Molecular Ecology (2003) **12**, 2153–2174 doi: 10.1046/j.1365-294X.2003.01894.x

Out of Anatolia: longitudinal gradients in genetic diversity support an eastern origin for a circum-Mediterranean oak gallwasp *Andricus quercustozae*

ANTONIS ROKAS,*¶ RACHEL J. ATKINSON,*† LUCY M. I. WEBSTER,‡ GYÖRGY CSÓKA§ and GRAHAM N. STONE*

Institute of Cell, Animal and Population Biology, Ashworth Laboratories, University of Edinburgh, West Mains Road, King's Buildings, Edinburgh EH9 3JT, UK, †Department of Biology, University of Stirling, Stirling FK9 4LA, UK, ‡Division of Environmental and Evolutionary Biology, University of Glasgow, Graham Kerr Building, Glasgow G12 8QQ, UK, §Hungarian Forest Research Institute, Mátrafüred Research Station, 3232 Mátrafüred, Hungary, ¶Howard Hughes Medical Institute and Laboratory of Molecular Biology, R.M. Bock Laboratories, University of Wisconsin-Madison, 1525 Linden Drive, Madison, WI 53706–1596, USA

Abstract

Many studies have addressed the latitudinal gradients in intraspecific genetic diversity of European taxa generated during postglacial range expansion from southern refugia. Although Asia Minor is known to be a centre of diversity for many taxa, relatively few studies have considered its potential role as a Pleistocene refugium or a potential source for more ancient westward range expansion into Europe. Here we address these issues for an oak gallwasp, *Andricus quercustozae* (Hymenoptera: Cynipidae), whose distribution extends from Morocco along the northern coast of the Mediterranean through Turkey to Iran. We use sequence data for a fragment of the mitochondrial gene cytochrome *b* and allele frequency data for 12 polymorphic allozyme loci to answer the following questions: (1) which regions represent current centres of genetic diversity for *A. quercustozae*? Do eastern populations represent one refuge or several discrete glacial refugia? (2) Can we infer the timescale and sequence of the colonization processes linking current centres of diversity? Our results suggest that *A. quercustozae* was present in five distinct refugia (Iberia, Italy, the Balkans, southwestern Turkey and northeastern Turkey) with recent genetic exchange between Italy and Hungary. Genetic diversity is greatest in the Turkish refugia, suggesting that European populations are either (a) derived from Asia Minor, or (b) subject to more frequent population bottlenecks. Although Iberian populations show the lowest diversity for putatively selectively neutral markers, they have colonized a new oak host and represent a genetically and biologically discrete entity within the species.

Keywords: Anatolia, *Andricus*, gallwasp, glacial refugia, *Quercus*, range expansion

Received 17 November 2002; revision received 2 May 2003; accepted 2 May 2003

Introduction

Many studies of a wide range of taxa have examined the latitudinal gradients in species richness and intraspecific genetic diversity generated by recent postglacial range expansion from southern refugia (Taberlet *et al*. 1998; Hewitt 1999, 2000). Many European species are native to more than one refugial region and refuge-specific genetic polymorphism has been used extensively in identifying putative origins for postglacial colonization of higher latitudes (Taberlet *et al*. 1998; Hewitt 1999). In general, however, there has been far less consideration of the pattern and timescale of *longitudinal* patterns in intraspecific genetic diversity. This is of particular significance, because while latitudinal trends in intraspecific genetic variation typically involve subsets of refugial diversity, in many species longitudinal variation involves transitions between discrete sets of genotypes. Significant populations of many western Palaearctic species are found to the east of Europe,

Correspondence: Graham Stone. Fax: + 44 131 650 6564; E-mail: graham.stone@ed.ac.uk

AR, RJA and GNS contributed equally to the preparation of this manuscript.

in Turkey, the Caucasus, Iran and the Levant (Cooper *et al*. 1995; King & Ferris 1998; Hewitt 1999). Although these more easterly populations are rarely considered, they may not only represent significant centres of genetic diversity, but also the potential origin of populations now occupying Europe. Estimates based on sequence divergence suggest that, in contrast to latitudinal gradients, longitudinal differentiation among refugial populations commonly predates the Pleistocene (reviewed in Taberlet *et al*. 1998; Hewitt 1999). The greater age of such longitudinal patterns increases the potential for significant evolutionary divergence between refugial populations. Here we examine the relationship among southern European refugia and illustrate the significance of Turkish populations for intraspecific diversity in an oak feeding insect — the gallwasp *Andricus quercustozae*.

Oak gallwasps are obligate parasites of their hosts, developing in galls induced on specific plant tissues (Stone *et al*. 2002). As such, their distributions are intimately dependent on those of their hosts, and these insects provide an appropriate model for the phylogeography of many invertebrates with similarly specific host associations. Western Palaearctic oaks are members of two sections — the section Cerris (e.g. *Q. ilex*, *Q. cerris*, *Q. suber*), and the section Quercus *sensu stricto* (e.g. *Q. robur*, *Q. petraea*, *Q. pubescens*) (Jalas & Suominen 1987). Phylogeographical analyses of oaks in both sections suggest that genetic divisions between oak populations in southern European refugia date from at least the beginning of the Pleistocene (Ferris *et al*. 1993; Ferris *et al*. 1995; Dumolin-Lapegue *et al*. 1997; Petit *et al*. 1997; Toumi & Lumaret 1998). Far less is known about populations east of the Balkans. This is significant because the ranges of several widespread European oaks (e.g. *Q. robur, Q. pubescens, Q. petraea* and *Q. infectoria*) extend eastwards through Turkey into the mountains of northern Iran, northwards into the Caucasus and south into Syria and Lebanon (Sternlicht 1968; Chodjai 1980; Yaltirik 1982). High levels of oak species endemism and the identification of numerous endemic subspecies and races suggest that both Turkey and the Caucasus may have provided discrete refugia for oaks during this period (Turrill 1929; Konarov 1936; Zohary 1966; Townsend & Guest 1980; Yaltirik 1982; Castroviejo 1986; Camus 1936–54), and the same regions represent both the current centre of diversity and the probable origin of the Western Palaearctic radiations of both host oak sections (Govaerts & Frodin 1998; Manos *et al*. 1999).

Most oak gallwasps have a cyclically parthenogenetic lifecycle, involving alternation between one sexual and one parthenogenetic generation each year (Atkinson 2000; Stone *et al*. 2002). *A. quercustozae* is a member of a large clade of *Andricus* species whose lifecycle also involves host alternation: the sexual generation develops in galls induced on oaks in the section Cerris, and the parthenogenetic generation only galls oaks in the section Quercus *sensu stricto* (Cook *et al*. 2002; Stone *et al*. 2002). Phylogenetic analysis of host associations in *Andricus* strongly supports a single evolution of the host-alternating lifecycle well before the start of the Pleistocene (Cook *et al*. 2002), and contemporary patterns of species richness suggest that, like their oak hosts, host-alternating gallwasps also diversified in Turkey and the Balkans (Stone *et al*. 2002).

