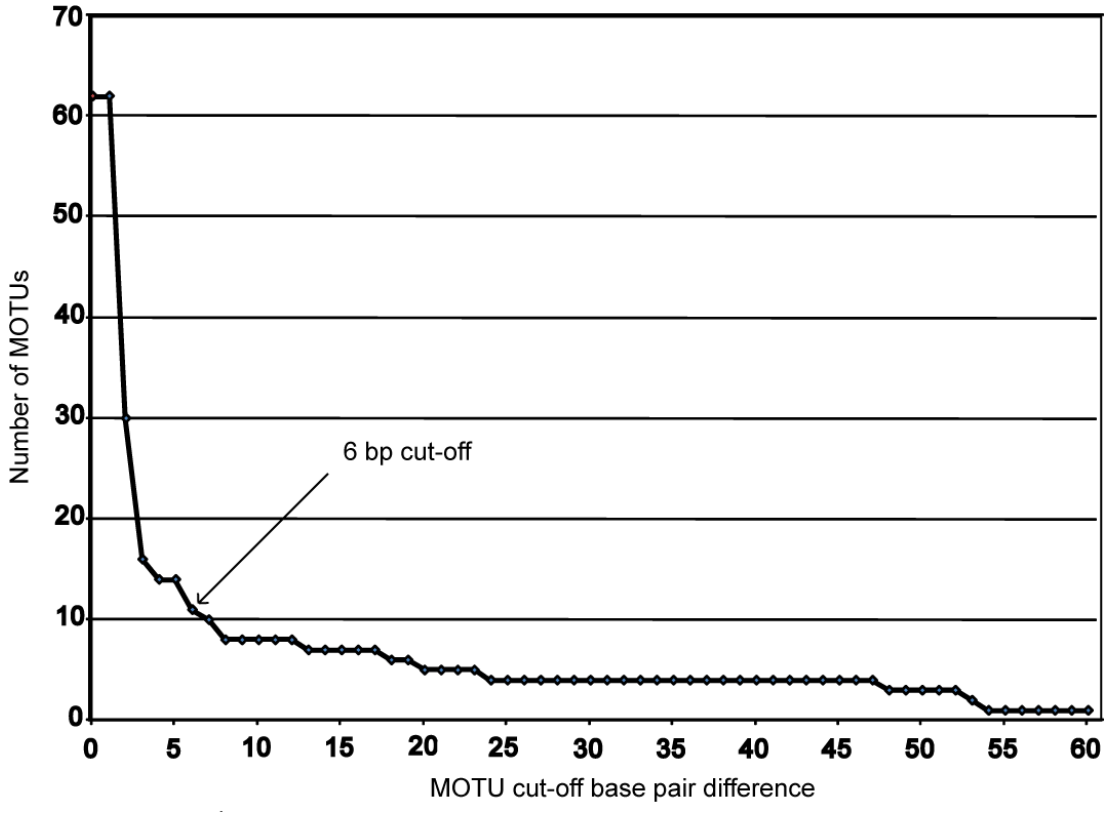


Figure 5.1. Relationship between the number of barcode MOTUs and defining threshold base pair difference between 0 and 60 base pairs. Arrow indicates the selected divisive threshold of 6 bp.

.

A neighbour-joining tree of all *Cecidostiba* barcode haplotypes is shown in Figure 5.2. Each of the three morpho-species was monophyletic within this tree, and none of the 10 MOTUs contained members of more than one morpho-species. One MOTU corresponded exactly with *Cecidostiba* species A, while *C. fungosa* contained 4 MOTUs and *C. semifascia* contained 5.

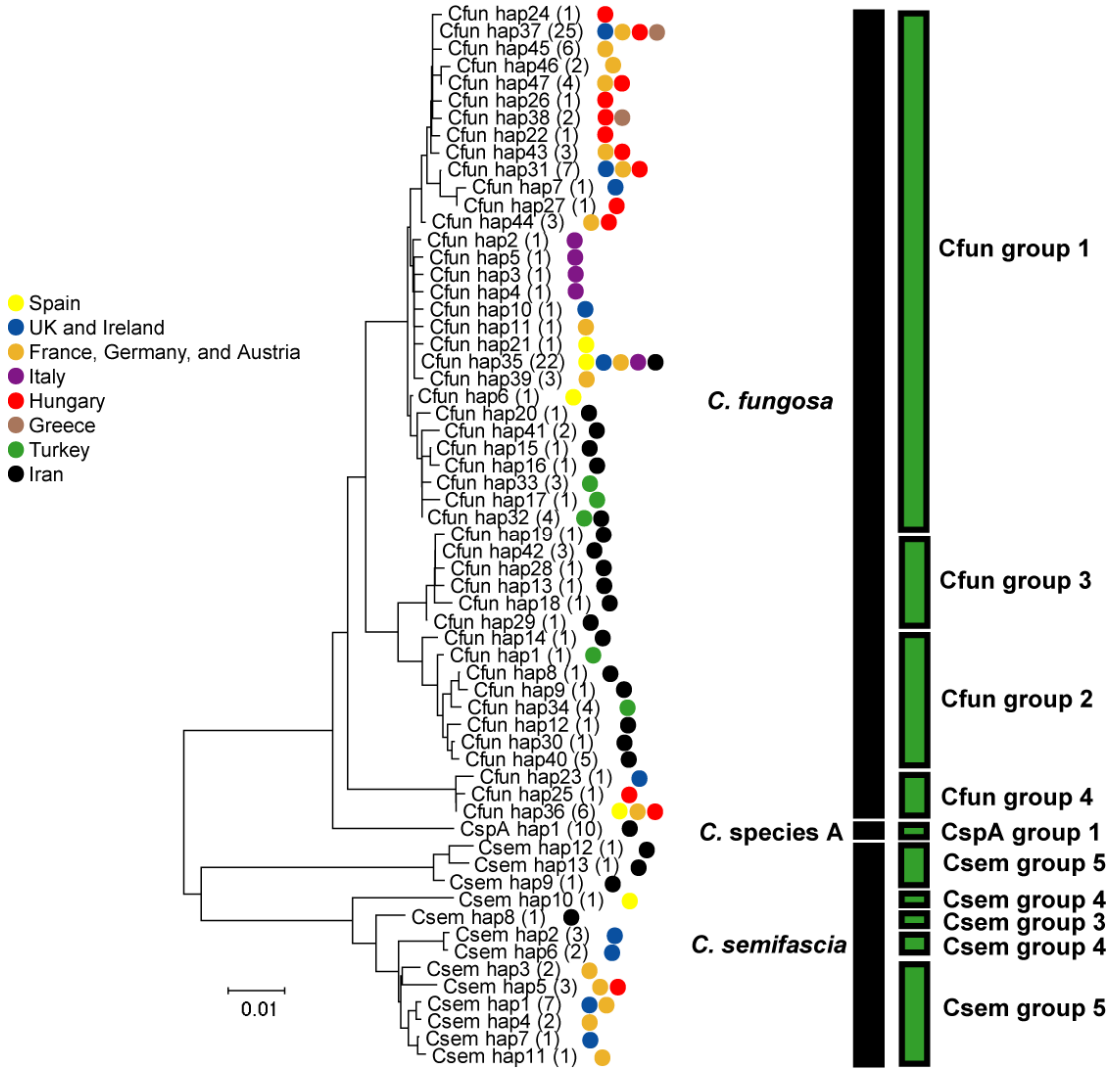


Figure 5.2. Neighbour-joining tree based on p-distances for *Cecidostiba* barcode haplotypes. Haplotype labels consist of a 4 letter morpho-species code and a 4-5 digit haplotype code, with numbers in brackets indicting how many parasitoid individuals shared a haplotype (Cfun = *C. fungosa*, Csem = *C. semifascia*, CspA = *C.* species A). Coloured circles to the right of haplotype codes indicate the countries from which individuals were sampled. Vertical bars to the right indicate membership of morpho-species (in black), and MOTUs as defined at the 6 bp threshold (in green). Scale bar indicates a p-distance of 0.01

5.3.2. Nuclear loci

A single allele from each of 21 selected individuals was successfully sequenced for the 10 EPIC loci in all but nine instances (i.e. 201 sequences were obtained from a maximum of 210). When combined with the sequences of six individuals from Lohse et al (2010), five sequences were missing at the *AntSesB* locus, two at the *RpS4* locus, and one at the *RpL15* and *RpS18* loci. These missing sequences could not be obtained due to a repeated lack of amplification during the PCR reaction, presumably due to polymorphism at sites within the priming regions that inhibited primer binding. Although 5 of the selected individuals were female, and therefore diploid, there were no obvious ambiguities at any of the sequence trace files for these individuals, suggesting that they were homozygous at the sampled loci.

The alignments of each of the 10 loci contained polymorphic sites (see Table 5.1), ranging from 34 in the shortest alignment (*RpL37a,* 14% of sites) to 131 in the longest (*RpL37,* 25% of sites). Polymorphic sites were present within the nominal species *C. fungosa* and *C. semifascia* for all 10 loci, which is consistent with these loci being suitable for intraspecific investigation of these nominal species.

The models of sequence evolution selected for gene tree estimation had clock constrained branch lengths and a GTR + I model applied to the intron partition for all 10 loci. For the exon partition, a GTR + I model was applied for three loci (*AntSesB, RpL15,* and *RpS4*), and a HKY + I model for the remaining seven. Gene trees for all EPIC loci are shown as cladograms in Figure 5.3.

Table 5.1. Summary of multiple sequence alignments used in analysis, including the number of sequences (# Seq), lengths (including indels), number of introns (# Intron), and the number of polymorphic sites within the full alignment (# S) and within each nominal *Cecidostiba* species.

		Length (bp)							
Locus	**# Seq**	**Exon**	**Intron**	**Total**	**# Intron**	**# S**	**# S_{Cfun}**	**# S_{Csem}**	**# S_{CspA}**
AntSesB	22	435	192	627	2	101	7	49	3
Ran	27	297	212	509	1	75	9	21	0
RpL15	26	215	452	667	2	102	21	23	0
RpL37	27	108	745	853	1	213	73	36	9
RpL37a	27	141	102	243	1	34	4	9	0
RpS4	25	369	436	805	2	129	24	47	12
RpS8	27	230	261	491	1	69	17	24	0
RpS18	26	257	587	844	2	148	51	59	2
RpS23	27	189	87	276	1	39	12	10	2
Sansfille	27	362	87	449	1	49	15	16	6
coxI	176	652	-	652	0	223	63	80	0

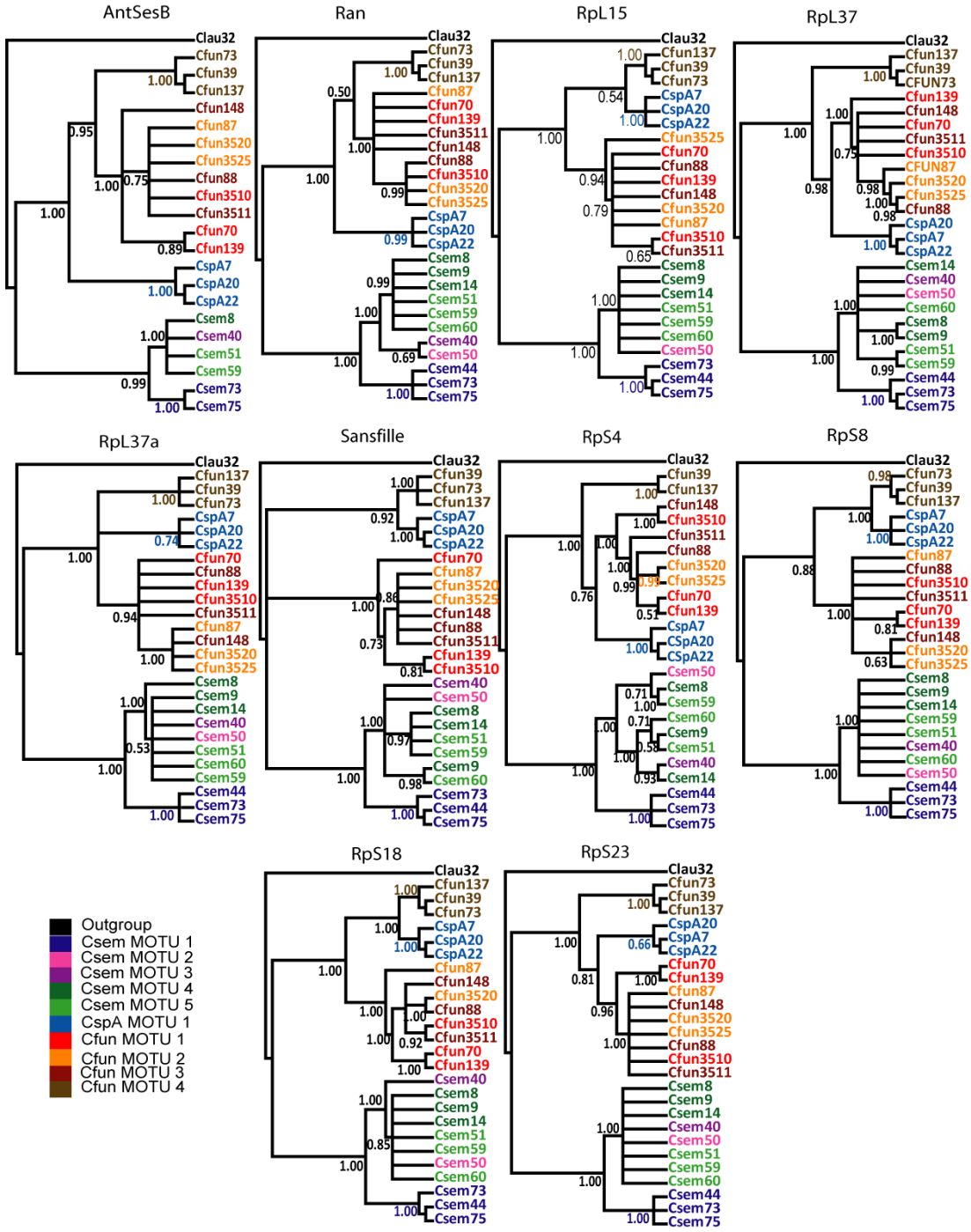


Figure 5.3. Bayesian majority consensus rectangular cladograms for 10 EPIC loci, rooted by outgroup (Clau32). The code names of individuals are coloured by the barcode MOTUs from which they were selected. Numbers to the left of nodes indicate posterior probability node support, and are coloured at nodes that support the monophyly of a particular barcode MOTU.

5.3.3. Assessment of monophyly across EPIC loci

Cecidostiba species A

The single MOTU from *Cecidostiba* species A (CspA MOTU 1) was monophyletic at all 10 EPIC loci, and the compound probability of observations across loci ($P_{(obs)}$) was sufficiently low to reject the null hypothesis that it was part of a larger taxonomic entity within the sample with a very high degree of confidence ($p<0.001$, see Table 5.2a). Support for the nodes that determined the monophyly of the CspA MOTU 1 was generally very high, and correcting the probabilities to incorporate uncertainty in gene tree estimation ($CP_{(obs)}$) resulted in only a slightly larger p-value that did not change the inference with regards to the null hypothesis. Thus, the status of this grouping as an independent lineage (i.e. a 'true' species) is very strongly supported.

Cecidostiba semifascia

Within the *C. semifascia* morpho-species, one of the barcode MOTUs (Csem MOTU 1) was monophyletic at all 10 loci, with a sufficiently low compound probability to reject the null hypothesis with a very high degree of confidence ($p<0.001$, see Table 5.2a). Correcting for gene tree uncertainty did not alter this inference, and its status as an independent lineage is therefore very strongly supported.

A limitation of using observations of monophyly to test for taxonomic distinctiveness is that monophyly can only be assessed for groups with at least two members. Two of the barcode MOTUs (Csem MOTUs 2 & 3) contained only a single individual and their monophyly could not therefore be assessed at any of the loci. However, evaluation of the sequence data revealed that the individuals of Csem MOTUs 2 & 3 shared identical haplotypes with members of at least one other MOTU at 4 and 5 of the 10 loci respectively, and so even if further members were available, the MOTUs could not have been monophyletic for these loci. It was therefore considered unlikely that the null hypothesis would have been rejected for these barcode MOTUs.

Neither of the remaining two *C. semifascia* barcode MOTUs (Csem MOTUs 4 & 5) were monophyletic at any of the loci, and the null hypothesis that they were part of a

larger taxonomic entity could not be rejected ($p>0.05$, Table 5.2a). Successive combinations of sister groupings were therefore attempted (i.e. those that shared a common but otherwise exclusive node in the barcode gene tree), and patterns of monophyly were reassessed (Table 5.2b). For the groupings of Csem MOTUs 4 and 5, and Csem MOTUs 3, 4 and 5, the compound probabilities were still not sufficiently low to reject the null hypothesis. However, the grouping of Csem MOTUs 2, 3, 4 and 5 was monophyletic at all 10 loci, and the compound probability for observations across loci was sufficiently low to reject the null hypothesis with a very high degree of confidence ($p<0.001$). The implication is therefore that these four barcode MOTUs represent a single taxonomic entity. The succession of barcode MOTU combinations is illustrated in Figure 5.4.

Cecidostiba fungosa

Within the *C. fungosa* morpho-species, one barcode MOTU (Cfun MOTU 4) was monophyletic at all 10 loci with a sufficiently low compound probability to reject the null hypothesis with a very high degree of confidence ($p<0.001$, see Table 5.2a). Correcting for gene tree uncertainty did not alter this inference, and its status as an independent lineage is therefore very strongly supported.

For each of the remaining three barcode MOTUs (Cfun MOTUs 1, 2, and 3) the null hypothesis of a larger taxonomic entity could not be rejected ($p>0.05$), and successive combinations of sister groupings were therefore attempted. The null hypothesis was not rejected for the grouping of Cfun MOTUs 2 & 3, but was rejected with a very high degree of confidence ($p<0.001$) for the grouping of Cfun MOTUs 1, 2, and 3 (Table 5.2b). The implication is therefore that these three barcode MOTUs represent a single taxonomic entity. The succession of barcode MOTU combinations is illustrated in Figure 5.4.

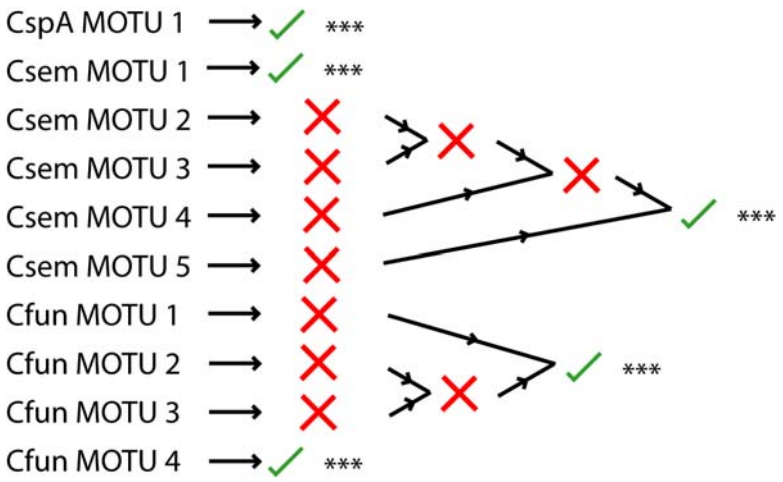


Figure 5.4. Illustration of how barcode MOTUs were assessed for monophyly and successively combined. Rejection of the null hypothesis that a particular grouping was part of a larger taxonomic entity based on the compound probability of observations across loci is indicated by a green tick with asterisks to show the confidence level (* = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$). A red cross indicates that the null hypothesis was not rejected for a particular grouping ($p>0.05$), and black arrows from left to right indicate the order that combinations of multiple barcode MOTUs were attempted.

Table 5.2. Summary of assessment of monophyly in EPIC loci gene trees for (a) barcode MOTUs, and (b) combinations of barcode MOTUs. The status of groups at each locus is indicated (monophyletic = Y, or not monophyletic = N), together with the probability of each observation when the group is considered against the null hypothesis that it is part of a larger taxonomic entity containing at least one other individual in the sample (i.e. *b*=1). The range of values for the number of individuals within each group (*a*) is provided, as are values of the binomial coefficient of number of loci over number of monophyletic loci (L over k). The final two columns contain the uncorrected (P(obs)) and corrected (CP(obs)) compound probabilities for the set of observations of each group.

Group	Range *a*	AntSes B	Ran	RpL15	RpL37	RpL37a	RpS4	RpS8	RpS18	RpS23	Sansfille	$\binom{L}{k}$	$P(_{obs})$	$CP_{(obs)}$
(a)														
CspA MOTU 1	3	Y 0.17	Y 0.17	Y 0.17	Y 0.17	Y 0.17	Y 0.17	Y 0.17	Y 0.17	Y 0.17	Y 0.17	1	1.65×10^{-8}	2.82×10^{-8}
Csem MOTU 1	2 - 3	Y 0.33	Y 0.17	Y 0.17	Y 0.17	Y 0.17	Y 0.17	Y 0.17	Y 0.17	Y 0.17	Y 0.17	1	3.31×10^{-8}	3.31×10^{-8}
Csem MOTU 2	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Csem MOTU 3	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Csem MOTU 4	1 - 3	NA	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	1	0.194	0.194
Csem MOTU 5	2 - 3	N 0.67	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.67	N 0.83	1	0.103	0.103
Cfun MOTU 1	3	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	1	0.162	0.162
Cfun MOTU 2	2 - 3	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	Y 0.33	N 0.83	N 0.67	N 0.83	N 0.83	1	0.517	0.522
Cfun MOTU 3	3	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	N 0.83	1	0.162	0.162
Cfun MOTU 4	2 - 3	Y 0.17	Y 0.17	Y 0.17	Y 0.17	Y 0.17	Y 0.33	Y 0.17	Y 0.17	Y 0.17	Y 0.17	1	3.31×10^{-8}	3.37×10^{-8}
(b)														
Csem MOTUs 4 & 5	3 - 6	N 0.83	Y 0.05	N 0.95	N 0.95	N 0.95	N 0.95	N 0.95	N 0.95	N 0.95	N 0.95	10	0.269	0.271
Csem MOTUs 3, 4 & 5	4 - 7	Y 0.1	N 0.96	N 0.95	N 0.96	N 0.96	N 0.96	N 0.96	N 0.96	N 0.96	N 0.96	10	0.712	0.712
Csem MOTUs 2, 3, 4 & 5	4 - 8	Y 0.1	Y 0.03	Y 0.04	Y 0.03	Y 0.03	Y 0.03	Y 0.03	Y 0.03	Y 0.03	Y 0.03	1	1.27×10^{-15}	1.27×10^{-15}
Cfun MOTUs 2 & 3	6	N 0.95	N 0.95	N 0.95	N 0.95	N 0.95	N 0.95	N 0.95	N 0.95	N 0.95	N 0.95	1	0.614	0.614
Cfun MOTUs 1, 2, 3	8 - 9	Y 0.02	Y 0.02	Y 0.02	Y 0.02	Y 0.02	Y 0.03	Y 0.02	Y 0.03	Y 0.02	Y 0.02	1	4.59×10^{-17}	5.36×10^{-17}

5.4. Discussion

5.4.1. Taxonomic inference

The presented DNA barcode data and analysis of observations of monophyly across EPIC loci strongly supported the status of *Cecidostiba* species A as a distinct taxonomic entity within the sample. Following the terminology of Vieites et al. (2009) and Padial et al. (2010), I therefore propose that it currently represents a confirmed candidate species. As distinguishable morphological characters have been identified (e.g. relating to the colouration of wing veins, legs and antenna, R. R. Askew, personal communication), it shall be formally described and included in appropriate taxonomic keys.

For the *Cecidostiba semifascia* morpho-species, the combination of DNA barcodes and analysis of observations of monophyly indicated the presence of two distinct taxonomic entities (i.e. independently evolving lineages). As these are supported by multiple independent lines of evidence (i.e. observations of distinctiveness at the barcode locus and multiple additional loci), I propose that they currently represent confirmed candidate species. Following the nomenclature recommended by Padial et al. (2010), the 3 individuals from Csem MOTU 1 shall henceforth be considered as *Cecidostiba semifascia* [Ca1], and the 23 individuals from Csem MOTUs 2, 3, 4, and 5, as *Cecidostiba semifascia* [Ca2]. Following the molecular analyses, representatives of these two taxa were sent for evaluation by Dr Richard Askew, a globally leading taxonomist of chalcid parasitoids, particularly those associated with cynipid galls. Initial indications are that although both fall under the morphological concept of *C. semifascia* and are currently morphologically cryptic, characters for distinguishing between them do exist. It is therefore possible that the morphological criteria of *C. semifascia* can be revised to allow for the description of *C. semifascia* [Ca1] and *C. semifascia* [Ca2] as distinct morphological species, which could then be included in morphological keys of the genus. Such a process of molecular assessment leading to morphological revision illustrates the reciprocally enlightened nature of

modern taxonomy, with DNA barcoding and other molecular approaches comfortably complementing more traditional morphological based taxonomy.

For the *Cecidostiba fungosa* morpho-species, the combination of DNA barcodes and analysis of observations of monophyly again indicated the presence of two distinct taxonomic entities (i.e. independently evolving lineages). As these are again supported by multiple independent lines of evidence, I propose that they currently represent confirmed candidate species. The 126 individuals from Cfun MOTUs 1, 2 and 3 shall henceforth be considered as *Cecidostiba fungosa* [Ca1], and the 8 individuals from Cfun MOTU 4 as *Cecidostiba fungosa* [Ca2]. Representatives of these two taxa were sent for evaluation by Dr Richard Askew, but initial indications are that they are truly morphologically cryptic species, with no reliable morphological characters for distinguishing between them.

5.4.2. Distributions and host records

The 10 individuals of *Cecidostiba* species A included in this study were reared from sexual generation galls of *Neuroterus saliens* and 3 unidentified gall types on oaks in the section *Cerris* (black oaks), at several sites in Iran (Figure 5.5, see Appendix 5.1 for collection details). There are presently no records of this morpho-species species from outside of Iran, although as it closely resembles *C. fungosa* and has yet to be included in morphological keys of the genus, it is possible that it has been recorded elsewhere as *C. fungosa.*

The new parasitoid species *C. semifascia* [Ca1] was widespread, with the 23 individuals included in this study having been reared from sexual generation galls of *Biorhiza pallida* (22 individuals) on oaks of the section *Quercus sensu stricto* (white oaks), and sexual generation galls of *Andricus quercuscalicis* (1) on a black oak, from six countries across the western Palaearctic from the UK to Iran. The species *C. semifascia* [Ca2] appears to have a more limited distribution, with the 3 included individuals having been reared from sexual generation galls of *Andricus cecconii* (1) and *Pseudoneuroterus macropterus* (2), on black oaks at sites in Iran (Figure 5.5).

The new parasitoid species *C. fungosa* [Ca1] appears to be a widespread bivoltine host generalist, with the 126 individuals included here having been reared from galls of 30 species including both sexual and asexual generations on both black and white oaks, collected in 10 countries spanning much of the western Palaearctic from the UK to Iran (Figure 5.5). Although much less numerous, *C. fungosa* [Ca2] was also widespread, with eight individuals reared from sexual generation galls of three species on white oaks, collected in five countries including Spain, Hungary, and the UK (Figure 5.5). The recorded geographical and host range of *C. fungosa* [Ca2] was completely over-laid by that of *C. fungosa* [Ca1], and at particular sites in Spain and France, both were reared from the same species and generation of gall.

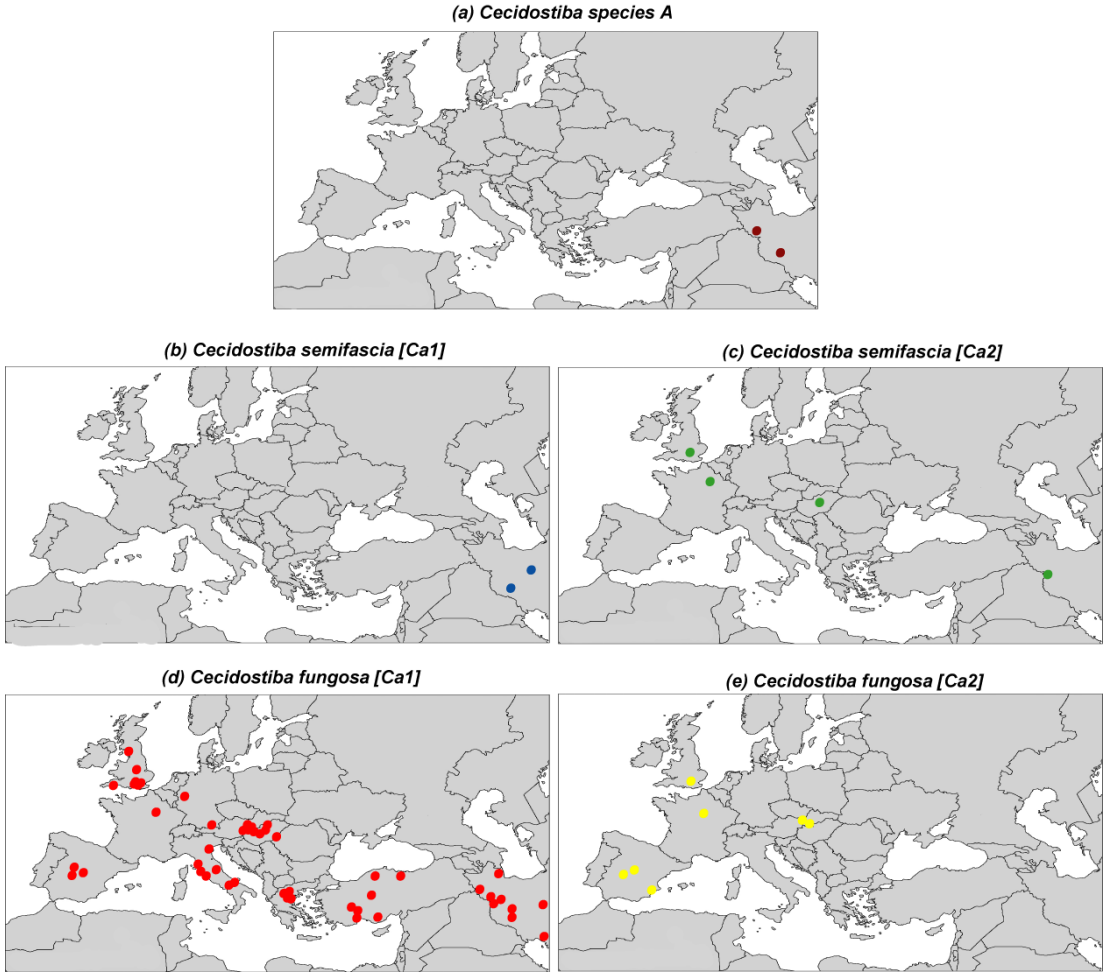


Figure 5.5. Maps of the Western Palaearctic showing the collection localities of the five *Cecidostiba* species recognised in this study (a-e).

The two pairs of cryptic species appeared to be sympatric, with *C. semifascia* [Ca1] and [Ca2] both present in western Iran, and *C. fungosa* [Ca1] and [Ca2] co-occurring at several sites in central and western Europe. The observation of genealogical distinctiveness within the pairs despite their sympatry, supports the inference that they are species level taxa rather than strongly structured populations within species (De Queiroz 2007).

Under the principle of competitive exclusion (Hardin 1960), it is expected that species persisting in sympatry will occupy distinct ecological niches. Consistent with this principle, there is a general trend for cryptic species to be more specialised in their host use than their parent morpho-species (Bickford et al. 2007). In extreme cases, supposedly generalist morpho-species of hymenopteran and dipteran parasitoids have been revealed as complexes of up to 32 cryptic species, each with a very limited and often non-overlapping host range (Smith et al. 2006, Smith et al. 2007, Smith et al. 2008). However, this trend has not been apparent in oak gall parasitoids with most pairs of cryptic species found to have at least partially overlapping host-gall ranges (Nicholls et al. 2010, Kaartinen and Roslin 2011, see Chapter 4). This is continued in the present study with the observation that *C. fungosa* [Ca1] and [Ca2] share several host gall-types at the same sites. The host gall ranges of *C. semifascia* [Ca1] and [Ca2] did not overlap, but the sampling is much too limited to infer that these species have distinct host-ranges. Further sampling would be required to establish the true extent of the host range for each cryptic species, but in the absence of host gall differentiation it would be expected that they differ in another aspect of their ecology, e.g. they may target different hosts within the same galls, may occupy different temporal niches, or may differ in their generation times (i.e. univoltine vs. bivoltine).

5.4.3. Implications for ecological study

Undetected taxonomic error is a potentially important source of bias in ecological studies (Kaartinen et al. 2010), see Chapter 4), and a principal objective of this study was to assess the accuracy of the established *Cecidostiba* morpho-species classifications so that the bias in existing ecological data might be considered, and

avoided in future studies of Palaearctic oak gall communities. The result that both the *C. semifascia* and *C. fungosa* morpho-species contain cryptic species means that existing ecological records for these species should be treated with caution, and that future studies should take measures to ensure accurate identification.

Although the morpho-species *C. semifascia* is known from 13 host gall species across much of the Western Palaearctic, it has rarely been included in comparative ecological studies. The one notable case is the study of Schönrogge and Crawley (2000), where it was recorded from asexual generation galls of *Cynips divisa* on a species of white oak at a single site in Scotland. Given that the cryptic species *C. semifascia* [Ca2] established in this study is so far only known from galls on black oaks in Iran, it seems probable that the individuals sampled by Schönrogge and Crawley (2000) all belonged to *C. semifascia* [Ca2], whose presence in the UK is confirmed. It is therefore unlikely that their inferences were at all biased by treating *C. semifascia* as a single taxon. However, the study also included the morpho-species *C. fungosa*, as have various others investigating the parasitoid communities of naturally invading oak gallwasps in western Europe (Schönrogge et al. 1995, Stone et al. 1995, Schönrogge et al. 1996, Schönrogge et al. 1999, Schönrogge and Crawley 2000, Schönrogge et al. 2000b, Schönrogge et al. 2007, Schönrogge et al. 2011). As the two new cryptic species *C. fungosa* [Ca1] and [Ca2] are both present in central and western Europe with overlapping host-gall ranges, it is not unlikely that they were both included in these studies under a single classification, leading to potential bias in the presented community parameters. Measures such as the diversity of parasitoid communities, and the number of parasitoid species per host (i.e. vulnerability) are potentially negatively biased, whereas the number of hosts per parasitoid species (i.e. generality) is potentially positively biased.

One of the benefits of the Western Palaearctic oak gall community as a model system for ecologically study is that the morphological taxonomy of the gall parasitoids is well developed, and a comprehensive morpho-species key is available (R. Askew and C. Thúroczy, unpublished). However, the discovery of cryptic species within the community, both here and in other studies (Kaartinen et al. 2010, Nicholls et al. 2010, see Chapter 4), undermines the ability of this key to return accurate species

level identifications. While it is intended that the key be modified to include *Cecidostiba* species A, and to distinguish between *C. semifascia* [Ca1] and [Ca2], this is not possible for the truly morphologically cryptic *C. fungosa* [Ca1] and [Ca2], and future studies that encounter the *C. fungosa* morpho-species may need to implement molecular methods to ensure correct identification.

It has already been described how DNA barcoding can be used to assess the accuracy of morpho-species classifications (see also Chapter 4), but its primary application is as a means of associating query individuals with established voucher taxa, based on sequence similarity criteria (Hebert et al. 2003, Ratnasingham and Hebert 2007). The accuracy of this approach is dependent on the assumptions that species level taxa are monophyletic at the barcode locus, and that inter-specific variation exceeds intra-specific variation. As these assumptions are supported by the data presented here for *C. fungosa* [Ca1] and [Ca2], and for the further 3 *Cecidostiba* species, DNA barcoding could be used in future studies to obtain accurate identifications for these species. To facilitate this, the 170 *Cecidostiba* barcode sequences generated here have been uploaded as vouchers to the Barcode Of Life Data Systems (http://www.boldsystems.org), an online database that returns an identification for a query barcode sequence if it differs by less than 1% from an established voucher (Ratnasingham and Hebert 2007). The p-distances in the barcode tree (Figure 5.2) indicate that this threshold difference of 1% is appropriate for distinguishing between the 5 *Cecidostiba* species (i.e. it is less than the inter-specific distance between all species pairs).

The cost of sequencing short specific fragments of DNA has fallen considerably in previous decades, but has recently stabilized at approximately US$ 5 for reagents and machine use but exclusive of labour (Cameron et al. 2006). While this may be easily affordable for small or moderate samples, it is likely to inhibit barcoding as a standard means of identification for larger ecological datasets that can contain thousands of individuals. In such cases an option could be to employ ‘integrative identification’, where specimens are initially identified based on their morphology, and only members of those morph-species known or suspected to contain cryptic species are then selected for DNA barcoding. To use the example of the *Cecidostiba*

species studied here, query specimens morphologically identified as *Cecidostiba* species A would be accepted as this is considered to contain no cryptic species, but any query specimens of *C. fungosa* would need to be barcoded to distinguish between the pair of cryptic species. As this approach is only practicable in systems where there has been detailed assessment of morphological taxonomy using molecular methods, further studies such as this will serve to increase the value of Western Palaearctic oak gall community as a model ecological system.

5.4.4. The utility of multi-locus monophyly for defining species

DNA barcodes offer a useful means for assessing and developing taxonomic hypotheses, but support from further independent lines of evidence is generally considered necessary to validate taxonomic inferences (Padial et al. 2010, Goldstein and DeSalle 2011). In this study, I opted to use statistical tests for taxonomic distinctiveness from observations of monophyly at further DNA loci to assess the support for hypotheses based on barcode data, resulting in the proposal of several confirmed candidate species. This approach is well grounded in population genetic theory (Baum and Shaw 1995, Hudson and Coyne 2002, Rosenberg 2007), and I consider it to be valid. There are however several important requirements and potential limitations to the approach that should be considered.

Firstly, it is not unlikely that an individual gene tree will show a conflicting topology with the underlying population or species tree (Pamilo and Nei 1988, Maddison 1997). Selection acting on a loci or at linked regions of the genome can systematically influence its rate of lineage sorting with divergent selection increasing the rate and balancing selection decreasing it, potentially to the extent where even anciently diverged species can remain paraphyletic for certain loci (Ayala and Escalante 1996). Lineage sorting may also be disrupted by introgression, where genes from one species are introduced into the gene pool of another through hybridisation (Funk and Omland 2003). Mitochondrial DNA in arthropods may be particularly affected by selective sweeps and introgression due to the influence of maternally inherited symbionts, potentially increasing error rates in identifications and inferences based on DNA barcodes (Hurst and Jiggins 2005). Further to the

systematic influences of selection and introgression, contrasting tree topologies will frequently occur simply due to the high stochastic variance of genetic processes, particularly where taxon or population divergence is relatively recent (Knowles and Carstens 2007). Consequently, taxonomic inferences drawn from observations of monophyly will be most reliable when considered across multiple independent loci, preferably ones that are selectively neutral. Such loci should also have sufficient information content to clearly differentiate between closely related species and allow for robust recovery of gene-trees (Funk and Omland 2003), criteria that may exclude many widely used exonic loci whose slow rate of sequence evolution limits their differentiation. An appropriate option is to use exon-primed intron-crossing (EPIC) loci, such as those in this study, that contain considerable intraspecific variation but are amplifiable across a range of taxa. Although not widely used, the development of primers for such loci from existing genomic and expressed sequence tag (EST) data is relatively straightforward (Lohse et al. 2011), and they are becoming increasingly available for non-model taxa (Bierne et al. 2000, Garrick et al. 2008, Tay et al. 2008, Li et al. 2010, Lohse et al. 2011).