At least 10 host-alternating *Andricus*, including *A. quercustozae*, have distributions extending from Morocco eastwards to Iran, with populations in all the major oak refugia (Atkinson 2000; Melika *et al*. 2000; Nieves-Aldrey 2001; Cook *et al*. 2002). A significant geographical pattern in oak distributions for these gallwasps is an east–west divide in their sexual generation hosts. In the Iberian peninsula and northwestern Africa, the sexual generation galls develop on cork oak, *Quercus suber*, while from Italy east to Turkey the dominant host is turkey oak, *Q. cerris* (Stone *et al*. 2001). The distributions of these two hosts, and of gallwasp populations exploiting them, meet north of the Pyrenees (Jalas & Suominen 1987; Stone *et al*. 2001). In a related gallwasp, *A. kollari*, populations associated with these two hosts represent genetically discrete ecotypes, and Iberian populations represent a single lineage that diverged from populations associated with *Q. cerris* well before the Pleistocene (Stone *et al*. 2001). The rarity of host shifts between oak sections is also supported by a genus-wide analysis of *Andricus* gallwasps (Cook *et al*. 2002). If this pattern is general, we expect a similar clear divide between populations of *A. quercustozae* associated with the two section Cerris oaks.

Here we use a combination of mitochondrial sequence and allozyme allele frequency data for populations of *A. quercustozae* ranging from Morocco to eastern Turkey to address the following specific questions.

1 What is the pattern and extent of refugial differentiation in *A. quercustozae?* Does Turkey contain significant genetic diversity absent from Europe?
2 Can we identify a putative geographical origin for *A. quercustozae*?
3 What is the timescale of longitudinal range expansion in *A. quercustozae*?
4 How frequent have host shifts been between *Q. suber* and *Q. cerris*?
5 A final question concerns the genetic basis of gall phenotype, which in *A. quercustozae* shows longitudinal variation. Asexual generation galls induced by all the Turkish populations sampled are bright red and very sticky when mature, while the galls induced by all the more western populations have a matt brown surface. Although the adults reared from the two forms do not differ significantly (Melika *et al*. 2000), the insects inducing the eastern gall type have sometimes been regarded as a separate species, *A. insana* (Dalla Torre & Kieffer 1910;

Chodjai 1980). If this difference in gall phenotype is indicative of a genuine split between gallwasp lineages, we expect the two phenotypes to represent monophyletic groups within the *A. quercustozae* phylogeny. This possibility is examined using mitochondrial sequence data.

Materials and methods

Collection of specimens

Parthenogenetic generation galls of *A. quercustozae* were collected at sites through the species' range (Table 1, Fig. 1) from Morocco to northeastern Turkey, corresponding to 50.2 degrees of longitude and a transect length of over 4000 km. Parthenogenetic females were reared from their galls under quarantine in Edinburgh, and were frozen from life and stored at –80 °C until required.

DNA extraction and sequencing

DNA was extracted using the DNeasy Tissue kit (Qiagen cat. 69504), following the manufacturer's protocol for insect DNA extraction. A 433 base pair (bp) fragment of the mitochondrial cytochrome b gene was amplified and sequenced for 47 individuals using previously described protocols (Stone & Cook 1998). Sample sizes for each population are shown in Table 1. Sequencing was carried out using Perkin-Elmer BigDye Terminator chemistry and an ABI 377 sequencer. Polymerase chain reaction (PCR) products for each specimen were sequenced in both directions to minimize PCR artefacts, ambiguities and base-calling errors. Chromatograph output was checked by eye. Following the detection of multiple sequences in PCR products for Turkish specimens from Gezende and Lysandra (see Results), PCR products were cloned in a blunt-end vector following the manufacturer's instructions (Zero Blunt® TOPO® Cloning kit, Invitrogen, cat. K4500-01). Ten colonies from each individual were selected, cultured and purified (QIAprep spin miniprep kit, Qiagen cat. 27104) for subsequent sequencing. All sequences are deposited in GenBank (Accession nos AY157269–AY157298).

Phylogenetic analysis of haplotype data

Sequences were aligned manually. Phylogenetic analysis was performed with maximum-likelihood and Bayesian inference algorithms, which allow hypothesis testing in a statistical framework (Huelsenbeck *et al*. 2001; Lewis 2001) and description of important aspects of sequence evolution.

Maximum likelihood. Phylogenetic relationships were estimated by maximum likelihood (ML), using the PAUP* package, version 4.0b10 (Swofford 2002). The best-fit ML model of sequence evolution for each data set was identified using likelihood ratio tests as implemented by MODELTEST, version 3.0 (Posada & Crandall 1998). Parameters allowed to vary in model-fitting were base composition, substitution rates and rate heterogeneity among sites. Data sets with and without the haplotypes from the two Turkish wasps that gave multiple products, and with and without outgroups, gave very similar best-fit models of sequence evolution (data not shown). For the full *A. quercustozae*

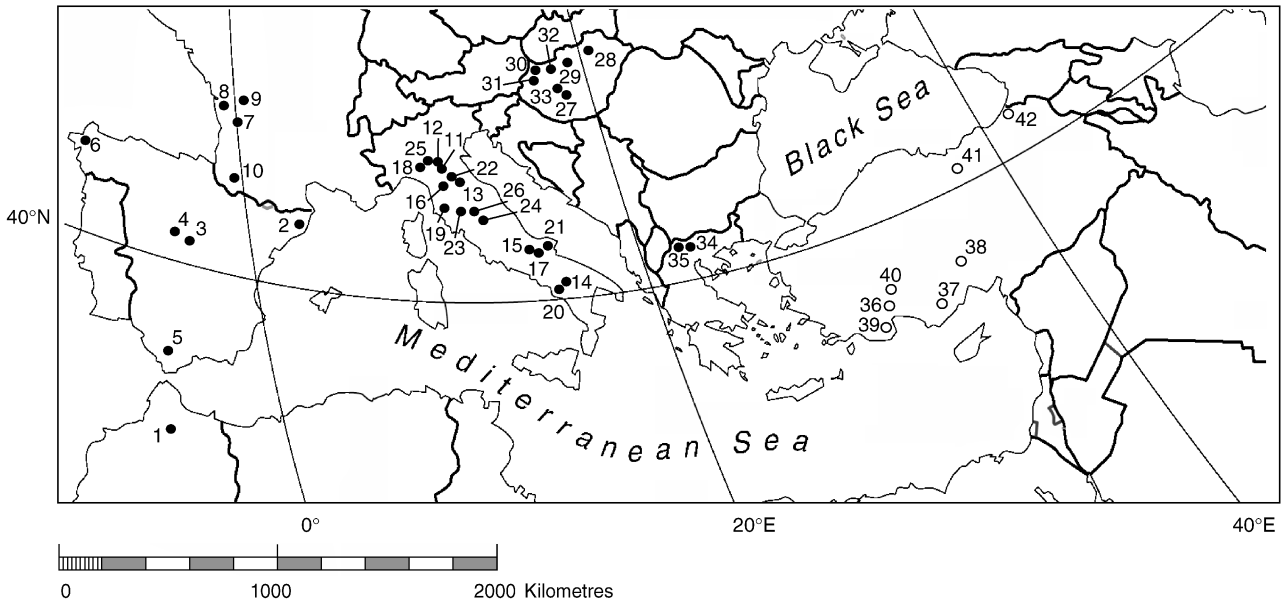

Fig. 1 The location of sample sites for *Andricus quercustozae*. Filled circles indicate populations showing the brown, nonsticky *quercustozae* asexual gall penotype, while empty circles indicate populations showing the red, sticky *insana* phenotype.