Secondly, the time to monophyly for a sample of genes is sensitive to population demography, and expectations based on population genetic theory suggest that this time will often be substantial. For example, for a single neutral nuclear loci in species with an effective population size (N_e) of 100,000, assuming one generation a year, it would take an expected 730,000 years for there to be a high probability (95%) of observing monophyly (Hudson and Coyne 2002). Taxonomic inference based on observations of monophyly are therefore likely to be conservative (Knowles and Carstens 2007, Padial et al. 2010), and while this may be desirable for minimising type I errors, it does mean that valid but relatively young species could be overlooked (type II error). Consequently, the decision of whether to employ observation of monophyly for taxonomic inference will be case specific, and should take into account any prior knowledge of the taxa under consideration. In cases involving long standing morpho-species with no *a priori* evidence for recent speciation, such as in this study, the minimal type I error offered by observation of monophyly may be appropriately prudent.

Finally, strong population structure within the species under consideration may cause various problems. The statistical tests developed by Rosenberg (2007) and applied in this study are based on the null hypothesis of a single panmictic population, and where the probability of observed patterns of monophyly for *a priori* groupings are sufficiently low, it is inferred that the null hypothesis does not hold because the sampled individuals are drawn from multiple distinct groups (Rosenberg 2007). If there truly is complete mixing within all species present, then these groups represent distinct taxa at or above the species level. However, if mixing is incomplete (i.e. a species is strongly structured) then the multiple distinct groups may be populations within a species. Although under the modern unified concept of species there is some uncertainty about what degree of differentiation is necessary for recognising species (De Queiroz 2007), it will usually be taxonomically imprudent to recognise geographically structured populations as distinct species when isolation is the only mechanism limiting gene-flow. Fortunately, the conservative nature of taxonomic inference from observations of monophyly that make it unsuitable for distinguishing between young species, also make it resilient to the issues of structured populations. As demonstrated in the earlier example, the time taken for an individual marker to become monophyletic following lineage separation will often be large (i.e. 730,000 years with a generation time of one year and N_e of 100,000), and this time would increase with the number of loci that were considered (Hudson and Coyne 2002). Such a degree of structure (i.e. isolation without gene-flow) is not plausible for species with even a moderate effective population size, although this time would fall to 7300 years for an isolated population with an N_e of 1000, or even to 730 years if N_e was 100. The risk of making inappropriate inferences due to population structure can be minimised by appropriate geographic sampling (i.e. the geographic range of a sample should be maximised), both during initial sampling, and for any sub-sampling within candidate taxa. Where one or more candidate taxa are allopatric, it may be useful to employ additional taxonomic characters to avoid type I error.

In summary, I consider that the combination of DNA barcoding with observation of monophyly at additional molecular markers, supported by statistical tests for taxonomic distinctiveness, offers a practical and effective means for identifying and correcting species level taxonomic error. For optimality, the approach should

incorporate multiple markers with substantial information content, and EPIC loci appear to be an appropriate option. Although unlikely to be inhibitive, the risks of failing to diagnose young species, or of inappropriately diagnosing structured populations, should be considered when making taxonomic inferences.

Chapter 6 – Concluding remarks

In this thesis I have presented a series of analytical studies that adress aspects of the Climate Matching strategy and develop an accurate taxonomic framework for further ecological investigation. Each of these studies includes a comprehensive discussion, but I conclude in this final chapter with a brief discussion of how Climate Matching might be practiced in light of these results. Finally, I suggest what I consider to be valuable avenues for future research.

6.1. How to match or not to match?

The use of Climate Matching to guide the selection of non-local provenances for planting has been advocated as a 'no-regret' option for promoting the adaptation of UK forests to expected changes in climate (Broadmeadow and Ray 2005). However, in the opening chapter of this thesis I identified two points that could limit the practical value of Climate Matching: firstly, that if tree provenances are not locally adapted to the climatic factors used to match sites then Climate Matching would be ineffective in promoting adaptation and could reduce population fitness; and secondly, that introduced provenances may influence associated organisms potentially impacting on forest biodiversity. In Chapters 2 and 3 I explored these issues empirically, using a model system of *Quercus petraea* and its associated community of herbivorous gallwasps. In light of the results of these studies, what can be said about how Climate Matching should or should not be practiced?

The results presented in Chapter 2 for *Q. petraea* suggest that the climatic factors used in Climate Matching analysis (i.e. temperature and precipitation) are involved in the local adaptation of particular phenotypic traits. Matched provenances are therefore likely to possess traits that would be of adaptive advantage under future climatic conditions, providing some justification for Climate Matching of this species. However, the study also highlighted that additional factors are likely to be involved in adaptation, and that while matched provenances may be well adapted to future climates at a planting site, they will not necessarily be adapted to the present climate. These present potentially major obstacles for Climate Matching, as there is

little to be gained from planting trees that are perfectly suited to the climate of the 2080s, but that die or perform poorly before that time due to maladaptation to earlier climates or to alternate factors such as pathogens or soil properties.

As discussed in Chapter 2, a possible mitigation strategy would be to adjust the Climate Matching analysis to also consider more geographical gradients associated with environmental variation such as latitude, longitude and altitude. Minimising the differences in these variables between planting sites and matched provenance sites could reduce the level of maladaptation to alternative aspects of the plantation environment. A possible further strategy would be to buffer seedling of matched provenances from unfavourable initial conditions by planting alongside shrubs or widely spaced mature trees that act as nurse plants (Castro et al. 2004). Alternatively, matched seed or seedlings could be planted at very high initial stocking density, allowing for high mortality and promoting selection for genotypes that are most adapted (or least maladapted) to the range of biotic and abiotic influences presented at the planting site. This latter strategy is also proposed for promoting the adaptation of native populations (Hubert and Cottrell 2007, Savolainen er al. 2007).

While these strategies may improve its effectiveness, it seems unlikely that Climate Matching will directly provide the 'holy grail' of adaptive forest management – i.e. a tree population that performs well at a planting site throughout its long lifetime and is resilient to any changes in climate. Such a population would presumably need to combine aspects of the native population such as adaptation to local photoperiod, soils, pests, and pathogens, with aspects of climate matched populations such as tolerance to summer drought and winter water logging. An alternative to planting Climate Matched seed or seedlings might therefore be to introduce Climate Matched pollen for crossing with trees of local provenance. At least some of the resulting offspring would hopefully exhibit traits well suited to both present and future environments at the planting site. Subsequent natural crossing and regeneration from these individuals, perhaps encouraged by regular gap creation and natural regeneration (Hubert and Cottrell 2007), could promote the spread of adaptive non-native traits (e.g. for improved tolerance of drought or water logging) in the local gene-pool.

In relation to the issue of associated biodiversity, the results presented in Chapter 3 suggest that while tree provenance can strongly influence the abundance and structure of an oak gallwasp community, Climate Matching is unlikely to have a severely negative impact. Tree phenological traits such as the timing of spring bud-burst appeared to be particularly influential in tree-gallwasp interactions, but gallwasps did not seem to be tightly synchronised with trees of local provenance and were generally found in greatest abundance on trees with non-local phenology. However, the dynamics of associations were potentially complex, with alternate generations of single species appearing in greater abundance on trees with different phenologies. The risk of negative impacts on this community could be minimised by ensuring that trees of matched provenance are well mixed with trees of local provenance, allowing for migration of alternate generations of gallwasps between trees of different provenance and phenotypes. Such mixed planting would also be likely to minimise the risk of more general biodiversity loss, ensuring that habitat remained available for those species less well able to interact with trees of non-local provenance.

The observation that gallwasps were able to interact with trees with non-local phenologies bodes well for Climate Matching and for sustainable forest management in general, suggesting that some components of local forest biodiversity will be resilient to the effects of introduced tree provenances, and to any phenological shifts in local tree populations. However, gallwasps are a very small and perhaps unrepresentative component of forest biodiversity, having a particularly intricate relationship with their hosts and being capable of manipulating host resources (Schönrogge et al. 2000a, Harper et al. 2004). It is yet unclear whether other forest organisms will be as resilient to Climate Matching, and precautionary steps to minimise the risk of biodiversity loss should be taken (e.g. mixed planting of local and matched provenances).

In summary, I propose the following recommendations for Climate Matching practitioners: (i) Climate Matching should only be applied to tree species with established evidence for local adaptation to temperature or precipitation gradients. (ii) The risk of maladaptation and consequent poor establishment of matched

provenances at planting sites should be considered and potentially mitigated by including additional environmental gradients in the Climate Matching analysis, exposing matched seedlings to high selection pressure at the planting site, or buffering young seedlings from initial environmental conditions. (iii) The option of introducing Climate Matched pollen rather than seed should be explored. (iv) The potential for negative impacts on associated biodiversity should be considered and the risks minimised by ensuring that trees of matched provenance are well mixed with trees of local provenance.

6.2. Future research

6.2.1. Population response functions

This thesis has focused on a single trial containing a wide variety of provenances. While analysis of single trial data can be informative about patterns of adaptation, it is not well suited to investigating the potentially complex genotype-by-environment interactions that determine how tree populations will perform across a range of environments. Such investigation requires data from reciprocal transplant or treatment experiments, where the same provenances are monitored at various planting sites or are subject to various experimental treatment conditions (Aitken 2004). In North America where provenance research has been extensive, analyses of multi-trial or multi-treatment data have been used to model the performance of individual populations across broad climatic and geographical ranges (Rehfeldt et al. 1999, Rehfeldt et al. 2002, St Clair et al. 2005, Wang et al. 2006, Bower and Aitken 2008). When combined with climate change projections, these models allow for quantified prediction of the degree of maladaptation of populations in their current location, and can guide the identification of populations that will perform well under future climates (Wang et al. 2006, St Clair and Howe 2007, Aitken et al. 2008, Wang et al. 2010).

In Europe, broad studies of population response functions are largely lacking, despite the availability of well reciprocated provenance trials (but see Matyas 1994). Their development from analysis of multi-trial data would be a valuable avenue for further

research, potentially illustrating the relative importance of various aspects of climate and allowing for better prediction of the effects of climate change on tree health and productivity.

6.2.2. Effects of host-tree provenance on multi-trophic communities and alternate herbivore guilds

The results presented in Chapter 3 suggest that gallwasp communities are unlikely to be severely negatively effected by Climate Matching. However, it remains to be seen whether other components of forest biodiversity will be similarly resilient. Oak gallwasps are nested within closed multi-trophic communities that include inquiline Cynipids and hymenopteran parasitoids. In Chapter 4, a taxonomic framework for these communities was developed based on DNA barcodes, but further investigation with multiple additional molecular markers is still required to determine the status of particular taxa. Following the completion of this taxonomic investigation, analysis of this data will reveal how the effects of Climate Matching under today's climates may apply throughout a trophically linked community. Further to this, empirical study at the Petite Charnie trial could be extended to additional components of diversity such as other herbivore guilds, entophytic fungi, and soil microbes. Such research would be of value, not only in terms of differentialting the implications of Climate Matching for different feeding guilds, but also in furthering a general understanding of how variation within foundation tree species may structure forest ecosystems.

6.2.3. Provenance deme formation

The adaptive deme hypothesis predicts that short-lived herbivores with long lived hosts may become locally adapted to the phenotypes of individual host plants (Edmunds and Alstad 1978). The hypothesis is supported by experimental evidence from various plant-herbivore systems (reviewed in van Zandt and Mopper 1998, Mopper 2005), including leaf-mining, leaf-folding, and gall-forming herbivores on oaks (Mopper et al. 2000, Egan and Ott 2007, Tack and Roslin 2010). Such fine scale adaptation may be an important force for maintaining adaptive genetic diversity

within herbivore populations, increasing their ability to withstand disturbance (Mopper 1996).

Although the adaptive deme hypothesis is usually considered in terms of adaptation to individual trees, it may extend to sub-sets of trees within a population that consistently exhibit particular phenotypes. Therefore, as discussed in Chapter 3, a possible result of Climate Matching is that adaptive demes will form within herbivore populations in response to the consistently differing phenotypes exhibited by trees of matched and local provenance. This prediction could be experimentally investigated at the Petite Charnie provenance trials by introducing seedlings of known provenance into established parcelles (i.e. blocks of 24 trees of a particular provenance that have been in place for approximately 20 years). If adaptive provenance demes have formed, it would be expected that herbivore abundance would be greater when seedlings of a particular provenance are introduced to parcelles of the same provenance, relative to when introduced to parcelles of different provenance. As gene-flow between herbivores from different host plants is expected to be a major factor in determining if and where deme formation will occur (Tack and Roslin 2010), it may also be of interest to apply this approach in trials where provenances have been planted in blocks of various sizes, so that spatial influences can be considered. Such research would not only further understanding of the implications of Climate Matching, but would be of general value in assessing the scale at which local adaptation may apply in field situations.

Bibliography

Ács, Z., R. J. Challis, P. Bihari, M. Blaxter, A. Hayward, G. Melika, G. Csoka, Z. Penzes, J. Pujade-Villar, J.-L. Nieves-Aldrey, K. Schonrogge, and G. N. Stone. 2010. Phylogeny and DNA barcoding of inquiline oak gallwasps (Hymenoptera: Cynipidae) of the Western Palaearctic. Mol Phylogenet Evol **55**:210-225.

Ács, Z., G. Melika, Z. Penzes, J. Pujade-Villar, and G. N. Stone. 2007. The phylogenetic relationships between Dryocosmus, Chilaspis and allied genera of oak gallwasps (Hymenoptera, Cynipidae : Cynipini). Systematic Entomology **32**:70-80.

Aitken, S. N. 2004. Genecology and Adaptation of Forest trees. Pages 197-2004 Encyclopedia of Forest Sciences Article 86. Elsevier.

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

Akaike, H. 1974. A new look at statistical model identification. Ieee Transactions on Automatic Control **AC19**:716-723.

Alberto, F., L. Bouffier, J. M. Louvet, J. B. Lamy, S. Delzon, and A. Kremer. 2011. Adaptive responses for seed and leaf phenology in natural populations of sessile oak along an altitudinal gradient. Journal of Evolutionary Biology **24**:1442-1454.

Ambrus, B. 1974. Cynipid galls.*in* S. Gusztáv, editor. Fauna Hungariae. Hyngary: Academic Press, Budapest.

Askew, R. R. 1961a. 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, R. R. 1961b. Some biological notes on the Pteromalid (Hym. Chalcidoidea) genera *Caenacis* Förster, *Cecidostiba* Thomson, and *Hobbya* Delucchi, with descriptions of two new species. Entomophaga **6**:57-67.

Askew, R. R. 1961c. A study of the biology of the species of the genus *Mesopolobus* Westwood (Hymenoptera: Pteromalidae) associated with cynipid galls on oak. Transactions of the Royal Entomological Society of London **113**:155-173.

Askew, R. R. 1962. The Distribution of Galls of Neuroterus (Hym, Cynipidae) on Oak. Journal of Animal Ecology **31**:439-455.

Askew, R. R. 1980. The Diversity of Insect Communities in Leaf-Mines and Plant Galls. Journal of Animal Ecology **49**:817-829.

Askew, R. R. 1984. The biology of gallwasps. Pages 223-271 *in* T. N. Anathakrishnan, editor. The biology of galling insects. Oxford and IBH publishing CO., New Dehli.

Askew, R. R., O. Plantard, J. F. Gomez, M. Hernandez Nieves, and J. L. Nieves-Aldrey. 2006. Catalogue of parasitoids and inquilines in galls of Aylacini, Diplolepidini and Pediaspidini (Hym., Cynipidae) in the West Palaearctic. Zootaxa:1-60.

Ayala, F. J. and A. A. Escalante. 1996. The evolution of human populations: A molecular perspective. Molecular Phylogenetics and Evolution **5**:188-201.

Baayen, R. H. 2011. languageR: Data sets and functions with "Analyzing Linguistic Data: A practical introduction to statistics". R package version 1.2. .

Bacilieri, R., A. Ducousso, and A. Kremer. 1995. Genetic, Morphological, Ecological and Phenological Differentiation between Quercus-Petraea (Matt) Liebl and Quercus-Robur L in a Mixed Stand of Northwest of France. Silvae Genetica **44**:1-10.

Bailey, J. K., R. Deckert, J. A. Schweitzer, B. J. Rehill, R. L. Lindroth, C. Gehring, and T. G. Whitham. 2005. Host plant genetics affect hidden ecological players: links among Populus, condensed tannins, and fungal endophyte infection. Canadian Journal of Botany-Revue Canadienne De Botanique **83**:356-361.

Bailey, J. K. and T. G. Whitham. 2003. Interactions among elk, aspen, galling sawflies and insectivorous birds. Oikos **101**:127-134.

Bailey, R., K. Schonrogge, J. M. Cook, G. Melika, G. Csoka, C. Thuroczy, and G. N. Stone. 2009. Host Niches and Defensive Extended Phenotypes Structure Parasitoid Wasp Communities. Plos Biology **7**.

Baldi, A. 2008. Habitat heterogeneity overrides the species-area relationship. Journal of Biogeography **35**:675-681.

Bangert, R. K., R. J. Turek, B. Rehill, G. M. Wimp, J. A. Schweitzer, G. J. Allan, J. K. Bailey, G. D. Martinsen, P. Keim, R. L. Lindroth, and T. G. Whitham. 2006. A genetic similarity rule determines arthropod community structure. Molecular Ecology **15**:1379-1391.

Barsoum, N., J. Cottrell, J. Hubert, S. Evans, D. Ray, and M. Petr. 2009. Concept Note - Forest adaptation to climate change: National Forest planting trial.

Barton, K. 2011. MuMIn: Multi-model inference. R package version 1.0.0.

Bates, D., M. Maechler, and B. Bolker. 2011. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-39.

Baum, D. and K. L. Shaw. 1995. Genealogical perspectives on the species problem.*in* P. C. S. Hoch, A. G., editor. Experimental and molecular aproaches to plant biosystematics Missouri Botanical Garden, St. Louis.

Bazin, E., S. Glemin, and N. Galtier. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science **312**:570-572.

Bersier, L. F., C. Banasek-Richter, and M. F. Cattin. 2002. Quantitative descriptors of food-web matrices. Ecology **83**:2394-2407.

Bickford, D., D. J. Lohman, N. S. Sodhi, P. K. L. Ng, R. Meier, K. Winker, K. K. Ingram, and I. Das. 2007. Cryptic species as a window on diversity and conservation. Trends in Ecology & Evolution **22**:148-155.

Bierne, N., A. Lehnert, E. Bedier, F. Bonhomme, and S. S. Moore. 2000. Screening for intron-length polymorphisms in penaeid shrimps using exon-primed intron-crossing (EPIC)-PCR. Molecular Ecology **9**:233-235.

Blaxter, M., J. Mann, T. Chapman, F. Thomas, C. Whitton, R. Floyd, and E. Abebe. 2005. Defining operational taxonomic units using DNA barcode data. Philosophical Transactions of the Royal Society B-Biological Sciences **360**:1935-1943.

Bolker, B. M., M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J. S. S. White. 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution **24**:127-135.

Bolte, A., C. Ammer, M. Lof, P. Madsen, G. J. Nabuurs, P. Schall, P. Spathelf, and J. Rock. 2009. Adaptive forest management in central Europe: Climate change impacts, strategies and integrative concept. Scandinavian Journal of Forest Research **24**:473-482.

Bower, A. D. and S. N. Aitken. 2008. Ecological genetics and seed transfer guidelines for Pinus albicaulis (Pinaceae). American Journal of Botany **95**:66-76.

Brewer, S., R. Cheddadi, J. L. de Beaulieu, and M. Reille. 2002. The spread of deciduous Quercus throughout Europe since the last glacial period. Forest Ecology and Management **156**:27-48.

Broadmeadow, M. and D. Ray. 2005. Climate Change and British Woodland. Forestry Commission Information Note 69 Forestry Commission, Edinburgh.

Broadmeadow, M., D. Ray, and L. Poulsom. 2003. Climate change and British woodland: what does the future hold? Forest Research Annual Report and Accounts 2002-2003.

Broadmeadow, M. S. J., D. Ray, and C. J. A. Samuel. 2005. Climate change and the future for broadleaved tree species in Britain. Forestry **78**:145-161.

Buhr, H. 1965. Bestimmungstabellen der Gallen (Zoo- und Phytocecidien) an Pflanzen Mittel- und Nordeuropas. Fischer Verlag,, Jena.

Bukovinszky, T., E. H. Poelman, R. Gols, G. Prekatsakis, L. E. M. Vet, J. A. Harvey, and M. Dicke. 2009. Consequences of constitutive and induced variation in plant nutritional quality for immune defence of a herbivore against parasitism. Oecologia **160**:299-308.

Bukovinszky, T., F. J. F. van Veen, Y. Jongema, and M. Dicke. 2008. Direct and indirect effects of resource quality on food web structure. Science **319**:804-807.

Calcagno, V. and C. de Mazancourt. 2010. glmulti: An R Package for Easy Automated Model Selection with (Generalized) Linear Models. Journal of Statistical Software **34**:1-29.

Cameron, S., D. Rubinoff, and K. Will. 2006. Who will actually use DNA barcoding and what will it cost? Systematic Biology **55**:844-847.

Castro, J., R. Zamora, J. A. Hodar, and J. M. Gomez. 2004. Seedling establishment of a boreal tree species (Pinus sylvestris) at its southernmost distribution limit: consequences of being in a marginal Mediterranean habitat. Journal of Ecology **92**:266-277.

Challis, R. J., S. Mutun, J. L. Nieves-Aldrey, S. Preuss, A. Rokas, A. Aebi, E. Sadeghi, M. Tavakoli, and G. N. Stone. 2007. Longitudinal range expansion and cryptic eastern species in the western Palaearctic oak gallwasp, Andricus coriarius. Molecular Ecology **16**:2103-2114.

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

Cornelissen, T., G. W. Fernandes, and J. Vasconcellos-Neto. 2008. Size does matter: variation in herbivory between and within plants and the plant vigor hypothesis. Oikos **117**:1121-1130.

Crawford, N. G. 2010. smogd: software for the measurement of genetic diversity. Molecular Ecology Resources **10**:556-557.

Crawley, M. J. 2007. The R Book. John Wiley & Sons Ltd, Chichester, England.

Crawley, M. J. and M. Akhteruzzaman. 1988. Individual Variation in the Phenology of Oak Trees and Its Consequences for Herbivorous Insects. Functional Ecology **2**:409-415.

Csóka, G. 1998. A Magyarországon honos tölgyek herbivor rovaregyüttese [Herbivore insects of oaks native in Hungary]. . Erdészeti Kutatások **88**:311-318.

Csóka, G., G. N. Stone, and G. Melika. 2005. Biology, ecology and evolution of gall-inducing Cynipinae. Pages 573-642 *in* A. Raman, C. W. Schaefer, and T. M. Withers, editors. Biology. Ecology and Evolution of Gall0inducing Arthropods. Science Publishers Inc, Enfield, New Hampshire.

De Queiroz, K. 2005. Ernst Mayr and the modern concept of species. Proceedings of the National Academy of Sciences of the United States of America **102**:6600-6607.

De Queiroz, K. 2007. Species concepts and species delimitation. Systematic Biology **56**:879-886.

Deans, J. D. and F. J. Harvey. 1995. Phenologies of sixteen European provenances of sessile oak growing in Scotland. Forestry **68**:265-274.

Debruyn, L. 1995. Plant stress and larval performance of a dipterous gall former. Oecologia **101**:461-466.

DEFRA and Forestry Commission. 2005. Keepers of time: a statement of policy for Englands ancient and native woodland.

Dickson, L. L. and T. G. Whitham. 1996. Genetically based plant resistance traits affect arthropods, fungi, and birds. Oecologia **106**:400-406.

Dormann, C. F., C. E. Grueber, and J. Fründ. 2008. Introducing the bipartite package: Analysing Ecological Networks. R News **8**:8-11.

Ducousso, A., J. P. Guyon, and A. Krémer. 1996. Latitudinal and altitudinal variation of bud burst in western populations of sessile oak (Quercus petraea (Matt) Liebl). Annales des Sciences Forestieres **53**:775-782.

Dungey, H. S., B. M. Potts, T. G. Whitham, and H. F. Li. 2000. Plant genetics affects arthropod community richness and composition: Evidence from a synthetic eucalypt hybrid population. Evolution **54**:1938-1946.

EC. 2005. Communication from the commission to the council and the European Parliament.

Reporting on the implementation of the EU forestry strategy.

Edmunds, G. F. and D. N. Alstad. 1978. Coevolution in Insect Herbivores and Conifers. Science **199**:941-945.

Edwards, S. V. and M. Kot. 1995. Comparative methods at the species level: Geographic variation in morphology and group size in grey-crowned babblers (Pomatostomus temporalis). Evolution **49**:1134-1146.

Egan, S. P. and J. R. Ott. 2007. Host plant quality and local adaptation determine the distribution of a gall-forming herbivore. Ecology **88**:2868-2879.

Folmer, O., M. Black, W. Hoeh, R. Lutz, and R. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology **3**:294-299.

Forest Research, National Forest Company, Forestry Commission, and A. Industries. 2010. Climate Change and Tree Provenance.

Forestry Commission. 2003. National Inventory of Woodland and Trees: Great Britain. Inventory Report. Forestry Commission, Edinburgh.

Forestry Commission. 2004. The UK Forestry Standard: The Governments Approach to Sustainable Forestry. Forestry Commission, Edinburgh.

Forsythe, W. C., E. J. Rykiel, R. S. Stahl, H. I. Wu, and R. M. Schoolfield. 1995. A Model Comparison for Daylength as a Function of Latitude and Day of Year. Ecological Modelling **80**:87-95.

Fritz, R. S. and P. W. Price. 1988. Genetic-Variation among Plants and Insect Community Structure - Willows and Sawflies. Ecology **69**:845-856.

Funk, D. J. and K. E. Omland. 2003. Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annual Review of Ecology Evolution and Systematics **34**:397-423.

Garrick, R. C., C. A. Meadows, A. N. Nicolas, J. D. Nason, and R. J. Dyer. 2008. A set of polymorphic nuclear intron markers for conservation genetics and phylogeography of Euphorbia species (Pedilanthus clade). Conservation Genetics **9**:1673-1676.

Gelman, A. 2008. Scaling regression inputs by dividing by two standard deviations. Statistics in Medicine **27**:2865-2873.

Gelman, A., Y.-S. Su, M. Yajima, J. Hill, M. G. Pittau, J. Kerman, and T. Zheng. 2011. arm: Data Analysis Using Regression and Multilevel/Hierarchical Models. R package version 1.4-13.

Gene Codes Corporation. 2009. Sequencher® version 4.9 sequence analysis software. Ann Arbor, MI USA http://www.genecodes.com.

Gibbs, M. J., J. S. Armstrong, and A. J. Gibbs. 2000. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics **16**:573-582.

Giorgi, F. and E. Coppola. 2009. Projections of twenty-first century climate over Europe. Pages 29-46 *in* C. Boutron, editor. Erca: From the Human Dimensions of Global Environmental Change to the Observation of the Earth from Space, Vol 8.

Godfray, H. C. J. 1994. Parasitoids: Behavioural and Evolutionary Ecology. Princeton University Press, Princeton New Jersey.

Goldstein, P. Z. and R. DeSalle. 2011. Integrating DNA barcode data and taxonomic practice: Determination, discovery, and description. Bioessays **33**:135-147.

Goudet, J. 1995. FSTAT (Version 1.2): A computer program to calculate F-statistics. Journal of Heredity **86**:485-486.

Grueber, C. E., S. Nakagawa, R. J. Laws, and I. G. Jamieson. 2011. Multimodel inference in ecology and evolution: challenges and solutions. Journal of Evolutionary Biology **24**:699-711.

Hamilton, A. J., Y. Basset, K. K. Benke, P. S. Grimbacher, S. E. Miller, V. Novotny, G. A. Samuelson, N. E. Stork, G. D. Weiblen, and J. D. L. Yen. 2010. Quantifying Uncertainty in Estimation of Tropical Arthropod Species Richness. American Naturalist **176**:90-95.

Hardin, G. 1960. The competitive exclusion principle. Science **131**:1292-1297.

Harper, L. J., K. Schönrogge, K. Y. Lim, P. Francis, and C. P. Lichtenstein. 2004. Cynipid galls: insect-induced modifications of plant development create novel plant organs. Plant Cell and Environment **27**:327-335.

Hassell, M. P. and R. M. May. 1986. Generalist and specialist natural enemies in insect predator prey interactions. Journal of Animal Ecology **55**:923-940.

Hayward, A. and G. N. Stone. 2006. Comparative phylogeography across two trophic levels: the oak gall wasp Andricus kollari and its chalcid parasitoid Megastigmus stigmatizans. Molecular Ecology **15**:479-489.

Hebert, P. D. N., A. Cywinska, S. L. Ball, and J. R. DeWaard. 2003. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London Series B-Biological Sciences **270**:313-321.

Hebert, R., S. Samuel, and S. Patterson. 1999. Using local stock for planting native trees and shrubs. Forestry Commission practice note 8. . Forestry Commission, Edinburgh.

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

Henneman, M. L. and J. Memmott. 2001. Infiltration of a Hawaiian community by introduced biological control agents. Science **293**:1314-1316.

Hickerson, M. J., C. P. Meyer, and C. Moritz. 2006. DNA barcoding will often fail to discover new animal species over broad parameter space. Systematic Biology **55**:729-739.

Hijmans, R. J., S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology **25**:1965-1978.

Howe, G. T., S. N. Aitken, D. B. Neale, K. D. Jermstad, N. C. Wheeler, and T. H. H. Chen. 2003. From genotype to phenotype: unraveling the complexities of cold adaptation in forest trees. Canadian Journal of Botany-Revue Canadienne De Botanique **81**:1247-1266.

Hubert, J. 2005. Selecting the Right Provenance of Oak for Planting in Britain. Forestry Commission Information Note 77. Forestry Commission, Edinburgh.

Hubert, J. and J. Cottrell. 2007. The role of forest genetic resources in helping British Forests respond to climate change. Forestry Commission Information Note 86. Forestry Commission, Edinburgh.

Hubert, J. and E. Cundall. 2006. Choosing Provenance in Broadleaved Trees. Forestry Commission Information Note 82. Forestry Commission, Edinburgh.

Hudson, R. R. and J. A. Coyne. 2002. Mathematical consequences of the genealogical species concept. Evolution **56**:1557-1565.

Hulme, M., Jenkins, G. J., Lu, X., Turnpenny, J. R., Mitchell, T. D., Jones, R. G., Lowe, J., Murphy, J. M., Hassell, D., Boorman, P., McDonald, R. and Hill, S. 2002. Climate Change Scenarios for the United Kingdom: The UKCIP02 Scientific Report Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia Norwich, UK. .

Hunter, M. D. and P. W. Price. 1992. Playing chutes and ladders - heterogeneity and the relative roles of bottom-up and top-down forces in natural communities. Ecology **73**:724-732.

Hunter, M. D., G. C. Varley, and G. R. Gradwell. 1997. Estimating the relative roles of top-down and bottom-up forces on insect herbivore populations: A classic study revisited. Proceedings of the National Academy of Sciences of the United States of America **94**:9176-9181.

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

Huxley, J. 1955. Morphism and evolution. Heredity **9**:1-52.

IPCC. 2007. Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, Pachauri, R.K and Reisinger, A. (eds.)]. Page 104 pp. IPCC,, Geneva, Switzerland,.

Ito, M. and N. Hijii. 2001. Effect of shoot size and phenological variation of host plants on the spatial patterns of cynipid galls. Journal of Forest Research **6**:147-151.

Jenkins, G. J., J. M. Murphy, D. M. H. Sexton, J. A. Lowe, P. Jones, and C. G. Kilsby. 2009. UK Climate Projections: Briefing report. Met Office Hadley Centre,, Exeter, UK.

Jenkins, G. J., M. C. Perry, and M. J. Prior. 2008. The climate of the United Kingdom and recent trends. Met Office Hadley Centre, Exeter, UK.

Jensen, J. S. 2000. Provenance variation in phenotypic traits in Quercus robur and Quercus petraea in Danish provenance trials. Scandinavian Journal of Forest Research **15**:297-308.

Jensen, J. S. and J. K. Hansen. 2008. Geographical variation in phenology of *Quercus petraea* (Matt.) Liebl and *Quercus robur* L. oak grown in a greenhouse. Scandinavian Journal of Forest Research **23**:179 - 188.

Johnson, M. T. J. 2008. Bottom-up effects of plant genotype on aphids, ants, and predators. Ecology **89**:145-154.

Jones, M., A. Ghoorah, and M. Blaxter. 2011a. jMOTU and Taxonerator: Turning DNA Barcode Sequences into Annotated Operational Taxonomic Units. Plos One **6**.

Jones, T. S., E. Allan, S. A. Haerri, J. Krauss, C. B. Müller, and F. J. F. van Veen. 2011b. Effects of genetic diversity of grass on insect species diversity at higher trophic levels are not due to cascading diversity effects. Oikos **120**:1031-1036.

Jost, L. 2008. G(ST) and its relatives do not measure differentiation. Molecular Ecology **17**:4015-4026.

Kaartinen, R. and T. Roslin. 2011. Shrinking by numbers: landscape context affects the species composition but not the quantitative structure of local food webs. Journal of Animal Ecology **80**:622-631.

Kaartinen, R., G. N. Stone, J. Hearn, K. Lohse, and T. Roslin. 2010. Revealing secret liaisons: DNA barcoding changes our understanding of food webs. Ecological Entomology **35**:623-638.

Kaltz, O., S. Gandon, Y. Michalakis, and J. A. Shykoff. 1999. Local maladaptation in the anther-smut fungus Microbotryum violaceum to its host plant Silene latifolia: Evidence from a cross-inoculation experiment. Evolution **53**:395-407.

Kamijo, K. 1981. Pteromalid wasps (Hymenoptera) reared from cynipid galls on oak and chestnut in Japan with descriptions of 4 new species. Kontyu **49**:272-282.

Kass, R. E. and A. E. Raftery. 1995. Bayes Factors. Journal of the American Statistical Association **90**:773-795.

Katoh, K. and H. Toh. 2008. Recent developments in the MAFFT multiple sequence alignment program. Briefings in Bioinformatics **9**:286-298.

Kawecki, T. J. and D. Ebert. 2004. Conceptual issues in local adaptation. Ecology Letters **7**:1225-1241.

Kennedy, C. E. J. and T. R. E. Southwood. 1984. The number of species of insects associated with British trees - a re-analysis. Journal of Animal Ecology **53**:455-478.

Kleinschmit, J. 1993. Intraspecific variation of growth and adaptive traits in European oak species. Annales des Sciences Forestieres **50**:166-185.

Knowles, L. L. and B. C. Carstens. 2007. Delimiting species without monophyletic gene trees. Systematic Biology **56**:887-895.

Konig, A. O. 2005. Provenance research: evaluating the spatial pattern of genetic variation.*in* T. T. Geburek, J., editor. Conservation and management of forest genetic resources in Europe. Arbora Publishers, Zvolen.

Kopelke, J. P., J. Amendt, and K. Schonrogge. 2003. Patterns of interspecific associations of stem gallers on willows. Diversity and Distributions **9**:443-453.

Koricheva, J., S. Larsson, and E. Haukioja. 1998. Insect performance on experimentally stressed woody plants: A meta-analysis. Pages 195-216 *in* M. R. Berenbaum, editor. Annual Review of Entomology.

Kremer, A., J. Kleinschmit, J. Cottrell, E. P. Cundall, J. D. Deans, A. Ducousso, A. O. Konig, A. J. Lowe, R. C. Munro, R. J. Petit, and B. R. Stephan. 2002. Is there a correlation between chloroplastic and nuclear divergence, or what are the roles of history and selection on genetic diversity in European oaks? Forest Ecology and Management **156**:75-87.

Kremer, A., A. Zanetto, and A. Ducousso. 1997. Multilocus and multitrait measures of differentiation for gene markers and phenotypic traits. Genetics **145**:1229-1241.

Larsson, S. 1989. Stressful times for the plant stress - insect performance hypothesis. Oikos **56**:277-283.

Le Corre, V., S. DumolinLapegue, and A. Kremer. 1997. Genetic variation at allozyme and RAPD loci in sessile oak Quercus petraea (Matt) Liebl: The role of history and geography. Molecular Ecology **6**:519-529.

LeRoy, C. J., T. G. Whitham, P. Keim, and J. C. Marks. 2006. Plant genes link forests and streams. Ecology **87**:255-261.

Levene, H. 1953. Genetic equilibrium when more than one ecological niche is available. American Naturalist **87**:331-333.

Levins, R. and R. MacArthur. 1966. Maintenance of genetic polymorphism in a spatially heterogenous environment - variations on a theme by Howard Levene. American Naturalist **100**:585-&.

Lewontin, R. C. and J. Krakauer. 1973. Distribution of gene frequency as a test of theory of selective neutrality of polymorphisms. Genetics **74**:175-195.

Li, C. H., J. J. M. Riethoven, and L. B. Ma. 2010. Exon-primed intron-crossing (EPIC) markers for non-model teleost fishes. Bmc Evolutionary Biology **10**.

Lindner, M., M. Maroschek, S. Netherer, A. Kremer, A. Barbati, J. Garcia-Gonzalo, R. Seidl, S. Delzon, P. Corona, M. Kolstrom, M. J. Lexer, and M. Marchetti. 2010. Climate change impacts, adaptive capacity, and vulnerability of European forest ecosystems. Forest Ecology and Management **259**:698-709.

Linhart, Y. B. and M. C. Grant. 1996. Evolutionary significance of local genetic differentiation in plants. Pages 237-277 *in* D. G. Fautin, editor. Annual Review of Ecology and Systematics.

Lohse, K. 2009. Can mtDNA Barcodes Be Used to Delimit Species? A Response to Pons et al. (2006). Systematic Biology **58**:439-441.

Lohse, K., B. Sharanowski, M. Blaxter, J. A. Nicholls, and G. N. Stone. 2011. Developing EPIC markers for chalcidoid Hymenoptera from EST and genomic data. Molecular Ecology Resources **11**:521-529.

Lukhtanov, V. A., A. Sourakov, E. V. Zakharov, and P. D. N. Hebert. 2009. DNA barcoding Central Asian butterflies: increasing geographical dimension does not significantly reduce the success of species identification. Molecular Ecology Resources **9**:1302-1310.