Table 1 Sampling locations and number of individuals sequenced for cytochrome *b* or electrophoresed for allozymes from each population. Location is given as decimal degrees latitude and longitude (negative values are west or south). A = Allelic diversity (mean number of alleles per locus), H = mean heterozygosity per locus expected from observed allele frequencies (±1 standard error). *In the genetic diversity column indicates that data were not available for this site

			Sample size		Genetic diversity	
No., Population	Country	Location	Allozymes	Haplotypes	H	A
1. Azrou	Morocco	33.45, –05.23	39	2	0.07 (0.02)	1.42
2. Barcelona	Spain	41.42, 02.17	0	1	*	*
3. Madrid	Spain	40.42, –03.72	9	1	0.06 (0.05)	1.25
4. Navacerrada	Spain	40.72, –04.02	0	1	*	*
5. Prado del Rey	Spain	36.80, –05.55	17	1	0.07 (0.03)	1.58
6. Santiago de Compostela	Spain	42.87, –08.55	0	1	*	*
7. Aire de Querane	France	45.37, –00.97	0	1	*	*
8. Bordeaux	France	44.83, –00.57	12	1	0.03 (0.03)	1.08
9. Perigeux	France	45.12, 00.73	0	1	*	*
10. Tarbes	France	43.23, 00.08	9	1	0.02 (0.03)	1.08
11. Bombiana	Italy	44.20, 10.95	16	1	0.12 (0.05)	1.58
12. Casina	Italy	44.52, 10.50	0	1	*	*
13. Chiusi	Italy	43.03, 11.95	0	1	*	*
14. Felitto	Italy	40.37, 15.25	0	1	*	*
15. Gildone	Italy	41.50, 14.67	0	1	*	*
16. Greve in Chianti	Italy	43.58, 11.32	39	1	0.11 (0.03)	1.92
17. Jelsi	Italy	41.53, 14.80	28	1	0.13 (0.04)	1.83
18. Lame	Italy	44.63, 09.70	0	1	*	*
19. Massa Marritimo	Italy	43.05, 10.88	27	1	0.11 (0.03)	1.42
20. Moio	Italy	40.15, 15.17	28	0	0.12 (0.03)	1.92
21. Monte Sant' Angelo	Italy	41.72, 15.97	40	0	0.15 (0.03)	1.67
22. Poppi	Italy	43.72, 11.77	39	1	0.12 (0.03)	1.75
23. Radicofani	Italy	42.90, 11.77	0	1	*	*
24. Rieti	Italy	42.40, 12.85	22	1	0.14 (0.04)	1.75
25. Salsomaggiore	Italy	44.19, 09.62	0	1	*	*
26. San Venanzo	Italy	42.87, 12.27	27	1	0.12 (0.03)	1.42
27. Lake Balaton	Hungary	47.10, 17.90	0	1	*	*
28. Eger	Hungary	47.88, 20.47	0	1	*	*
29. Gödöllő	Hungary	47.60, 19.33	0	2	*	*
30. Sopron	Hungary	47.67, 16.58	40	1	0.06 (0.02)	1.67
31. Szeghalom	Hungary	47.23, 16.70	39	1	0.05 (0.01)	1.58
32. Tatabanya	Hungary	47.52, 18.42	10	1	0.03 (0.02)	1.25
33. Veszprem	Hungary	47.1, 17.90	40	1	0.06 (0.01)	1.75
34. Arnissa	Greece	40.46, 21.56	8	3	0.12 (0.06)	1.67
35. Pisoderi	Greece	40.46, 21.13	0	2	*	*
36. Ağlasun	SW Turkey	37.65, 30.53	15	0	0.25 (0.07)	1.83
37. Gezende	SW Turkey	36.53, 33.15	40	1 (cloned)	0.25 (0.04)	2.83
38. Küllüce	C Turkey	38.20, 34.60	7	2	0.26 (0.11)	2.08
39. Lysandra	SW Turkey	36.48, 30.05	6	1 (cloned)	0.26 (0.12)	1.92
40. Madenli	SW Turkey	38.13, 31.02	40	0	0.30 (0.04)	2.75
41. Refahiye	NE Turkey	39.90, 38.75	7	3	0.27 (0.11)	1.92
42. Yeniyol	NE Turkey	41.40, 41.63	5	2	0.24 (0.12)	1.67

haplotype data set discussed below, the best fit model incorporated unequal base frequencies, one rate for transitions and one rate for transversions (ti/tv ratio = 6.9) as well as among-site variation (proportion of invariable sites = 0.72, shape parameter α of the Γ distribution = 0.84). The parameter values suggested by MODELTEST were used to search for the ML topology, using a heuristic search with TBR swapping and 100 random-taxon-addition replications. Bootstrap values were generated by a heuristic search with tree bisection and reconnection on 100 pseudoreplicate data sets using the same parameter values.

Bayesian inference. Inference of phylogenetic relationships and calculation of posterior probabilities for the branches

separating *A. quercustozae* haplotypes were carried out using the program MRBAYES, version 3.0B (Huelsenbeck & Ronquist 2001). In MRBAYES, sampling of trees from the posterior probability distribution is achieved by implementation of the Metropolis-coupled Markov chain Monte Carlo algorithm — $(MC)^3$ for short. The $(MC)^3$ algorithm allows running of multiple Markov chains. A run with four chains was performed for 2 000 000 generations, under a general time-reversible model (all six types of substitution occurring at different rates) with parameter value estimation for base frequencies, substitution matrix values and rate heterogeneity. Rate heterogeneity was estimated both by using a gamma distribution for the variable sites and by assuming a certain portion of sites to be invariable. The burn-in time was 100 000 generations.

Tree rooting. For identification of early branching clades within *A. quercustozae*, six species were selected as outgroups on the basis of their position within the *Andricus* phylogeny (Stone & Cook 1998; Cook *et al*. 2002; Rokas *et al*. 2003). Four closely related outgroups were selected from within the species group containing *A. quercustozae* (*A. quercuscalicis*: AJ228459; *A. coronatus*: AJ228461; *A. dentimitratus*: AJ228450; and *A. caputmedusae:* AJ228456) together with two more distantly related outgroups (*A. conificus*: AJ228460; and *A. conglomeratus:* AJ228468).

Alignments and topologies reported in this study are available electronically from TreeBASE (http://www.herbaria.harvard.edu/treebase/, TreeBASE Study Accession no. S887).