Macel, M., C. S. Lawson, S. R. Mortimer, M. Smilauerova, A. Bischoff, L. Cremieux, J. Dolezal, A. R. Edwards, V. Lanta, T. M. Bezemer, W. H. van der Putten, J. M. Igual, C. Rodriguez-Barrueco, H. Muller-Scharer, and T. Steinger. 2007. Climate vs. soil factors in local adaptation of two common plant species. Ecology **88**:424-433.

Maddison, W. P. 1997. Gene trees in species trees. Systematic Biology **46**:523-536.

Martin, D. P., P. Lemey, M. Lott, V. Moulton, D. Posada, and P. Lefeuvre. 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics **26**:2462-2463.

Martin, D. P., D. Posada, K. A. Crandall, and C. Williamson. 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. Aids Research and Human Retroviruses **21**:98-102.

Martinez, N. D. and J. H. Lawton. 1995. Scale and food-web structure - from local to global. Oikos **73**:148-154.

Matyas, C. 1994. Modeling climate-change effects with provenance test data. Tree Physiology **14**:797-804.

Maynard Smith, J. 1992. Analyzing the Mosaic Structure of Genes. Journal of Molecular Evolution **34**:126-129.

McCauley, D. E. and W. F. Eanes. 1987. Hierarchical population-structure analysis of the milkweed beetle, *Tetrapoes tetraophthalmus* (Forster). Heredity **58**:193-201.

McKay, J. K., C. E. Christian, S. Harrison, and K. J. Rice. 2005. "How local is local?" - A review of practical and conceptual issues in the genetics of restoration. Restoration Ecology **13**:432-440.

Memmott, J., N. D. Martinez, and J. E. Cohen. 2000. Predators, parasitoids and pathogens: species richness, trophic generality and body sizes in a natural food web. Journal of Animal Ecology **69**:1-15.

Meyer, C. P. and G. Paulay. 2005. DNA barcoding: Error rates based on comprehensive sampling. Plos Biology **3**:2229-2238.

Millar, C. I., N. L. Stephenson, and S. L. Stephens. 2007. Climate change and forests of the future: Managing in the face of uncertainty. Ecological Applications **17**:2145-2151.

Mimura, M. and S. N. Aitken. 2007. Adaptive gradients and isolation-by-distance with postglacial migration in Picea sitchensis. Heredity **99**:224-232.

Monaghan, M. T., M. Balke, T. R. Gregory, and A. P. Vogler. 2005. DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. Philosophical Transactions of the Royal Society B-Biological Sciences **360**:1925-1933.

Mopper, S. 1996. Adaptive genetic structure in phytophagous insect populations. Trends in Ecology & Evolution **11**:235-238.

Mopper, S. 2005. Phenology - how time creates spatial structure in endophagous insect populations. Annales Zoologici Fennici **42**:327-333.

Mopper, S. and D. Simberloff. 1995. Differential Herbivory in an Oak Population - the Role of Plant Phenology and Insect Performance. Ecology **76**:1233-1241.

Mopper, S., P. Stiling, K. Landau, D. Simberloff, and P. van Zandt. 2000. Spatiotemporal variation in leafminer population structure and adaptation to individual oak trees. Ecology **81**:1577-1587.

Morris, R. J., O. T. Lewis, and H. C. J. Godfray. 2004. Experimental evidence for apparent competition in a tropical forest food web. Nature **428**:310-313.

Müller, C. B., I. C. T. Adriaanse, R. Belshaw, and H. C. J. Godfray. 1999. The structure of an aphid-parasitoid community. Journal of Animal Ecology **68**:346-370.

New, M., D. Lister, M. Hulme, and I. Makin. 2002. A high-resolution data set of surface climate over global land areas. Climate Research **21**:1-25.

Nicholls, J. A., S. Preuss, A. Hayward, G. Melika, G. Csoka, J. L. Nieves-Aldrey, R. R. Askew, M. Tavakoli, K. Schonrogge, and G. N. Stone. 2010. Concordant phylogeography and cryptic speciation in two Western Palaearctic oak gall parasitoid species complexes. Molecular Ecology **19**:592-609.

Nieves-Aldrey, J. L. and R. R. Askew. 1988. A new species of *Cecidostiba* Thomson (Hym. Pteromalidae), a key to species of the genus and rearing records of other Pteromalidea associated with oak gall wasps (Hym. Cynipidae). Entomological Monthly Gazette **124**:1-6.

O'Hara, R. B. and J. Merila. 2005. Bias and precision in Q(ST) estimates: Problems and some solutions. Genetics **171**:1331-1339.

Ode, P. J. 2006. Plant chemistry and natural enemy fitness: Effects on herbivore and natural enemy interactions. Annual Review of Entomology **51**:163-185.

Oksanen, J., F. G. Blanchet, R. Kindt, P. Legendre, R. B. O'Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, and H. Wagner. 2011. vegan: Community Ecology Package. R package version 1.17-11.

Padial, J. M., A. Miralles, I. De la Riva, and M. Vences. 2010. The integrative future of taxonomy. Frontiers in Zoology **7**.

Padidam, M., S. Sawyer, and C. M. Fauquet. 1999. Possible emergence of new geminiviruses by frequent recombination. Virology **265**:218-225.

Pamilo, P. and M. Nei. 1988. Relationships between Gene Trees and Species Trees. Molecular Biology and Evolution **5**:568-583.

Peakall, R. and P. E. Smouse. 2006. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Molecular Ecology Notes **6**:288-295.

Petit, R. J., S. Brewer, S. Bordacs, K. Burg, R. Cheddadi, E. Coart, J. Cottrell, U. M. Csaikl, B. van Dam, J. D. Deans, S. Espinel, S. Fineschi, R. Finkeldey, I. Glaz, P. G. Goicoechea, J. S. Jensen, A. O. Konig, A. J. Lowe, S. F. Madsen, G. Matyas, R. C. Munro, F. Popescu, D. Slade, H. Tabbener, S. G. M. de Vries, B. Ziegenhagen, J. L. de Beaulieu, and A. Kremer. 2002. Identification of refugia and post-glacial colonisation routes of European white oaks based on chloroplast DNA and fossil pollen evidence. Forest Ecology and Management **156**:49-74.

Pfenninger, M. and K. Schwenk. 2007. Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. Bmc Evolutionary Biology **7**.

Prado, P. and E. M. Vieira. 1999. The interplay between plant traits and herbivore attack: a study of a stem galling midge in the neotropics. Ecological Entomology **24**:80-88.

Pregitzer, C. C., J. K. Bailey, S. C. Hart, and J. A. Schweitzer. 2010. Soils as agents of selection: feedbacks between plants and soils alter seedling survival and performance. Evolutionary Ecology **24**:1045-1059.

Price, P. W. 1991. The Plant Vigor Hypothesis And Herbivore Attack. Oikos **62**:244-251.

Price, P. W., C. E. Bouton, P. Gross, B. A. McPheron, J. N. Thompson, and A. E. Weis. 1980. Interactions among 3 trophic levels - influence of plants on interactions between insect herbivores and natural enemies. Annual Review of Ecology and Systematics **11**:41-65.

Pujade-Villar, J., G. Melika, P. Ros-Farre, Z. Acs, and G. Csoka. 2003. Cynipid inquiline wasps of Hungary, with taxonomic notes on the Western Palaearctic Fauna (Hymenoptera: Cynipidae, Cynipinae, Synergini). Folia Entomologica Hungarica **64**:121-170.

R Development Core Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rambaut, A. and A. J. Drummond. 2007. Tracer v1.5, Available from http://beast.bio.ed.ac.uk/Tracer

Ratnasingham, S. and P. D. N. Hebert. 2007. BOLD: The Barcode of Life Data System (www.barcodinglife.org). Molecular Ecology Notes **7**:355-364.

Redfern, M. and P. Shirley. 2002. British plant galls: identification of galls on plants and fungi. Field Studies, 10, 207-531.

Rehfeldt, G. E., N. M. Tchebakova, Y. I. Parfenova, W. R. Wykoff, N. A. Kuzmina, and L. I. Milyutin. 2002. Intraspecific responses to climate in Pinus sylvestris. Global Change Biology **8**:912-929.

Rehfeldt, G. E., C. C. Ying, D. L. Spittlehouse, and D. A. Hamilton. 1999. Genetic responses to climate in Pinus contorta: Niche breadth, climate change, and reforestation. Ecological Monographs **69**:375-407.

Rhoades, D. F. 1985. Offensive-defensive interactions between herbivores and plants - their relevance in herbivore population-dynamics and ecological theory. American Naturalist **125**:205-238.

Roca, A. L., N. Georgiadis, J. Pecon-Slattery, and S. J. O'Brien. 2001. Genetic evidence for two species of elephant in Africa. Science **293**:1473-1477.

Rokas, A., R. J. Atkinson, L. Webster, G. Csoka, and G. N. Stone. 2003a. Out of Anatolia: longitudinal gradients in genetic diversity support an eastern origin for a circum-Mediterranean oak gallwasp Andricus quercustozae. Molecular Ecology **12**:2153-2174.

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

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

Ronquist, F., P. van der Mark, and J. P. Huelsenbeck. 2009. Bayesian phylogenetic analysis using MrBayes.*in* P. Lemey, M. Salemi, and A. Vandamme, M, editors. The Phylogenetic Handbook: a Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press.

Rosenberg, N. A. 2006. The mean and variance of the numbers of r-pronged nodes and r-caterpillars in yule-generated genealogical trees. Annals of Combinatorics **10**:129-146.

Rosenberg, N. A. 2007. Statistical tests for taxonomic distinctiveness from observations of monophyly. Evolution **61**:317-323.

Roy, B. A. 1998. Differentiating the effects of origin and frequency in reciprocal transplant experiments used to test negative frequency-dependent selection hypotheses. Oecologia **115**:73-83.

Saether, S. A., P. Fiske, J. A. Kalas, A. Kuresoo, L. Luigujoe, S. B. Piertney, T. Sahlman, and J. Hoglund. 2007. Inferring local adaptation from Q(ST)-F-ST comparisons: neutral genetic and quantitative trait variation in European populations of great snipe. Journal of Evolutionary Biology **20**:1563-1576.

Salminen, M. and D. Martin. 2009. Detecting and characterising individual recombination events.*in* P. Lemey, M. Salemi, and A. Vandamme, editors. The Phylogenetic Handbook: a Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press, Cambridge.

Samuel, C. J. A. 2003. Recent changes to the control of forest reproductive material. Forestry Commission information note 53., Forestry Commission, Edinburgh.

Savolainen, O., F. Bokma, R. Garcia-Gil, P. Komulainen, and T. Repo. 2004. Genetic variation in cessation of growth and frost hardiness and consequences for adaptation of Pinus sylvestris to climatic changes. Forest Ecology and Management **197**:79-89.

Savolainen, O., T. Pyhajarvi, and T. Knurr. 2007. Gene flow and local adaptation in trees. Annual Review of Ecology Evolution and Systematics **38**:595-619.

Schielzeth, H. 2010. Simple means to improve the interpretability of regression coefficients. Methods in Ecology and Evolution **1**:103-113.

Schönrogge, K., T. Begg, R. Williams, G. Melika, Z. Randles, and G. N. Stone. 2011. Range expansion and enemy recruitment by eight alien gall wasp species in Britain. Insect Conservation and Diversity **doi: 10.1111/j.1752-4598.2011.00161.x**.

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

Schönrogge, K. and M. J. Crawley. 2000. Quantitative webs as a means of assessing the impact of alien insects. Journal of Animal Ecology **69**:841-868.

Schönrogge, K., L. J. Harper, and C. P. Lichtenstein. 2000a. The protein content of tissues in cynipid galls (Hymenoptera : Cynipidae): Similarities between cynipid galls and seeds. Plant Cell and Environment **23**:215-222.

Schönrogge, K., S. Moriya, G. Melika, Z. Randle, T. Begg, A. Aebi, and G. N. Stone. 2007. Early parasitoid recruitment in invading cynipid galls.

Schönrogge, K., G. N. Stone, and R. R. Askew. in prep. Oak gall communities.*in* G. N. Stone, G. Csóka, and G. Melika, editors. The oak gallwasps of the Western Palearctic. The Ray Society, London.

Schonrogge, K., G. N. Stone, and M. J. Crawley. 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., G. N. Stone, and M. J. Crawley. 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., G. N. Stone, and M. J. Crawley. 1996. Abundance patterns and species richness of the parasitoids and inquilines of the alien gall-former Andricus quercuscalicis (Hymenoptera: Cynipidae). Oikos **77**:507-518.

Schönrogge, K., P. Walker, and M. J. Crawley. 1998. Invaders on the move: parasitism in the sexual galls of four alien gall wasps in Britain (Hymenoptera : Cynipidae). Proceedings of the Royal Society of London Series B-Biological Sciences **265**:1643-1650.

Schönrogge, K., P. Walker, and M. J. Crawley. 1999. Complex life cycles in Andricus kollari (Hymenoptera, Cynipidae) and their impact on associated parasitoid and inquiline species. Oikos **84**:293-301.

Schönrogge, K., P. Walker, and M. J. Crawley. 2000b. Parasitoid and inquiline attack in the galls of four alien, cynipid gall wasps: host switches and the effect on parasitoid sex ratios. Ecological Entomology **25**:208-219.

Schweitzer, J. A., J. K. Bailey, B. J. Rehill, G. D. Martinsen, S. C. Hart, R. L. Lindroth, P. Keim, and T. G. Whitham. 2004. Genetically based trait in a dominant tree affects ecosystem processes. Ecology Letters **7**:127-134.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 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, M. C. and C. Parmesan. 2010. Phenological asynchrony between herbivorous insects and their hosts: signal of climate change or pre-existing adaptive strategy? Philosophical Transactions of the Royal Society B-Biological Sciences **365**:3161-3176.

Sites, J. W. and J. C. Marshall. 2003. Delimiting species: a Renaissance issue in systematic biology. Trends in Ecology & Evolution **18**:462-470.

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

Smith, M. A., E. S. Eveleigh, K. S. McCann, M. T. Merilo, P. C. McCarthy, and K. I. van Rooyen. 2011. Barcoding a quantified food web: crypsis, concepts, ecology and hypotheses. Plos One **6**:e14424.

Smith, M. A., J. J. Rodriguez, J. B. Whitfield, A. R. Deans, D. H. Janzen, W. Hallwachs, and P. D. N. Hebert. 2008. Extreme diversity of tropical parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology, and collections. Proceedings of the National Academy of Sciences of the United States of America **105**:12359-12364.

Smith, M. A., D. M. Wood, D. H. Janzen, W. Hallwachs, and P. D. N. Hebert. 2007. DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Proceedings of the National Academy of Sciences of the United States of America **104**:4967-4972.

Smith, M. A., N. E. Woodley, D. H. Janzen, W. Hallwachs, and P. D. N. Hebert. 2006. DNA barcodes reveal cryptic host-specificity within the presumed

polyphagous members of a genus of parasitoid flies (Diptera : Tachinidae). Proceedings of the National Academy of Sciences of the United States of America **103**:3657-3662.

Sork, V. L., K. A. Stowe, and C. Hochwender. 1993. Evidence for local adaptation in closely adjacent subpopulations of northern red oak (*Quercus rubra* L) expressed as resistance to leaf herbivores. American Naturalist **142**:928-936.

Spittlehouse, D. L. and R. B. Stewart. 2003. Adaptation to climate change in forest management. BC Journal of Ecosystems and Management **4**:1-11.

Spitze, K. 1993. Population structure in *Daphnia obtusa* - quantitative genetic and allozymic variation. Genetics **135**:367-374.

St Clair, J. B. and G. T. Howe. 2007. Genetic maladaptation of coastal Douglas-fir seedlings to future climates. Global Change Biology **13**:1441-1454.

St Clair, J. B., N. L. Mandel, and K. W. Vance-Boland. 2005. Genecology of Douglas fir in western Oregon and Washington. Annals of Botany **96**:1199-1214.

Stone, G., A. Hernandez-Lopez, J. A. Nicholls, E. Pierro, J. Pujade-Villar, G. Melika, and J. M. Cook. 2009. Extreme Host Plant Conservatism During at Least 20 Million Years of Host Plant Persuit by Oak Gallwasps. Evolution **63**:854-869.

Stone, G. N., R. J. Atkinson, and G. Csóka. in prep. Host plant associations of oak gallwasps.*in* G. N. Stone, G. Csóka, and G. Melika, editors. The oak gallwasps of the Western Palearctic. The Ray Society, London.

Stone, G. N., R. J. Atkinson, A. Rokas, J. L. Nieves-Aldrey, G. Melika, Z. Acs, G. Csoka, A. Hayward, R. Bailey, C. Buckee, and G. A. T. McVean. 2008. Evidence for widespread cryptic sexual generations in apparently purely asexual Andricus gallwasps. Molecular Ecology **17**:652-665.

Stone, G. N. and J. M. Cook. 1998. The structure of cynipid oak galls: patterns in the evolution of an extended phenotype. Proceedings of the Royal Society of London Series B-Biological Sciences **265**:979-988.

Stone, G. N., S. Nee, and J. Felsenstein. 2011. Controlling for non-independance in comparative analysis of patterns across populations within species. Philosophical Transactions of the Royal Society B-Biological Sciences **366**:1410-1424.

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

Stone, G. N., K. Schönrogge, R. J. Atkinson, D. Bellido, and J. Pujade-Villar. 2002. The population biology of oak gall wasps (Hymenoptera : Cynipidae). Annual Review of Entomology **47**:633-668.

Stone, G. N., K. Schonrogge, M. J. Crawley, and S. Fraser. 1995. Geographic and between-Generation Variation in the Parasitoid Communities Associated with an Invading Gallwasp, Andricus-Quercuscalicis (Hymenoptera, Cynipidae). Oecologia **104**:207-217.

Swofford, D. L. 2003. PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Symonds, M. R. E. and A. Moussalli. 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behavioral Ecology and Sociobiology **65**:13-21.

Tack, A. J. M., S. Gripenberg, and T. Roslin. 2011. Can we predict indirect interactions from quantitative food webs? - an experimental approach. Journal of Animal Ecology **80**:108-118.

Tack, A. J. M., O. Ovaskainen, P. Pulkkinen, and T. Roslin. 2010. Spatial location dominates over host plant genotype in structuring an herbivore community. Ecology **91**:2660-2672.

Tack, A. J. M. and T. Roslin. 2010. Overrun by the neighbors: Landscape context affects strength and sign of local adaptation. Ecology **91**:2253-2260.

Tamura, K., J. Dudley, M. Nei, and S. Kumar. 2007. MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. Molecular Biology and Evolution **24**:1596-1599.

Tautz, D., P. Arctander, A. Minelli, R. H. Thomas, and A. P. Vogler. 2003. A plea for DNA taxonomy. Trends in Ecology & Evolution **18**:70-74.

Tay, W. T., G. T. Behere, D. G. Heckel, S. F. Lee, and P. Batterham. 2008. Exon-primed intron-crossing (EPIC) PCR markers of Helicoverpa armigera (Lepidoptera: Noctuidae). Bulletin of Entomological Research **98**:509-518.

Team, R. C. D. 2011. R Core Development Team. 2011. R: A language and environment for statistical computing. R Foundation for Statistical Computing, , Vienna, Austria. .

Tylianakis, J. M., T. Tscharntke, and O. T. Lewis. 2007. Habitat modification alters the structure of tropical host-parasitoid food webs. Nature **445**:202-205.