Allozyme data generation and analysis

Allozyme electrophoresis. A total of 609 individuals were screened at 12 variable allozyme loci using cellulose acetate gel electrophoresis (Zip Zone, Helena Laboratories) and methods described in full in Stone & Sunnucks (1993) and Stone *et al*. (2001). Sample sizes for each site are shown in Table 1. The loci screened were αGPD 1 and 2 (EC 1.1.1.8), GOT-m and -s (EC 2.6.1.1) HK (EC 2.7.1.1), PGM (EC 2.7.5.1), MDHm (EC 1.1.1.37), ME (EC 1.1.1.40), 6PGD (EC 1.1.1.44), AK (EC 2.7.4.3), GPI (EC 5.3.1.9), PEPb (EC 3.4.11). Running buffers for each system, as described in Richardson *et al*. (1986), were Tris glycine pH 8.6 for HK and PGM, Tris-EDTA-maleate-$MgCl_2$ pH 8.3 for 6PGD, GPI, MDHm, ME and GOT and phosphate buffer pH 6.3 for PEPb, αGPD and AK.

Allozyme data analysis. Allele frequencies for all sites and loci are shown in Appendix I. Genotypic data were tested for deviations from Hardy–Weinberg equilibrium (HWE) and linkage equilibrium (LD) using the multisample test in GENEPOP (Raymond & Rousset 1995). Dunn–Sidak correction with an experiment-wise error rate of 0.01 was used to correct for multiple comparisons. There was no evidence for significant deviation from Hardy–Weinberg equilibrium, or evidence for linkage disequilibrium between loci, in any population, and both equilibria were assumed for the STRUCTURE analysis below. Two measures of genetic diversity — mean heterozygosity per locus calculated from observed allele frequencies (H_E) and allelic diversity (the average number of alleles per locus) were calculated for each population using GENEPOP (Raymond & Rousset 1995) and GENETIX (Belkhir *et al*. 2000). These indices allow direct comparison with other published studies on gallwasps (Stone & Sunnucks 1993; Stone *et al*. 2001).

Analyses of population substructure. Similarities in gene frequencies among populations were analysed using three different clustering methods.

(a) Pairwise F_{ST}. F_{ST} values between sites were calculated using the methods of Weir & Cockerham (1984). Permutation tests were used to determine whether the values of theta obtained were significantly different from those expected for random allocation of data for two populations into two groups. Dunn–Sidak correction with an experiment-wise error rate of 0.01 was used to correct for multiple comparisons.

(b) Population-based pairwise-distance matrix method. Relationships between populations were analysed using three different criteria (neighbour-joining, least squares and maximum likelihood) on two different pairwise distance measures; Nei's genetic distance (Nei 1972) and Cavalli-Sforza's chord measure (Cavalli-Sforza & Edwards 1967). For neighbour-joining and least squares, and on allele frequency data for maximum likelihood, 100 bootstrap replicates were generated for each method using PHYLIP (Felsenstein 1993). The topologies returned by the different approaches were extremely similar (data not shown), and for brevity we present the topology produced using the least squares criterion on Cavalli-Sforza's chord distance (Cavalli-Sforza & Edwards 1967).

(c) Individual-based modelling method. The number of discrete clusters of genotypes in *A. quercustozae* was determined from the data (i.e. without geographical preconception) using the programme STRUCTURE (Pritchard *et al*. 2000). STRUCTURE assumes a model in which a specified number of genotype pools are characterized by a set of allele frequencies derived from multilocus genotype data. Individuals are assigned probabilistically to the pools under the assumptions of Hardy–Weinberg and linkage equilibrium (shown above to be appropriate for our data) using Markov Chain Monte Carlo (MCMC) simulation. The simulation is rerun for models specifying different numbers of pools (K) and the posterior probabilities for

each simulation are compared to infer the pool number that is best supported by the underlying genotypic data.

We determined simulation run-lengths for $K = 1–5$ by varying the number of replications between 1×10^3 and 1×10^6 in the burn-in period and the actual run, and looking for convergence in the estimated parameter values for 3–10 runs at each K. The simulations were run using only individuals from the putative refuge areas. Having determined the best estimate of K, the simulation was rerun to include individuals from French populations (Tarbes and Bordeaux) to determine their most likely origin. Structure allows fitting of models with or without admixture, the latter allowing genotypes to arise through mating between individuals derived from different populations. The results with and without admixture were indistinguishable (as indicated by the small value of the Dirichlet parameter) and we report only the results without admixture.

Results

Analyses of haplotype sequence data

Sequence variation. PCR amplification and sequencing produced a single sequence in 45 of 47 individuals, and a total of 25 haplotypes (Table 2). Of the 433 nucleotide sites in these 25 haplotypes, 35 (8.08%) were polymorphic and 24 (5.54%) were parsimony-informative. For two specimens from southwestern Turkey (Gezende and Lysandra), multiple cytochrome *b*-like sequences were recovered. Six distinct 'haplotypes' of correct length and reading frame (one of which was shared by both individuals) were recovered by sequencing of 10 clones from each of these two specimens. Although sequence divergence among these haplotypes was high both within (uncorrected: 0.5–4.9%; ML-estimated: 0.5–8.4%) and between these two individuals (uncorrected: 0.2–5.3%; ML-estimated: 0.2–8.5%), this group of haplotypes were more similar to each other than to any other Turkish sequences (uncorrected: 4.4–7.6%; ML-estimated: 9.2–18.3%). There are two plausible alternative explanations for the existence of multiple cytochrome-*b*-like sequences in these specimens. First, one of the sequences is the true mitochondrial copy, while the others are nuclear pseudogenes (numts) (Bensasson *et al*. 2001). Second, all the sequences are recently derived pseudogenes from a true mitochondrial copy (given that they are not, as yet, mutationally degenerated). Alternative explanations for the existence of multiple mitochondrial-like sequences (including heteroplasmy and intramitochondrial duplications (Mirol *et al*. 2000; Bensasson *et al*. 2001) are unlikely for reasons discussed in detail by Mirol *et al*. (2000). Although our data do not allow formal discrimination between these two hypotheses, we suggest that (i) the absence of multiple sequences for any other specimens, (ii) the far higher copy number of mitochondrial than nuclear templates and (iii) the rarity of numts without indels and/or stop codons in other studies (e.g. Bensasson *et al*. 2000; Mirol *et al*. 2000) including work on gallwasps (Rokas *et al*. 2001, 2003; Stone *et al*. 2001) make it likely that these sequences are derived from truly distinct regional haplotypes (or alternatively, one of the sequences is the 'true' mitochondrial haplotype). This interpretation is supported by concordance between the phylogenetic patterns seen when these sequences are included, and patterns seen in the allozyme data (see below). Nevertheless, because the possibility that all of the sequences are numts cannot be excluded, we refer to them as pseudohaplotypes (Table 2). Because the alternative sequences for this pair of individuals are more similar to each other than to those from any other locations, the same phylogeographical inferences result from inclusion of any single sequence from each of them.