Underwood, N. and M. D. Rausher. 2000. The effects of host-plant genotype on herbivore population dynamics. Ecology **81**:1565-1576.

van Asch, M., P. H. Tienderen, L. J. M. Holleman, and M. E. Visser. 2007. Predicting adaptation of phenology in response to climate change, an insect herbivore example. Global Change Biology **13**:1596-1604.

van Asch, M. and M. E. Visser. 2007. Phenology of forest caterpillars and their host trees: The importance of synchrony. Annual Review of Entomology **52**:37-55.

van Veen, F. J. F., R. J. Morris, and H. C. J. Godfray. 2006. Apparent competition, quantitative food webs, and the structure of phytophagous insect communities. Annual Review of Entomology **51**:187-208.

van Zandt, P. A. and S. Mopper. 1998. A Meta-Analysis of Adaptive Deme Formation in Phytophagous Insect Populations. The American Naturalist **152**:595-604.

Via, S. 1991. The genetic-structure of host plant adaptation in a spatial patchwork - demographic variability among reciprocally transplanted pea aphid clones. Evolution **45**:827-852.

Vieites, D. R., K. C. Wollenberg, F. Andreone, J. Koehler, F. Glaw, and M. Vences. 2009. Vast underestimation of Madagascar's biodiversity evidenced by an integrative amphibian inventory. Proceedings of the National Academy of Sciences of the United States of America **106**:8267-8272.

Vitasse, Y., C. C. Bresson, A. Kremer, R. Michalet, and S. Delzon. 2010. Quantifying phenological plasticity to temperature in two temperate tree species. Functional Ecology **24**:1211-1218.

Vitasse, Y., S. Delzon, E. Dufrene, J. Y. Pontailler, J. M. Louvet, A. Kremer, and R. Michalet. 2009. Leaf phenology sensitivity to temperature in European trees: Do within-species populations exhibit similar responses? Agricultural and Forest Meteorology **149**:735-744.

Vogler, A. P. and M. T. Monaghan. 2007. Recent advances in DNA taxonomy. Journal of Zoological Systematics and Evolutionary Research **45**:1-10.

Wang, T., A. Hamann, A. Yanchuk, G. A. O'Neill, and S. N. Aitken. 2006. Use of response functions in selecting lodgepole pine populations for future climates. Global Change Biology **12**:2404-2416.

Wang, T. L., G. A. O'Neill, and S. N. Aitken. 2010. Integrating environmental and genetic effects to predict responses of tree populations to climate. Ecological Applications **20**:153-163.

Weir, B. S. and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution **38**:1358-1370.

White, T. C. R. 1969. An Index to Measure Weather-Induced Stress of Trees Associated With Outbreaks of Psyllids in Australia. Ecology **50**:905-909.

White, T. C. R. 1974. Hypothesis to Explain Outbreaks of Looper Caterpillars, with Special Reference to Populations of Selidosema-Suavis in a Plantation of Pinus-Radiata in New-Zealand. Oecologia **16**:279-301.

White, T. C. R. 1984. The Abundance of Invertebrate Herbivores in Relation to the Availability of Nitrogen in Stressed Food Plants. Oecologia **63**:90-105.

White, T. C. R. 2009. Plant vigour versus plant stress: a false dichotomy. Oikos **118**:807-808.

Whitham, T. G., J. K. Bailey, J. A. Schweitzer, S. M. Shuster, R. K. Bangert, C. J. Leroy, E. V. Lonsdorf, G. J. Allan, S. P. DiFazio, B. M. Potts, D. G. Fischer, C. A. Gehring, R. L. Lindroth, J. C. Marks, S. C. Hart, G. M. Wimp, and S. C. Wooley. 2006. A framework for community and ecosystem genetics: from genes to ecosystems. Nature Reviews Genetics **7**:510-523.

Whitham, T. G., W. P. Young, G. D. Martinsen, C. A. Gehring, J. A. Schweitzer, S. M. Shuster, G. M. Wimp, D. G. Fischer, J. K. Bailey, R. L. Lindroth, S. Woolbright, and C. R. Kuske. 2003. Community and ecosystem genetics: A consequence of the extended phenotype. Ecology **84**:559-573.

Whitlock, M. C. 2008. Evolutionary inference from Q(ST). Molecular Ecology **17**:1885-1896.

Whitlock, M. C. and F. Guillaume. 2009. Testing for Spatially Divergent Selection: Comparing Q(ST) to F-ST. Genetics **183**:1055-1063.

Whittingham, M. J., P. A. Stephens, R. B. Bradbury, and R. P. Freckleton. 2006. Why do we still use stepwise modelling in ecology and behaviour? Journal of Animal Ecology **75**:1182-1189.

Wiemers, M. and K. Fiedler. 2007. Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). Frontiers in Zoology **4**.

Wimp, G. M., G. D. Martinsen, K. D. Floate, R. K. Bangert, and T. G. Whitham. 2005. Plant genetic determinants of arthropod community structure and diversity. Evolution **59**:61-69.

Worrell, R. 1992. A Comparison between European Continental and British Provenances of Some British Native Trees - Growth, Survival and Stem Form. Forestry **65**:253-280.

Wright, J. W. 2007. Local adaptation to serpentine soils in Pinus ponderosa. Plant and Soil **293**:209-217.

Wright, S. 1951. The genetical structure of populations. Annals of Eugenics **15**:323-354.

Yukawa, J. 2000. Synchronization of gallers with host plant phenology. Population Ecology **42**:105-113.

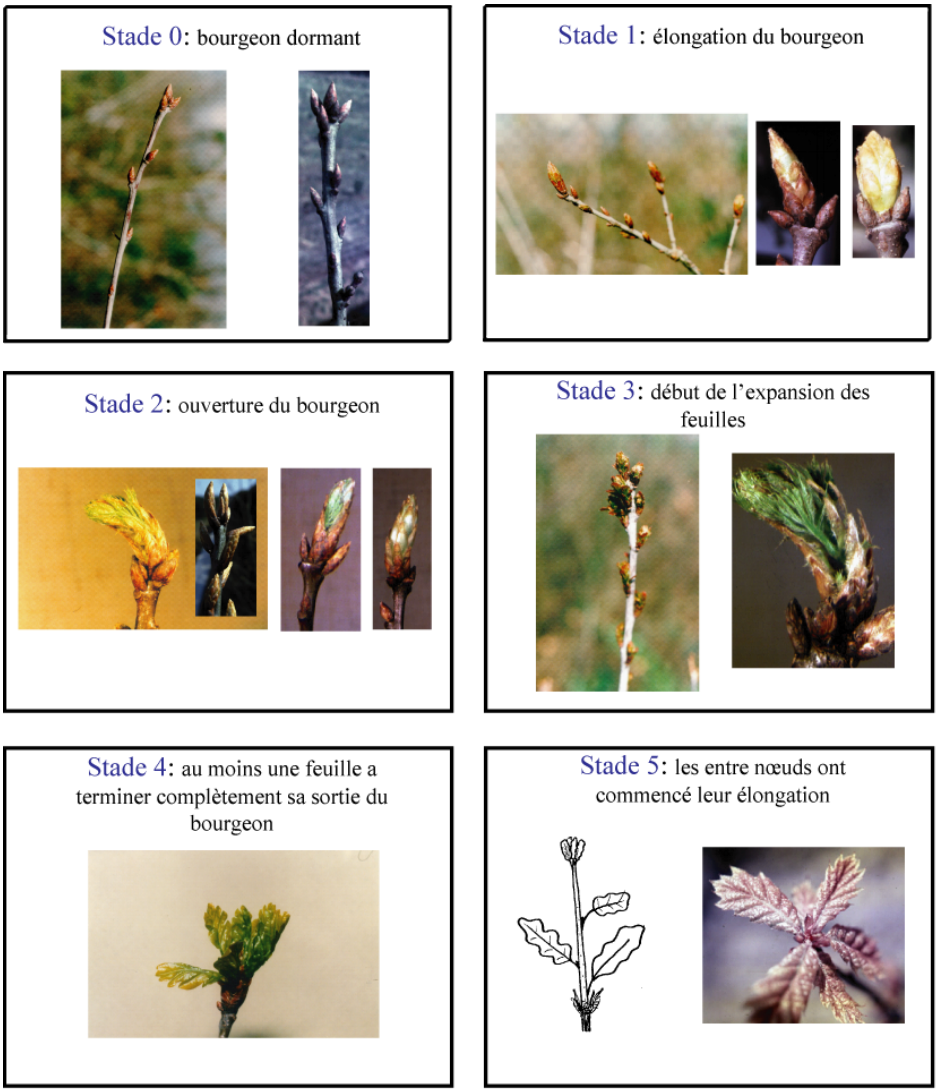
Zhu, S., J. H. Degnan, and M. Steel. 2011. Clades, clans, and reciprocal monophyly under neutral evolutionary models. Theoretical Population Biology **79**:220-227.

Appendices

Appendix 2.1. Procedures for measuring tree phenotypic traits

A2.1.1 Spring bud-burst phenology

The bud-burst phase of each tree in tranch 4 of the Petite Charnie provenance trials was assessed by INRA researchers on a single day in spring 1995, with the following 6 stage scoring system. The scoring ranges from late flushing trees still with dormant buds (stage 0), to early flushing trees with open leaflets that hung separately (stage 5).



A2.1.2. Autumn leaf-fall phenology

The degree of leaf retention for each tree in tranch 4 of the Petite Charnie provenance trials was assessed by INRA researchers on a single day in late autumn 2001, with the following 6 stage scoring system. The scoring ranges The scoring ranges from all/most leaves shed (stage 0), to all/most leaves retained (stage 5).

Note 0 : aucune feuille

Note 1 : quelques feuilles

Note 2 : feuilles uniquement sur la partie centrale de l Õarbre

Note 3 : des feuilles pr˘sentent des feuilles en densit˘ moyenne sur sur une grande partie de l Õarbre

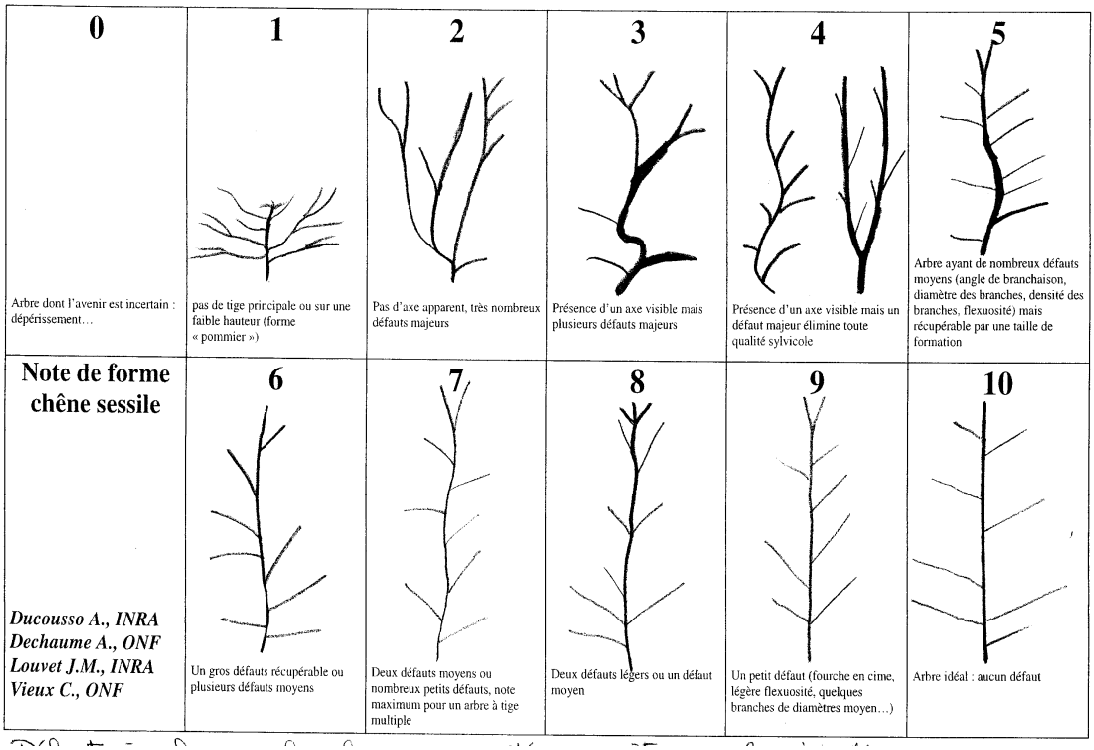
Note 4 : absence de feuille uniquement sur l Õextr mit˘ des rameau ou la densit˘ est faible , le reste est tr¸s dense

Note 5 : toutes les feuilles pr˘sentes

NOTATION MARCESCENCE : ch ne sessile

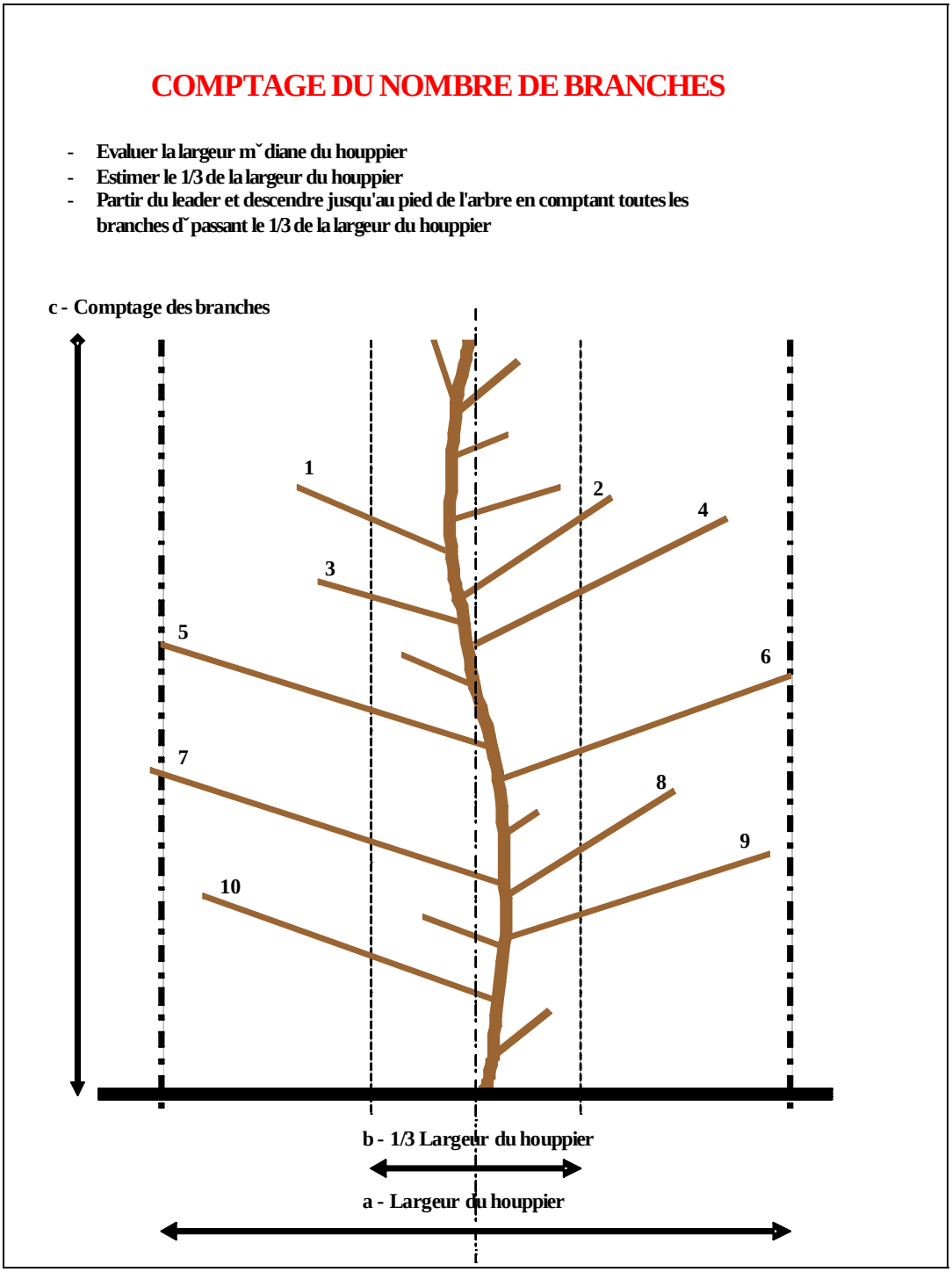
A2.1.3. Tree form

The overall 'form' of each tree was assessed by INRA researchers during winter 2001-02 with the following 10 stage scoring system. Form in this sense is related to the quality of the tree from a commercial perspective, and the scoring ranges from bushy with poor apical dominance (score 1), to straight stemmed with even branching and strong apical dominance (score 10).



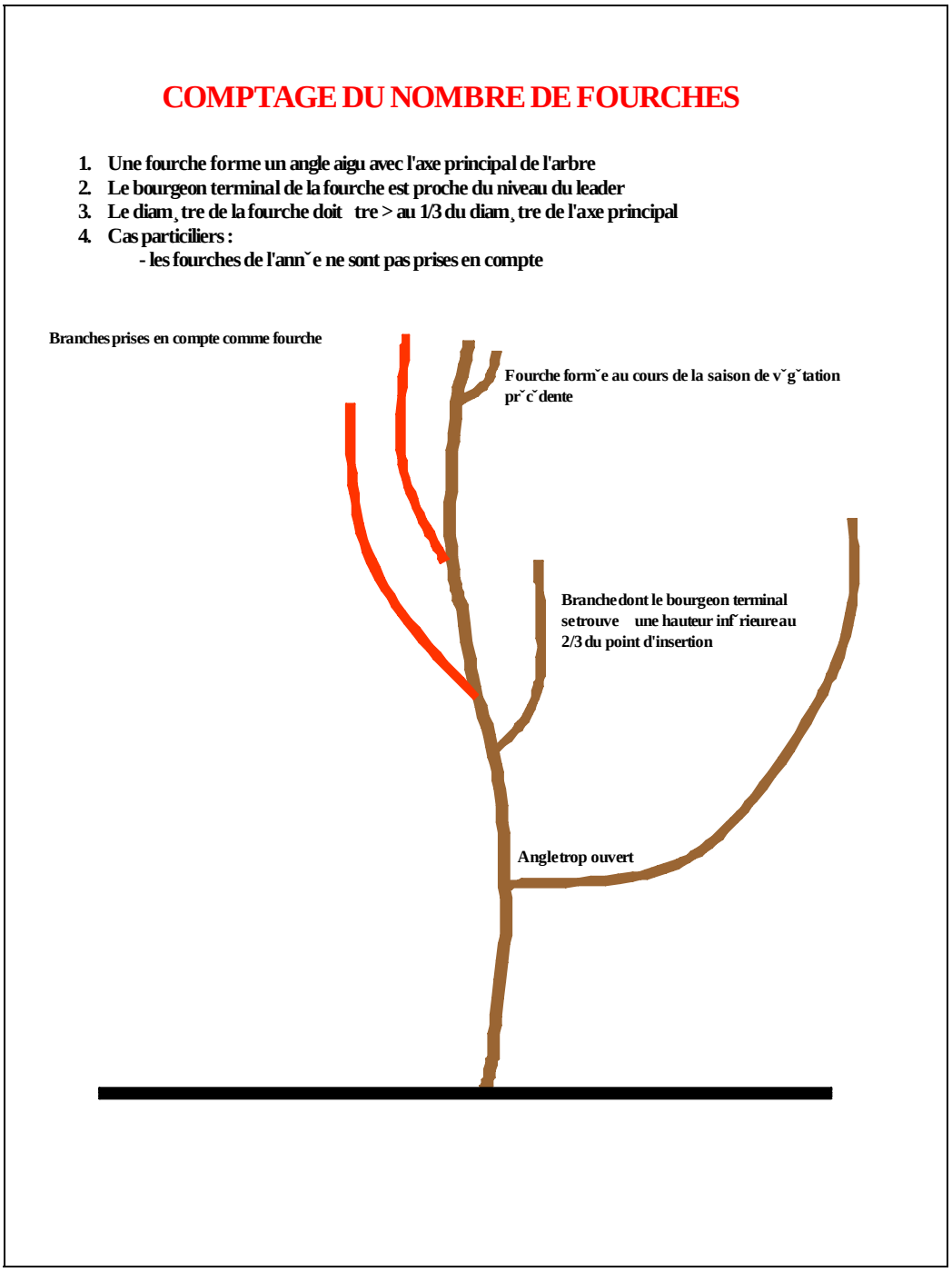
A2.1.4. Number of branches

The number of branches for each tree in the Petite Chanie provenance trial was assessed by INRA researchers during winter 2001-02. Branches were defined as limbs originating from the trunk that protruded to at least 1/3 of the radius of the tree crown, as illustrated in the following diagram.



A2.1.5 Number of forks

The number of forks for each tree in the Petite Charnie provenance trials was assessed by INRA researchers during winter 2001-02. Forks were defined as limbs protruding from the main axis at an acute angle, that were at least 1/3 the diameter of the main axis, and whose terminal bud reached to approximately the same height as the main axis, as illustrated in the following diagram.



Appendix 4.1. Host association data

The following tables contain the source data used to construct the bipartite host association networks presented in Chapter 4.

Table A4.1.1. Table showing the number if individuals of each of 23 parasitoid morpho-species (columns) reared from each of 17 host gall-types (rows). Taxon codes follow Tables 4.1 and 4.2

	Aaet	**Aars**	**Acer**	**Aski**	**Cdia**	**Eann**	**Ebru**	**Espl**	**Euro**	**Mdor**	**Mdub**	**Mfas**	**Mfus**	**Mmed**	**Mtib**	**Onit**	**Opom**	**Over**	**Plys**	**Svar**	**Taur**	**Tfla**	**Tger**
AcallAsex	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AcurvSex	0	0	0	0	0	0	0	0	1	0	0	1	0	0	4	0	0	0	0	0	0	1	0
AfecAsex	0	0	0	0	0	0	2	0	0	7	0	0	0	0	0	1	0	0	0	0	0	0	0
AglanAsex	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0
AkollAsex	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	10	0	1
AsolAsex	0	0	0	0	0	0	82	0	10	0	0	0	0	0	0	5	0	0	0	1	0	0	0
BpalSex	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	8	0
CdivAsex	0	0	0	0	0	0	5	0	0	0	0	5	0	0	0	0	0	0	0	0	1	0	4
CqfAsex	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	0
NalbAsex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	154	0	80	0	0	118	0
NalbSex	11	1	1	0	0	0	0	0	0	0	0	1	0	0	49	0	0	12	0	0	0	4	0
NantAsex	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1
NantSex	3	1	0	0	0	1	0	3	13	9	1	1	4	2	179	0	0	43	0	2	0	4	0
NnAsex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	89	0	0	0	0
NnSex	3	8	0	0	1	0	2	0	4	0	1	50	0	0	52	0	0	0	0	0	0	2	0
NqbAsex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0
NqbSex	1	14	0	4	0	0	12	0	30	5	1	112	0	2	62	0	1	1	0	3	0	26	11

Table A4.1.2. Table showing the number if individuals of each of 35 parasitoid MOTUs (as defined at the decisive limit (13 bp) of the barcoding gap, columns) reared from each of 17 host gall-types (rows). Taxon codes follow Tables 4.1 and 4.2. Table is continued on the following page.

	Aaet_1	**Aaet_2**	**Aars_1**	**Acer_1**	**Aski_1**	**Cdia_1**	**Eann_3**	**Ebru_1**	**Ebru_2**	**Ebru_3**	**Espl_1**	**Euro_1**	**Euro_2**	**Mdor_1**	**Mdor_2**	**Mdub_1**	**Mfas_1**	**Mfus_1**	**Mmed_1**	**Mtib_1**
AcallAsex	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
AcurvSex	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	4
AfecAsex	0	0	0	0	0	0	0	0	1	1	0	0	0	6	1	0	0	0	0	0
AglanAsex	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
AkollAsex	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0
AsolAsex	0	0	0	0	0	0	1	75	7	0	0	8	1	0	0	0	0	0	0	0
BpalSex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
CdivAsex	0	0	0	0	0	0	0	4	1	0	0	0	0	0	0	0	5	0	0	0
CqfAsex	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
NalbAsex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NalbSex	11	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	49
NantAsex	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0
NantSex	2	1	1	0	0	0	0	0	0	0	3	14	0	0	9	1	1	4	2	179
NnAsex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NnSex	3	0	8	0	0	1	0	2	0	0	0	4	0	0	0	1	50	0	0	52
NqbAsex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NqbSex	0	1	14	0	4	0	0	11	1	0	0	30	0	0	5	1	112	0	2	62

Table A4.1.2. Continued

	Onit_1	Opom_1	Opom_2	Opom_3	Opom_4	Opom_5	Over_1	Over_2	Plys_1	Plys_2	Sbig_1	Taur_1	Tfla_1	Tfla_2	Tger_1
AcallAsex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AcurvSex	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
AfecAsex	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AglanAsex	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
AkollAsex	0	0	0	0	0	0	0	0	0	0	0	10	0	0	1
AsolAsex	5	0	0	0	0	0	0	0	0	0	1	0	0	0	0
BpalSex	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0
CdivAsex	0	0	0	0	0	0	0	0	0	0	0	1	0	0	4
CqfAsex	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0
NalbAsex	0	40	96	16	1	1	0	0	25	55	0	0	91	27	0
NalbSex	0	0	0	0	0	0	3	9	0	0	0	0	4	0	0
NantAsex	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
NantSex	0	0	0	0	0	0	0	43	0	0	2	0	4	0	0
NnAsex	0	0	0	0	0	0	0	0	0	89	0	0	0	0	0
NnSex	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
NqbAsex	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0
NqbSex	0	0	0	1	0	0	0	1	0	0	3	0	26	0	11

Table A4.1.3. Table showing the number if individuals of each of 10 inquiline MOTUs (as defined at the decisive limit (11 bp threshold) of the barcoding gap, columns) reared from each of 17 host gall-types (rows). Taxon codes follow Tables 4.1 and 4.2.

	Inq_1	Inq_2	Inq_3	Inq_4	Inq_5	Inq_6	Inq_7	Inq_8	Inq_9	Inq_10
NalbSex	0	0	0	0	5	1	0	0	0	0
NantSex	1	0	0	1	0	2	0	0	0	0
NnSex	46	0	0	0	154	7	0	0	0	0
NqbSex	40	0	1	161	78	285	0	0	0	0
AcurvSex	0	0	0	1	0	4	0	0	0	0
BpalSex	0	0	0	0	0	0	0	0	0	0
NalbAsex	6	0	0	0	1	0	0	0	0	0
NantAsex	0	0	0	26	0	8	0	0	0	0
NnAsex	0	0	0	0	0	0	0	0	0	0
NqbAsex	0	0	0	0	0	0	0	0	0	0
AcallAsex	0	0	0	3	0	0	0	0	0	0
AfecAsex	0	0	7	0	0	0	0	42	0	0
AglanAsex	0	0	0	0	0	0	0	0	2	0
AkollAsex	0	0	0	0	0	0	138	0	0	0
AsolAsex	0	0	1	0	0	0	0	0	0	2
CdivAsex	0	0	0	2	0	3	0	0	0	0
CqfAsex	0	10	0	0	0	2	0	0	0	0

Appendix 4.2. Complied host association data

The following table details the compiled host association data for the parasitoid and inquiline MOTUs (as defined at the decisive thresholds of 13 and 11 bp respectively) that were matched with the studies of Ács et al. (2010) and Kaartinen et al. (2010) in Chapter 4.

Table A4.2.1. Summary of the compiled host association data for parasitoid and inquiline MOTUs that could be matched with sequences published by Ács et al. (2010), and Kaartinen et al. (2010), including morpho-species names, any alternative names applied to these taxa, the countries where they were recorded, and their host gall-types. Taxon codes follow Table 4.2. Two letter country codes follow those provided by the International Organization for Standardization, and are presented in approximate geographic order from west to east. Records from France (FR) are from this study, records labelled as [1] are from Kaartinen et al. (2010), and [2] are from Ács et al (2010). Numbers within brackets indicate the individuals from a particular host species recorded within a particular country. The table is continued on following pages.

Taxon code	Morpho-species	Other names	Distribution	Host gall-type
Parasitoids				
Aaet_2	*Aprostocetus aethiops*		FI[1] (13)	*Andricus callidoma* (Sex)
			FI[1] (2)	*Andricus glandulae* (Sex)
			FR (1)	*Neuroterus anthracinus* (Sex)
			FR (1)	*Neuroterus quercusbaccarum* (Sex)
Acer	*Aprostocetus cerricola*		FR (1)	*Neuroterus albipes* (Sex)
			HU[1] (1)	*Neuroterus numismalis* (Sex)
Aars	*Aulogymnus arsames*		FI[1] (11)	*Andricus curvator* (Sex)
			HU[1] (1)	*Andricus multiplicatus* (Sex)
			FR (1)	*Neuroterus albipes* (Sex)
			FR (1)	*Neuroterus anthracinus* (Sex)
			FR (8)	*Neuroterus numismalis* (Sex)
			FR (14), FI[1] (10)	*Neuroterus quercusbaccarum* (Sex)
			HU[1] (1)	Unknown bud gall (Sex)
Cdia	*Cirrospilus diallus*		FI[1] (2)	*Ectoedemia albifasciella*
			FR (1)	*Neuroterus numismalis* (Sex)
			FI[1] (2)	*Phyllonorycter sp.*
Eann	*Eupelmus annulatus*		FR (1)	*Andricus solitarius* (Asex)
			ES[1] (1)	*Cynips quercus* (Asex)

Table A4.2.1. Continued

Taxon code	Morpho-species	Other names	Distribution	Host gall-type
Euro_1	*Eupelmus urozonus*	*Eupelmus urozonus A*[1]	ES[1] (1), HU[1] (1)	*Andricus burgundus* (Sex)
			FI[1] (1)	*Andricus callidoma* (Asex)
			HU[1] (1)	*Andricus caputmedusae* (Asex)
			FR (1), FI[1] (10)	*Andricus curvator* (Sex)
			PT[1] (1)	*Andricus kollari* (Asex)
			HU[1] (1)	*Andricus lucidus* (Asex)
			HU[1] (1)	*Andricus multiplicatus* (Asex)
			ES[1] (2)	*Andricus quercustozae* (Asex)
			FR (8)	*Andricus solitarius* (Asex)
			ES[1] (4)	*Cynips quercus* (Asex)
			ES[1] (1)	*Diplolepis mayri* (Sex)
			FR (14)	*Neuroterus anthracinus* (Sex)
			FR (4)	*Neuroterus numismalis* (Sex)
			FR (30), FI[1] (2)	*Neuroterus quercusbaccarum* (Sex)
			HU[1] (1)	Unknown bud gall (Sex)
Onit	*Ormyrus nitidulus*		ES[1] (1)	*Andricus grosulariae* (Sex)
			ES[1] (1)	*Andricus grosulariae* (Asex)
			FR (1)	*Andricus feccundator* (Asex)
			FR (5)	*Andricus solitarius* (Asex)
			FR[1] (1)	*Biorhiza pallida* (Sex)
Opom_3	*Ormyrus pomaceus*		HU[1] (1)	*Andricus crispator* (Sex)
			HU[1] (2)	*Andricus multiplicatus* (Sex)
			FI[1] (8)	*Cynips longiventris* (Asex)
			HU[1] (1)	*Chilaspis nitida* (Asex)
			FR (16)	*Neuroterus albipes* (Asex)
			FR (1)	*Neuroterus quercusbaccarum* (Sex)
Mdub	*Mesopolobus dubius*		UK[1] (2)	*Andicus quercuscalicis* (Sex)
			FR (1)	*Neuroterus anthracinus* (Sex)
			FR (1)	*Neuroterus numismalis* (Sex)
			FR (1)	*Neuroterus quercusbaccarum* (Sex)

Table A4.2.1. Continued

Taxon code	Morpho-species	Other names	Distribution	Host gall-type
Mfas	*Mesopolobus fasciventris*		FI[1] (1)	*Andricus callidoma* (Sex)
			FR (1)	*Andricus curvator* (Sex)
			HU[1] (1)	*Andricus kollari* (Asex)
			FI[1] (1)	*Andricus paradoxus* (Sex)
			ES[1] (1)	*Andricus quercustozae* (Asex)
			HU[1] (1)	*Aphelonyx cerricola* (Sex)
			FR (5)	*Cynips divisa* (Asex)
			FI[1] (19)	*Cynips longiventris* (Asex)
			HU[1] (1)	*Chilaspis nitida* (Asex)
			ES[1] (1)	*Cynips quercus* (Asex)
			FR (1)	*Neuroterus albipes* (Sex)
			FR (1)	*Neuroterus anthracinus* (Sex)
			FR (1)	*Neuroterus anthracinus*
			FR (50)	*Neuroterus numismalis* (Sex)
			FR (120)	*Neuroterus quercusbaccarum* (Sex)
Mfus	*Mesopolobus fuscipes*		HU[1] (1)	*Andricus multiplicatus* (Sex)
			UK[1] (2)	*Andricus quercuscalisis* (Sex)
			FR (4)	*Neuroterus anthracinus* (Sex)
Mmed	*Mesopolobus mediterraneus*		UK[1] (1)	*Andicus quercuscalicis* (Sex)
			FR (2)	*Neuroterus anthracinus* (Sex)
			FR (2)	*Neuroterus quercusbaccarum* (Sex)
Mtib	*Mesopolobus tibialis*		FI[1] (4)	*Andricus callidoma* (Sex)
			HU[1] (2)	*Andricus crispator* (Sex)
			FR (4), FI[1] (2), HU[1] (2)	*Andricus curvator* (Sex)
			FI[1] (4)	*Andricus glandulae* (Sex)
			FI[1] (5)	*Andricus quadrilineatus* (Asex)
			HU[1] (1)	*Andricus schroeckingeri* (Sex)
			HU[1] (1)	*Andricus singularis* (Sex)
			HU[1] (1)	*Cynips divisa* (Sex)
			FR (49)	*Neuroterus albipes* (Sex)
			FR (179)	*Neuroterus anthracinus* (Sex)
			FR (52), HU[1] (1)	*Neuroterus numismalis* (Sex)
			FR (52)	*Neuroterus quercusbaccarum* (Sex)
			HU[1] (1)	Unknown Cynips sp. (Sex)

Table A4.2.1. Continued

Taxon code	Morpho-species	Other names	Distribution	Host gall-type
Taur	*Torymus auratus*		FR (10)	*Andricus kollari* (Asex)
			ES[1] (1), HU[1] (1)	*Biorhiza pallida* (Sex)
			ES[1] (1)	*Cynips quercus* (Asex)
			FR (27), HU[1] (1)	*Cynips quercusfolii* (Asex)
Tfla_1	*Torymus flavipes*	*Torymus flavipes A*	FI[1] (2)	*Andricus callidoma* (Sex)
			FR (1), FI[1] (10), HU[1] (2)	*Andricus curvator* (Sex)
			FI[1] (2)	*Andricus glandulae* (Sex)
			FI[1] (12)	*Andricus pseudoinflator* (Sex)
			FI[1] (14)	*Andricus quadrilineatus* (Asex)
			HU[1] (3)	*Andricus quercusramuli* (Sex)
			FR (8), FI[1] (4)	*Biorhiza pallida* (Sex)
			FR (91)	*Neuroterus albipes* (Asex)
			FR (4)	*Neuroterus anthracinus* (Sex)
			FR (26), FI[1] (5), HU[1] (3)	*Neuroterus quercusbaccarum* (Sex)
			FI[1] (2)	*Neuroterus quercusbaccarum* (Asex)
Tfla_2	*Torymus flavipes*	*Torymus flavipes B*	FI[1] (1)	*Andricus pseudoinflator* (Sex)
			FR (27)	*Neuroterus albipes* (Asex)
			FI[1] (5)	*Neuroterus quercusbaccarum* (Asex)
Tger	*Torymus geranii*		FI[1] (2)	*Andricus curvator* (Sex)
			HU[1] (1)	*Biorhiza pallida* (Sex)
			HU[1] (1)	*Cynips divisa* (Asex)
			FI[1] (5), HU[1] (1)	*Cynips longiventris* (Asex)
			HU[1] (1)	*Cynips quercusfolii* (Asex)
			FI[1] (1)	*Neuroterus quercusbaccarum* (Sex)
Inquilines				
Inq_1		*MOTU 4*[1]	FR (6)	*Neuroterus albipes* (Asex)
			FR (1)	*Neuroterus anthracinus* (Sex)
			FR (46)	*Neuroterus numismalis* (Sex)
			FR (40)	*Neuroterus quercusbaccarum* (Sex)
			FI[1] (8)	*Neuroterus quercusbaccarum* (Asex)
Inq_2		*MOTU 3*[1]	FI[1] (3)	*Cynips longiventris* (Asex)
			FR (10)	*Cynips quercusfolii* (Asex)

Table A4.2.1. Continued

Taxon code	Morpho-species	Other names	Distribution	Host gall-type
Inq_4		*MOTU 2*[1]	FR (3), FI[1] (5)	*Andricus callidoma* (Asex)
			FI[1] (7)	*Andricus curvator* (Sex)
			FI[1] (4)	*Andricus nudus* (Asex)
			ES[2] (2), FI[1] (27)	*Andricus quadrilineatus* (Sex)
			FI[1] (2)	*Andricus quercusramuli* (Asex)
			FR (1)	*Neuroterus anthracinus* (Sex)
			FR (26), FI[1] (2)	*Neuroterus anthracinus* (Asex)
			FR (161), FI[1] (11)	*Neuroterus quercusbaccarum* (Sex)
			FR (2)	*Cynips divisa* (Asex)
Inq_6		*Synergus sp. 3*[2]	FR (4), FI[1] (1)	*Andricus curvator* (Sex)
			FR (3)	*Cynips divisa* (Asex)
			FI[1] (12)	*Cynips longiventris* (Asex)
			FR (2)	*Cynips quercusfolii* (Asex)
			FR (1)	*Neuroterus albipes* (Sex)
			FR (2)	*Neuroterus anthracinus* (Sex)
			FR (8), HU[2] (1)	*Neuroterus anthracinus* (Asex)
			FR (7)	*Neuroterus numismalis* (Sex)
			FR (285)	*Neuroterus quercusbaccarum* (Sex)
Inq_10	*Ceroptres clavicornis*		HU[1] (1)	*Andricus conglomerates* (Asex)
			HU[1] (1)	*Andricus lignicolus* (Asex)
			FR (2)	*Andricus solitarius* (Asex)

Appendix 5.1. *Cecidostiba s*pecimen details

The following table provides collection details for the 171 parasitoid specimens analysed in Chapter 5.

Table A5.1. Summary of collection details for 171 parasitoid specimens, including morpho-species classifications, inferred species identities following the analyses, sex (m = male, f = female), host gall species and generation, the country, locality and year of collection, and host oak species. Unknown data are indicated as '?'. Code numbers follow the Stone laboratory parasitoid database. Table is continued on the following pages.