A. quercustozae haplotypes show disjunct distributions corresponding to recognized refugial regions (Table 2): of the 25 haplotypes, six were restricted to the region corresponding to the natural distribution of *Q. suber* in Iberia, Morocco and southwestern France. Within the native range of *Q. cerris*, seven haplotypes (7–10 and 12–14) were restricted to Italy, seven to Greece and the Balkans (15–21) and four (not including the pseudohaplotypes) to Turkey (22–25). Assumption that one of the pseudohaplotypes is the correct sequence for the two individuals from southwestern Turkey increases the number of haplotypes specific to this region to a minimum of five.

Genetic differentiation among regions in A. quercustozae. ML and Bayesian phylogenetic inference of the haplotype sequence data support the existence of four clades within *A. quercustozae* (Fig. 2a): clade 1 (haplotypes 1–6) contains all the individuals from Iberia and Morocco, while clade 2 (haplotypes 7–23) contains one French individual, all the individuals from Italy, Hungary and Greece and two from Central Turkey. Clade 3 contains pseudohaplotypes 1–6 from southwestern Turkey, while clade 4 (haplotypes 24–25) contains two individuals from northeastern Turkey. If we accept that one of the pseudohaplotypes is the true mitochondrial copy for each of the specimens in clade 3, then this topology suggests a deep split between southwestern and northeastern Turkish lineages.

In addition to the subdivision shown by these haplotypes, there is regional variation in the sequence diversity they represent, in terms of maximum divergence between any two sequences from a region, and the nucleotide diversity across all sites from a region. The greatest diversity is present in Turkey (uncorrected divergence: 0.2–4.2%; ML-estimated: 0.2–6.0%; pseudohaplotypes not included in calculations), with lower divergence and diversity in the Balkans (uncorrected: 0.2–1.4%; ML-estimated: 0.2–1.6%), Italy (uncorrected: 0.2–0.7%; ML-estimated: 0.2–0.8%) and Iberia (uncorrected: 0.2–1.0%; ML-estimated: 0.2–1.1%).

Table 2 Haplotypes for the 433 bp fragment of cytochrome *b* from *Andricus quercustozae*. Each column corresponds to a variable nucleotide site. The number above each column indicates the position of the nucleotide site along the 433 bp cytochrome *b* fragment. The parsimony informative sites are indicated by an asterisk in the first row of the table

	11111	1111111122	2222222222	2233333333	3333333334	44
	11124445	7788900011	4678899900	0122223456	6702222223	3345677881	22
	1703950365	0425107928	2467801625	8403692105	8740125892	5708469255	67
Parsimony-informative?	**** ** ** * *** ***	*	** *	* * **	******* **	** ** ***	
Haplotype 1	TAGCATATAG	TTGGATCTTT	GTATTAAATT	TATATATGTA	TATGTTATGT	GATATAATTT	TA
Haplotype 2	....	....	..G.......	....	....	A.........	..
Haplotype 3	.......C..	....	....	....	....	....C.....	..
Haplotype 4	....	....	....	....	........A.	....	..
Haplotype 5	.........A	....	....	....	....	....	..
Haplotype 6	???.......	...A......	....	....	........A.	....	..
Haplotype 7	..AT.....A	..AAG.TA..	A......G..	......C...	......G...	.T........	..
Haplotype 8	?.AT.....A	..AAG.TA..	A....G.G..	......C...	......G...	.T........	..
Haplotype 9	..AT......	..AAG.TA..	A......G..	......C...	......G.A.	.T........	..
Haplotype 10	..AT.....A	..AAG.TA..	A......G..	......C...	......G.A.	.T........	..
Haplotype 11	..AT......	..AAG.TA..	A......G..	......C...	......G...	.T........	..
Haplotype 12	..AT......	..AAG.TA..	A......G..	..C...C...	......G...	.T........	..
Haplotype 13	..AT.....C	..AAG.TA..	A......G..	......C...	....	.T........	..
Haplotype 14	..AT.....C	..AAG.TA..	A......G..	....	......G...	.T........	..
Haplotype 15	..AT.....C	..AAG.TA..	A......G..	......C...	......G...	.T........	..
Haplotype 16	C.AT.....T	..AA..TA..	A......G..	......C...	......G...	.T........	..
Haplotype 17	C.AT.....C	..AA..TA..	A......G..	......C...	......G...	.T........	..
Haplotype 18	..AT.....T	..AA..TA..	AC.....G..	......C...	......G...	.T........	..
Haplotype 19	..AT.....C	..AA..TA..	A......G..	......C...	......G...	.T........	..
Haplotype 20	..AT.....C	..AA..TA..	A......G..	......C...	......G.A.	.TC.......	..
Haplotype 21	..AT......	.CAAG.TA..	A......G..	......C...	......G.A.	.T.G......	..
Haplotype 22	..AT......	..A.G.TA..	A.........	....	.....CG...	.TC.......	..
Haplotype 23	..AT......	..AAG.TA..	A.........	....	.....CG...	.TC.......	..
Haplotype 24	..AT.....A	..AA....C.	A.........	....	C..AC...A.	.T....C...	..
Haplotype 25	...TG....A	..AA....C.	A......G.C	....C...C.	C..AC.G.A.	.TC.C.....	..
Pseudohaplotype 1	C..T.C..T.	A.AAT....C	A.....CG.C	.T.....T..	.GC...T...	.T...T....	..
Pseudohaplotype 2	C..T.C...A	A...GC....	A......G.C	.T.GCG.C..	.GC...TC..	.T..C.....	..
Pseudohaplotype 3	CG.T.CT.TA	A...G.....	A......G.C	.T.....T..	.GC...T..C	.TC..C.CCC	..
Pseudohaplotype 4	C..T.C..T.	A.AAT.....	A.....CG.C	.T.....T..	.GC...T...	.T...T....	..
Pseudohaplotype 5	CG.T.CT.T.	A...G.....	A..CC..GCC	CT.....C.G	..C...T.AC	.TC..T.CCC	..
Pseudohaplotype 6	C..T.C..T.	A.AAT.....	A.....CG.C	.T.....T..	.GC...T...	.T...T....	AC

Haplotype locations: haplotypes 1–6 are from Iberia, Morocco and France, haplotypes 7–14 are mainly from Italy, haplotypes 15–17 from Greece, haplotypes 18–21 from the Balkans and haplotypes 22–25 from Turkey. 1 — Barcelona, Prado del Rey, Santiago de Compostela (S), Tarbes (F); 2 — Madrid (S); 3 — Azrou (M); 4 — Azrou (M); 5 — Bordeaux (F); 6 — Perigeux (F); 7 — Lame (I); 8 — Casina (I); 9 — Chiusi, Salsomaggiore, San Venanzo (I); 10 — Jelsi (I); 11 — Aire de Querane (F), Felitto, Greve, Poppi, Radicofani, Rieti (I), Gödöllö, Sopron, Szeghalom (H); 12 — Bombiana (I); 13 — Greve in Chianti (I); 14 — Gildone, Massa Marritimo (I); 15 — Arnissa (G); 16 — Pisoderi (G); 17 — Pisoderi (G); 18 — Lake Balaton (H); 19 — Eger, Gödöllö, Tatabanya (H), Arnissa (G); 20 — Veszprem (H); 21 — Sopron (H); 22 — Küllüce (T); 23 — Küllüce (T); 24 — Yeniyol (T); 25 — Refahiye (T). Pseudohaplotype locations: 1, 2, 3 — Lysandra (T); 2, 4, 5, 6 — Gezende (T). Letters in parentheses indicate the country of origin (M: Morocco, S: Spain, F: France, I: Italy, H: Hungary, G: Greece, T: Turkey).