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0001	f	*Biorhiza pallida*	Sexual	UK	Bovey Tracey	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca2]*	Cfun0003	f	*Biorhiza pallida*	Sexual	UK	Puttenham	2007	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0004	f	*Andricus quercuscalicis*	Asexual	UK	Silwood	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0005	f	*Biorhiza pallida*	Sexual	UK	Silwood	2007	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0006	f	*Andricus quercuscalicis*	Asexual	UK	Puttenham	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0007	f	*Biorhiza pallida*	Sexual	UK	Hainhault Forest	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0008	f	*Andricus quercuscalicis*	Asexual	UK	Hainhault Forest	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0009	f	*Andricus quercuscalicis*	Asexual	UK	Rufford	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0010	f	*Andricus quercuscalicis*	Asexual	UK	Lancaster	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0011	f	*Andricus grossulariae*	Sexual	UK	Puttenham	2006	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0012	f	*Andricus grossulariae*	Sexual	UK	Silwood	2006	*Quercus cerris*

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0013	m	*Andricus grossulariae*	Sexual	UK	Maidenhead	2006	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0014	f	*Andricus grossulariae*	Sexual	UK	Farnham Park	2006	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0015	f	*Andricus grossulariae*	Sexual	UK	Farnham Park	2006	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0018	m	*Andricus grossulariae*	Sexual	UK	Puttenham	2007	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0019	f	*Andricus grossulariae*	Sexual	UK	Silwood	2007	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0020	f	*Andricus quercuscalicis*	Asexual	France	Arboretum Nogent	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0021	f	*Andricus quercuscalicis*	Asexual	France	Arboretum Nogent	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0022	f	*Andricus quercuscalicis*	Asexual	France	Arboretum Nogent	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0023	f	*Andricus quercuscalicis*	Asexual	France	Arboretum Nogent	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0024	f	*Andricus quercuscalicis*	Asexual	France	Arboretum Nogent	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0025	f	*Andricus quercuscalicis*	Asexual	France	Arboretum Nogent	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0026	f	*Andricus quercuscalicis*	Asexual	France	Arboretum Nogent	2005	*Quercus robur*

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0028	f	*Andricus quercuscalicis*	Asexual	France	Arboretum Nogent	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0039	f	*Biorhiza pallida*	Sexual	Austria	Unterlois	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0043	m	*Biorhiza pallida*	Sexual	Austria	Ober Pullendorf	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0044	f	*Biorhiza pallida*	Sexual	Austria	Ober Pullendorf	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0045	f	*Andricus grossulariae*	Sexual	Austria	Unterlois	2006	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0046	f	*Biorhiza pallida*	Sexual	Germany	Ludwigsburg	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0047	f	*Andricus quercuscalicis*	Asexual	Germany	Ludwigsburg	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0048	f	*Andricus quercuscalicis*	Asexual	Germany	Ludwigsburg	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0049	f	*Andricus quercuscalicis*	Asexual	Germany	Ludwigsburg	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0050	f	*Andricus quercuscalicis*	Asexual	Germany	Ludwigsburg	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0051	f	*Andricus quercuscalicis*	Asexual	Germany	Ludwigsburg	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0052	f	*Andricus lignicolus*	Asexual	Germany	Ludwigsburg	2005	*Quercus robur*

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0053	f	*Andricus quercuscalicis*	Asexual	Germany	Munich	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0054	f	*Andricus quercuscalicis*	Asexual	Germany	Munich	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0055	m	*Andricus quercuscalicis*	Asexual	Germany	Munich	2005	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0056	f	*Biorhiza pallida*	Sexual	Germany	Ludwigsburg	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0058	f	*Biorhiza pallida*	Sexual	Germany	Ludwigsburg	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0059	f	*Biorhiza pallida*	Sexual	Germany	Ludwigsburg	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0060	f	*Biorhiza pallida*	Sexual	Germany	Ludwigsburg	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0061	f	*Biorhiza pallida*	Sexual	Germany	Ludwigsburg	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0062	f	*Biorhiza pallida*	Sexual	Germany	Ludwigsburg	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0063	f	*Biorhiza pallida*	Sexual	Germany	Ludwigsburg	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0064	f	*Biorhiza pallida*	Sexual	Germany	Ludwigsburg	2006	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0065	f	*Callirhytis glandium*	Sexual	Hungary	Szentkut	?	*Quercus cerris*

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0066	f	*Andricus gallaetinctoriae*	Sexual	Hungary	Godollo	2002	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0067	f	*Andricus grossulariae*	Sexual	Hungary	Godollo	2002	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0068	f	*Andricus lucidus*	Sexual	Hungary	Szentkut	2002	*Quercus pubescens*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0069	f	*Neuroterus saliens*	Sexual	Hungary	Godollo	2001	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0070	m	*Callirhytis glandium*	Sexual	Hungary	Szentkut	2002	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0071	m	*Andricus caputmedusae*	Sexual	Hungary	Matrafured	2002	*Quercus pubescens*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0072	f	*Andricus hungaricus*	Asexual	Hungary	Godollo	2002	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca2]*	Cfun0073	m	*Biorhiza pallida*	Sexual	Hungary	Szentkut	2001	*Quercus pubescens*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0074	f	*Andricus crispator*	Sexual	Hungary	Matrafured	2001	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0075	m	*Biorhiza pallida*	Sexual	Hungary	Matrafured	2001	*Quercus petrea*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0076	m	*Andricus coriarius*	Asexual	Hungary	Matrafured	2000	*Quercus petrea*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0077	m	*Neuroterus saliens*	Sexual	Hungary	Matrafured	2001	*Quercus cerris*

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0078	m	*Andricus grossulariae*	Sexual	Hungary	Godollo	2001	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0079	f	*Andricus burgundus*	Sexual	Hungary	Godollo	2001	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0081	f	*Andricus quercuscalicis*	Asexual	Hungary	Godollo	2001	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0083	m	*Andricus quercuscalicis*	Asexual	Hungary	Sopron	2001	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0084	f	*Biorhiza pallida*	Asexual	Hungary	Szentkut	2001	*Quercus pubescens*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0086	m	*Andricus megalucidus*	?	Iran	?	?	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0087	m	*Chilaspis israeli*	?	Iran	?	?	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0088	m	*Andricus lucidus*	?	Iran	?	?	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0092	f	*Neuroterus saliens*	Asexual	Iran	?	?	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0097	f	*Andricus grossulariae*	?	Iran	Piran Shahr	?	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0098	f	*Aphelonyx persica*	?	Iran	Phars province	?	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0099	f	*Andricus lucidus*	?	Turkey	Aglasun	?	*Quercus* sp

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0112	f	*Andricus coriarius*	Asexual	Turkey	Gezende	?	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0113	f	*Andricus seckendorffi*	Asexual	Turkey	Madenli	?	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0115	f	*Andricus grossulariae*	Asexual	Turkey	Egirdir	?	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0116	m	*Andricus grossulariae*	Asexual	Turkey	Aglasun	?	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0117	f	*Andricus coriarius*	Asexual	Turkey	North of Antalya	?	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0119	m	*Andricus dentimitratus*	Asexual	Turkey	Beybesli	?	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0121	f	*Andricus lucidus*	Asexual	Italy	Massa Maritima	1998	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0122	f	*Andricus dentimitratus*	Sexual	Italy	Capalbio Cane	2005	*Quercus pubescens*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0123	m	*Biorhiza pallida*	Sexual	Italy	Volterra	1999	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0125	m	*Andricus coriarius*	Asexual	Italy	Piedimonte	?	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0126	f	*Andricus dentimitratus*	?	Italy	Capulbro Campolae	?	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0127	f	*Andricus dentimitratus*	Sexual	Italy	Capalbio Cane	2005	*Quercus pubescens*

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0128	m	*Andricus coriarius*	Asexual	Italy	Lame (Gazzo)	1998	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0129	f	*Andricus coronatus*	Asexual	Italy	Monte Sant'Angelo	2000	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0130	f	*Andricus lucidus*	Asexual	Italy	Monte Sant'Angelo	2000	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0131	f	*Andricus coronatus*	Asexual	Italy	Gildone	2000	*Quercus robur*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0132	f	*Andricus quercustozae*	Sexual	Spain	Puerto de Villatoro	2002	*Quercus pyrenaica*
Cecidostiba fungosa	*C. fungosa [Ca2]*	Cfun0133	f	*Biorhiza pallida*	Sexual	Spain	La Cañada	2002	*Quercus pyrenaica*
Cecidostiba fungosa	*C. fungosa [Ca2]*	Cfun0134	f	*Andricus quercustozae*	Sexual	Spain	Fresnedoso	2002	*Quercus pyrenaica*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0135	m	*Andricus quercustozae*	Sexual	Spain	Puerto de Villatoro	2002	*Quercus pyrenaica*
Cecidostiba fungosa	*C. fungosa [Ca2]*	Cfun0137	m	*Biorhiza pallida*	Sexual	Spain	El Escorial	2002	*Quercus pyrenaica*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0139	m	*Andricus quercustozae*	Sexual	Spain	Puerto de Villatoro	2002	*Quercus pyrenaica*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0140	f	*Andricus quercustozae*	Sexual	Spain	Puerto de Villatoro	2002	*Quercus pyrenaica*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0141	f	*Andricus quercustozae*	Sexual	Spain	Puerto de Villatoro	2002	*Quercus pyrenaica*

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0142	m	*Andricus quercustozae*	Sexual	Spain	El Escorial	2002	*Quercus pyrenaica*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0143	m	*Andricus quercustozae*	Sexual	Spain	Fresnedoso	2002	*Quercus pyrenaica*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0144	f	*Andricus quercustozae*	Sexual	Spain	Puerto de Villatoro	2002	*Quercus pyrenaica*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0146	f	*Andricus tomentosus*	Asexual	Greece	Arnissa	2000	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0147	f	*Andricus tomentosus*	Sexual	Greece	Edessa	2001	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0148	m	*Andricus coronatus*	Asexual	Greece	Edessa	2000	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0149	f	*Andricus tomentosus*	Asexual	Greece	Florina	2000	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0150	f	Unknown	Asexual	Greece	Arnissa	2000	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun0151	f	*Andricus coriarius*	?	Greece	Pisoderi	?	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun2897	m	*Aphelonyx persica*	Asexual	Iran	Piran Shahr	2002	*Quercus brantii*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun2900	m	*Andricus tomentosus*	Asexual	Iran	Ghelaie	?	*Quercus infectoria*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun2923	f	*Neuroterus lanuginosus*	Asexual	Iran	?	?	*Quercus brantii*

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun2931	f	*Aphelonyx persica*	?	Iran	?	?	*Quercus brantii*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun2955	f	*Andricus caputmedusae*	?	Iran	Bane	?	*Quercus infectoria*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun2961	f	*Synophrus syriacus*	?	Iran	Bane	?	*Quercus libani*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun2967	f	*Andricus megalucidus*	?	Iran	Ghelaie	?	*Quercus infectoria*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun2968	f	*Aphelonyx persica*	?	Iran	Javanrod	?	*Quercus brantii*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun2969	f	Unknown	?	Iran	?	?	*Quercus macranthera*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3018	f	*Neuroterus lanuginosus*	?	Iran	Khalkalsharaf	?	*Quercus brantii*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3026	f	*Andricus askewi*	?	Iran	Piran Shahr	2004	*Quercus infectoria*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3027	f	*Andricus grossulariae*	Asexual	Iran	Ghelaie	?	*Quercus infectoria*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3052	f	*Andricus caputmedusae*	Asexual	Turkey	Kirazoglu	2000	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3054	f	*Andricus seckendorffi*	Asexual	Turkey	Madenli	1998	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3055	m	*Andricus grossulariae*	Asexual	Turkey	Aglasun	1998	*Quercus* sp

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3059	f	*Andricus dentimitratus*	Asexual	Turkey	Beybesli	2000	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3510	m	*Andricus polycerus*	?	Iran	Marivan	?	*Quercus infectoria*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3511	m	*Andricus insana*	?	Iran	Marivan	?	*Quercus infectoria*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3519	m	*Aphelonyx cerricola*	?	Hungary	Márkó	2007	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3520	m	*Andricus multiplicatus*	?	Iran	Ghelaie	?	*Quercus brantii*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3521	m	*Chilaspis nitida*	Sexual	Hungary	Magyarkeszi	2008	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3522	m	*Chilaspis nitida*	Sexual	Hungary	Kópháza	2008	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3523	m	young acorn	Sexual	Hungary	Márkó	2008	*Quercus cerris*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3524	f	Unknown	?	Iran	Zagross Mountains	?	*Quercus brantii*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3525	m	*Pseudoneuroterus macropterus*	?	Iran	?	?	*Quercus brantii*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3526	f	*Pseudoneuroterus macropterus*	?	Iran	Taff	?	*Quercus brantii*
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3527	m	Unknown	?	Spain	?	?	*Quercus* sp

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3528	m	Unknown	?	Spain	?	?	*Quercus* sp
Cecidostiba fungosa	*C. fungosa [Ca1]*	Cfun3529	m	Unknown acorn gall	?	Hungary	Gyula	2008	*Quercus cerris*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem001	f	*Biorhiza pallida*	Sexual	UK	Silwood Park	2007	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem003	f	*Biorhiza pallida*	Sexual	UK	Silwood Park	2007	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem007	f	*Biorhiza pallida*	Sexual	UK	Silwood Park	2007	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem008	f	*Biorhiza pallida*	Sexual	UK	Puttenham	2007	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem009	f	*Biorhiza pallida*	Sexual	UK	Puttenham	2007	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem010	f	*Biorhiza pallida*	Sexual	UK	Puttenham	2007	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem011	f	*Biorhiza pallida*	Sexual	UK	Puttenham	2007	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem014	f	*Biorhiza pallida*	Sexual	UK	Puttenham	2007	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem016	f	*Biorhiza pallida*	Sexual	Austria	Ober Pullendorf	2006	*Quercus* sp
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem040	m	*Biorhiza pallida*	Sexual	Iran	Marivan	?	*Quercus infectoria*

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba semifascia	*C. semifascia [Ca2]*	Csem044	f	*Andricus cecconii*	Sexual	Iran	Ghelaie	?	*Quercus brantii*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem050	m	*Biorhiza pallida*	Sexual	Spain	El Escorial	2002	*Quercus pyrenaica*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem051	m	*Andricus quercuscalicis*	Sexual	UK	Puttenham	2006	*Quercus cerris*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem052	f	*Biorhiza pallida*	Sexual	France	Reims	2006	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem053	m	*Biorhiza pallida*	Sexual	France	Reims	2006	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem054	m	*Biorhiza pallida*	Sexual	France	Reims	2006	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem055	f	*Biorhiza pallida*	Sexual	France	Reims	2006	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem056	f	*Biorhiza pallida*	Sexual	France	Reims	2006	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem057	f	*Biorhiza pallida*	Sexual	France	Reims	2006	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem058	f	*Biorhiza pallida*	Sexual	France	Reims	2006	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem059	m	*Biorhiza pallida*	Sexual	France	Reims	2006	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem060	m	*Biorhiza pallida*	Sexual	Hungary	Godollo	2001	*Quercus robur*

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem061	f	*Biorhiza pallida*	Sexual	France	Reims	2006	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca1]*	Csem062	m	*Biorhiza pallida*	Sexual	Austria	Unterlois	2006	*Quercus robur*
Cecidostiba semifascia	*C. semifascia [Ca2]*	Csem073	m	*Pseudoneuroterus macropterus*	Sexual	Iran	Shurab	?	*Quercus brantii*
Cecidostiba semifascia	*C. semifascia [Ca2]*	Csem075	f	*Pseudoneuroterus macropterus*	Sexual	Iran	Ghelaie	?	*Quercus brantii*
Cecidostiba species A	*C. species A*	CspA003	f	Unknown small leaf gall	Sexual	Iran	Ghelaie	?	*Quercus brantii*
Cecidostiba species A	*C. species A*	CspA004	f	Unknown small leaf gall	Sexual	Iran	Ghelaie	?	*Quercus brantii*
Cecidostiba species A	*C. species A*	CspA005	f	Unknown small leaf gall	Sexual	Iran	Ghelaie	?	*Quercus brantii*
Cecidostiba species A	*C. species A*	CspA007	m	Unknown small leaf gall	Sexual	Iran	Ghelaie	?	*Quercus brantii*
Cecidostiba species A	*C. species A*	CspA008	m	Unknown small leaf gall	Sexual	Iran	Ghelaie	?	*Quercus brantii*
Cecidostiba species A	*C. species A*	CspA010	f	Unknown	?	Iran	?	?	*Quercus infectoria*
Cecidostiba species A	*C. species A*	CspA012	f	Unknown	?	Iran	Bane	?	*Quercus libani*
Cecidostiba species A	*C. species A*	CspA015	f	*Neuroterus saliens*	?	Iran	Ghelaie	?	*Quercus brantii*
Cecidostiba species A	*C. species A*	CspA020	m	Unknown small leaf gall	Sexual	Iran	Ghelaie	?	*Quercus brantii*

Table A5.1.
Continued

Morpho-Species	Inferred species	Code number	Sex	Host gall species	Host gall generation	Locality country	Locality site	Collection year	Host tree species
Cecidostiba species A	*C. species A*	CspA022	m	*Neuroterus saliens*	?	Iran	Ghelaie	?	*Quercus brantii*
Caenacis lauta		Clau32	m	*Andricus coriarius*	Asexual	Hungary	Matrafured	?	*Quercus petraea*

Appendix 5.2. EPIC loci PCR details

Details of the PCR recipes, conditions, and primers used for amplification of the 10 EPIC loci in Chapter 5 are provided below. These methods follow Lohse et al. (2011).

Recipe for 20 µl reaction:

2.0 µl 10· Bioline PCR buffer

2.0 µl bovine serum albumin (10 mg/mL)

0.8 µl $MgCl_2$ (50 mM)

0.16 µl dNTPs (25 mM each)

0.1 µl Taq Polymerase (5 U/µL, Bioline)

0.3 µl of each primer (20 µM)

1.0 µl of DNA template

13.34 µl milipure H_2O

PCR conditions:

A generic touchdown PCR protocol was used for all loci with an initial step of 94 °C for 3 minutes, followed by 10 cycles of 94 °C for 30 seconds, an annealing step of 40 seconds, and 72 °C for 1 minute, where the annealing temperature started at 65 °C and decreased by 1 °C each cycle, then a further 30 cycles with an annealing temp of 55 °C, and a final step of 72 °C for 10 minutes.

Table A5.2.1. Forward and reverse PCR primer sequences, used for the amplification of 10 EPIC loci following Lohse et al. (2011).

Locus	Primers	Forward	Reverse
AntSesB	40Fb/Rb	GCCAAYGTYATCMGDTACTTC	TACKGTRTCRAAKGGATAGGA
Ran	32F/R	TAYATTCARGGMCARTGYGC	GGRTCCATTGTRACTTCTGG
RpL15	2F/R	GGGTGCNACTTAYGGHAARC	GCGMAGYTCACGRTGYTTDTG
RpL37	27F/R	GAARGGTACNTCVAGYTTTGG	GACCRGTDCCRGTRGTCTTCCT
RpL37a	36F/R	CGHACVAAGAAGGTTGGAATCAC	GTYCTYTTGCAYCGYTTGC
RpS4	11F/R	BAARGCATGGATGTTRGACA	GGTCWGGRTADCGRATRGT
RpS8	5F/R	GAAGAGGAAGTWYGARTTRGGWC	TTCRTACCAYTGBCTGAADGG
RpS18	22F/R	GTYATGTTYGCYATGACNGC	KRAGRCCCCAGTARTGWCG
RpS23	21F/R	ACVMGVTGGAAGGCYAATCC	ATGACCYTTACGHCCRAATCC
Sansfille	35F/R	CHWTVAAAATGCGTGGWCAAG	CDGGGAAYTGATTRAACARCAT