Two further conclusions emerge from Fig. 2a. First, populations able to exploit *Q. cerris* in their sexual generation are represented by at least two lineages (clades 2 and 3), while populations able to exploit *Q. suber* in Iberia, Morocco and southwestern France possess haplotypes belonging to a single lineage (clade 1). Second, galls showing the eastern *insana* gall morphology (represented here by all the Turkish haplotypes) do not form a monophyletic group. Haplotypes 22 and 23 from central Turkey cluster within a large clade of sequences from *quercustozae* gall morphs in clade 2. The posterior probability from the Bayesian analysis that all the Turkish haplotypes (and so the *insana* gall morphology) indeed represent a monophyletic group is extremely low ($P < 0.0001$).

When did lineages in these regions diverge? If we use the widely applied approximation of 2.3% mitochondrial sequence divergence per million years (Brower 1994), the pairwise sequence distances between members of the clades in Fig. 2a can be used to provide an approximate estimate

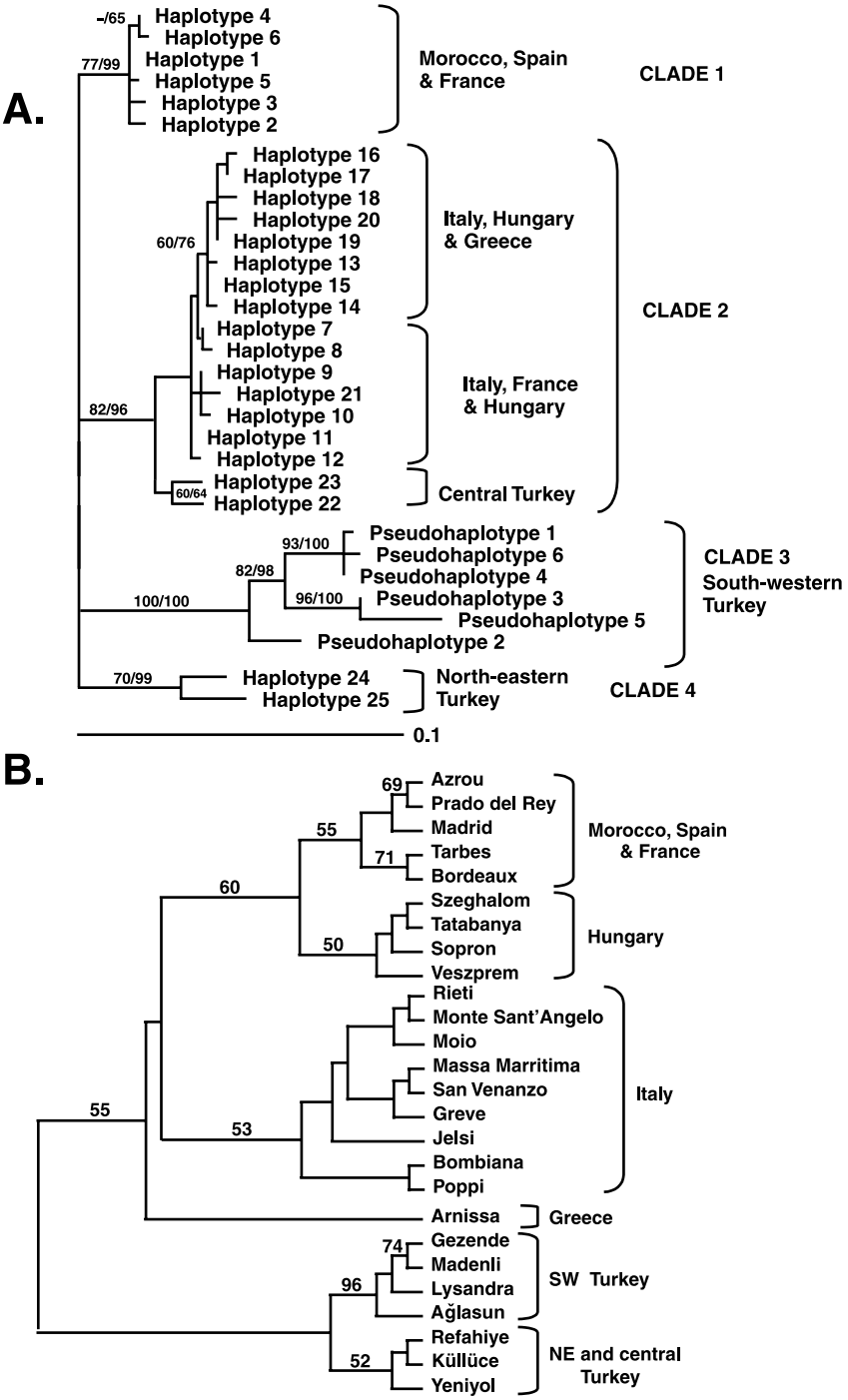

Fig. 2 (a) Phylogram representing the ML topology (100 random-taxon-addition replications with TBR swapping). Bootstrap support values > 50% are indicated on the left of the slash and posterior probabilities > 50% on the right. (b) A 50% majority-rule consensus tree based on allozyme allele frequencies in *A. quercustozae* populations. The phylogeny was generated using the least squares criterion on Cavalli-Sforza's chord distance (Cavalli-Sforza & Edwards 1967) in PHYLIP. Numbers at nodes indicate bootstrap support.

of the time since these lineages diverged (Table 3). These estimates imply that divergence among the European refugia is not recent, but predates the Pleistocene glaciations. Furthermore, if one of the sequences from southwestern Turkey (clade 4) represents the true mitochondrial sequence, then the divergence between eastern and western Turkey is even more ancient.

Which of the A. quercustozae *clades is basal?* Inclusion of outgroups, either singly or together, failed to resolve which of the *A. quercustozae* lineages shown in Fig. 2a is basal (data not shown). In trees rooted with closely related *Andricus* species [members of the *A. quercuscalicis* clade *sensu* Stone & Cook (1998)], the four *A. quercustozae* clades did not form a monophyletic group but were placed in a basal polytomy

Table 3 Sequence divergence (lower diagonal) and estimated date in millions of years since most recent common ancestry (upper diagonal) within and between the four *A. quercustozae* clades in Fig. 1A. Along the diagonal, data are presented as divergence/estimated date. Dates since most recent common ancestry are estimated assuming a 2.3% rate of divergence every million years (Brower 1994)

	Clade 1	Clade 2	Clade 3	Clade 4
Clade 1	1.9/0.8	2.7	8.0	2.6
Clade 2	6.2	2.9/1.3	7.9	3.2
Clade 3	18.5	18.2	8.5/3.7	7.0
Clade 4	5.9	7.3	16.0	1.1/0.5

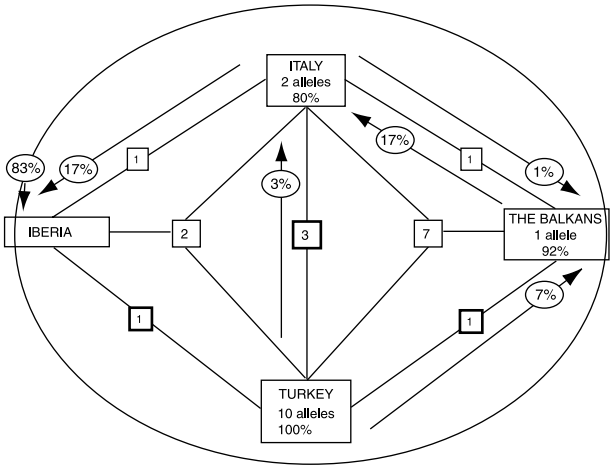

Fig. 3 A diagrammatic representation of the distribution of alleles among the three major European refugial regions (Iberia and Morocco, Italy, the Balkans) and Turkey, excluding those found in all regions. Alleles joined by links from more than one region are shared by those regions. Alleles shared only by Turkey and one other region are in boxes with a bold border, and would be private alleles for that region were Turkish sites excluded. Percentage figures within boxes are the proportion of individuals attributable to that region's genotype pool by STRUCTURE. Percentage values on arrows indicate the proportion of individuals in the targeted genotype pool attributable to the arrow's source.

together with the outgroups. In trees rooted using the two more distantly related outgroups (*A. conificus*, *A. conglomeratus*), *A. quercustozae* was always a monophyletic taxon, but no single clade was consistently supported as basal. There is thus inadequate resolution in this data set either to reveal the geographical origin of *A. quercustozae*, or which of the *Q. cerris*/*Q. suber* lifecycles is ancestral.

Analyses of allozyme allele frequency data

Longitudinal patterns in allele frequencies and genetic diversity. Allele frequency data for all loci and populations are given in Appendix I. Over the entire sampled range of *A. quercustozae* the 12 polymorphic enzyme systems had a total of 48 alleles. Of these, 16 were found throughout the distribution of *A. quercustozae* and probably represent ancestral polymorphism. Of the remaining 32 alleles, 13 were found in only one of the four main regions sampled (Fig. 3); one was restricted to the Balkans, two were restricted to Italy and 10 were found only in Turkey. The number of alleles found only in Turkey is striking, and because the numbers of specimens and populations sampled from Turkish sites is similar to or lower than those in Italy and the Balkans, this high richness is unlikely to represent a sampling artefact. The inclusion of Turkish populations is significant because of its impact on the apparent distribution of region-specific alleles in Europe. If the Turkish populations are excluded (as they have been through limited sampling in many previous studies of widespread taxa), the number of alleles private to each of the recognized refugial areas in southern Europe increases (to one in Iberia, five in Italy and two in the Balkans). The east–west decline in the diversity of regionally private alleles is repeated in both measures of genome-wide genetic diversity (Fig. 4, Table 1): allelic diversity halves from Turkey to Morocco, and average heterozygosity falls by *c.* 75% (from 0.24 to 0.30–0.07–0.08) over the same range.

Pairwise F_{ST} *A. quercustozae* shows a strong signature of isolation by distance across its longitudinal range (Fig. 5, Appendix II), with values for comparisons between Turkey and Spain/southwestern France reaching 0.635. Pairwise values were generally higher for comparisons involving Turkish sites than for comparisons over an equivalent geographical distance in Europe.

How many refugia are there for A. quercustozae*?*

(i) Pairwise distance-matrix methods (Fig. 2b). Pairwise distance-matrix methods support the division of *A. quercustozae* into the same five groups of populations suggested by analysis of haplotype sequences (compare Fig. 2a,b), with the difference that the allele frequency data gather the Turkish populations into a single group. This analysis supports the southwest–northeast division within Turkey evident in the haplotype data, and also provides additional resolution of population divergence for populations within haplotype clade 2. Here Iberian populations are grouped not with geographically closer populations in Italy, but with Balkan populations.

(ii) Model-based method. The results of applying STRUCTURE for differing numbers of populations (genotype pools) provide strong support for the division of *A. quercustozae*

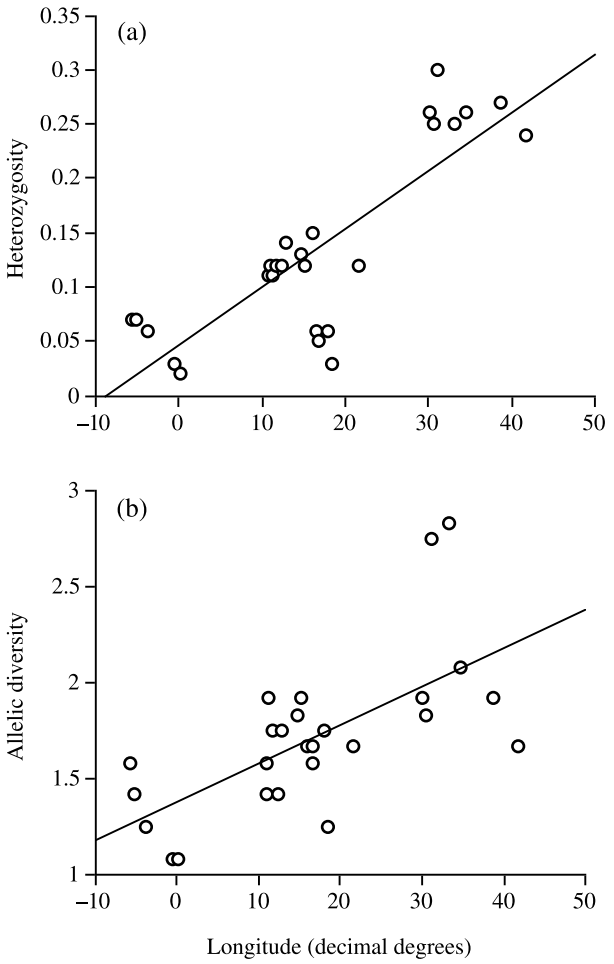

Fig. 4 The correlation between longitude (in decimalized degrees) and genetic diversity indices for *A. quercustozae*. The fitted lines are least squares regressions. (a) Expected heterozygosity: $y = 0.0047$ (longitude) + 0.047, $R^2 = 0.684$, $P < 0.001$. (b) Allelic diversity: $y = 0.020$ (longitude) + 1.382, $R^2 = 0.423$, $P < 0.01$.

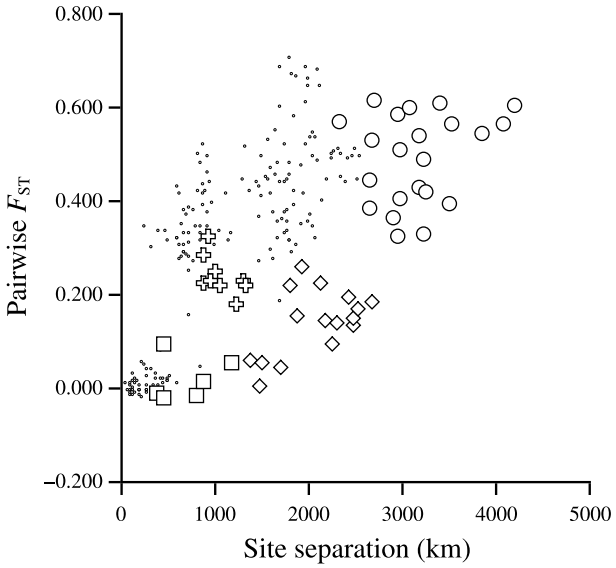

Fig. 5 The relationship between pairwise F_{ST} and site separation. The signature of isolation by distance in the data is illustrated by highlighting (with large symbols) comparisons between sites in southwestern Europe (Morocco, Spain, southwestern France) and each other (squares), Italy (crosses), Hungary (diamonds) and Turkey (circles).

Table 4 Conditional and posterior probabilities from STRUCTURE associated with each pool number (K) from a model without admixture. ln Pr(X | K) is the estimated conditional probability of observing the data under a model with the specified number of pools. Pr(K | X) is the posterior probability (calculated from Bayes' s rule with a uniform prior) of each pool number given the underlying genotypic information

K	ln Pr(X | K)	Pr(K | X)
1	−5061	~0
2	−4138	~0
3	−4009	~0
4	–3795	~1
5	–3954	~0

into four groups (Table 4, Fig. 6). Examination of the composition of these four groups shows them to correspond to Italy, Hungary, southwestern Turkey and central and northeastern Turkey, in general agreement with those supported by pairwise distance-matrix methods (Table 4, Fig. 4). The addition of a fifth population did not result in a distinct Iberian group, as might be expected, but instead resulted in division of the Italian population into two. The assignment of individuals to populations suggests very limited gene flow between the two Turkish genotype groups, and from Europe to Turkey, with all individuals allocated to their regional groups with probabilities near 1. However, there is evidence of migration of individuals into Europe from Turkey, and between regions within Europe. All but four of the Hungarian individuals were allocated to the Hungarian genotype group, but two (both from Veszprem) were assigned to the Italian genotype group and two (both from Sopron) were assigned to the southwestern Turkey genotype group. The inferred origins of individuals in Italy and Iberia are more mixed. All Italian populations included individuals assigned to the Hungarian genotype group. One individual from Greve (Italy) was assigned to the southwestern Turkish genotype group with a probability of 0.99, while for the other 38 individuals from this site the association probability for this genotype group is zero. Seven individuals from Monte Sant' Angelo (Italy) and seven of eight individuals from Arnissa (Greece) were assigned with probabilities between 0.996 and one to

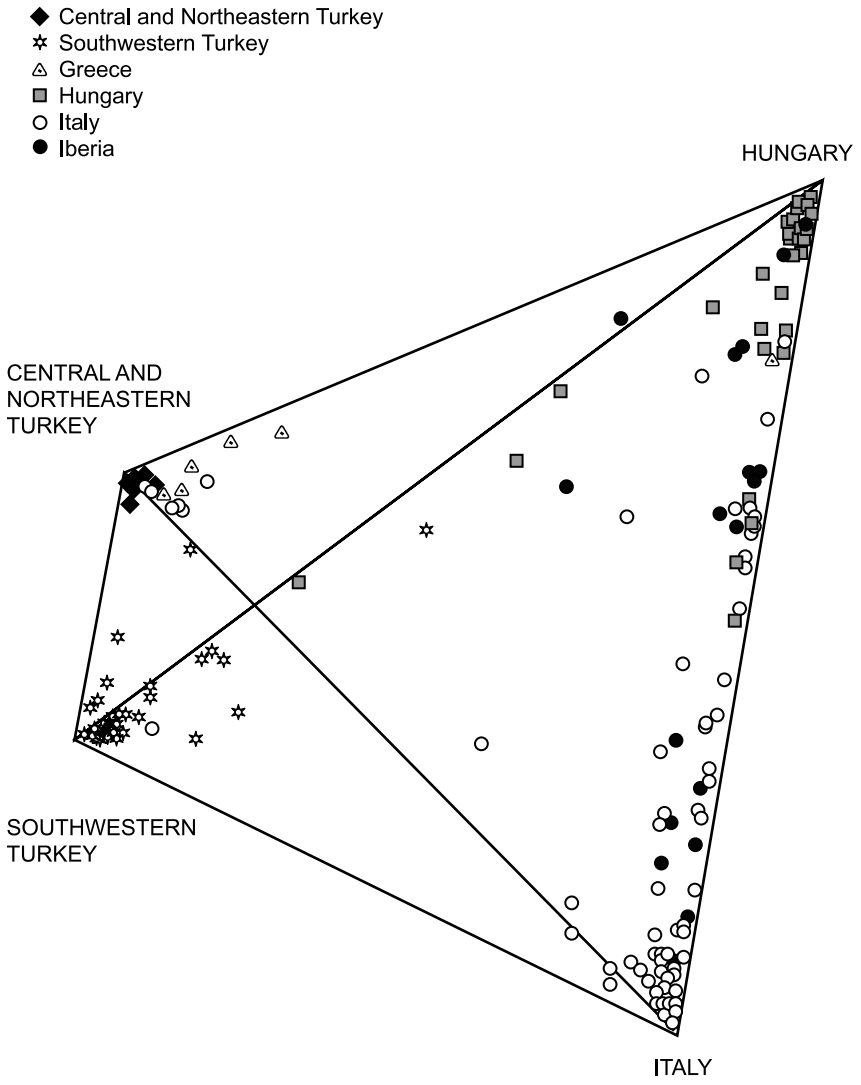

Fig. 6 Clustering of data from the STRUCTURE analysis with four discrete pools. Each axis represents the probability of membership of one of the four groups (represented by the four corners of the tetrahedron).

the genotype group that otherwise includes individuals only from central and northeastern Turkey. Most individuals sampled in Iberia were assigned to the Hungarian genotype group, but a small proportion of individuals in all sites (7/39 for Azrou, 3/17 for Prado del Rey and 2/9 for Madrid) were more strongly clustered with the Italian genotype group. No Iberian populations showed any affinity with either Turkish grouping.

Discussion

How many refugia were occupied by A. quercustozae*?*

Analyses based on allele frequencies and haplotype sequences both reveal considerable genetic subdivision in *A. quercustozae*. Turkey, the Balkans and Italy each have region-specific haplotypes and allozyme alleles, while populations in Spain, southwest France and Morocco within the native range of *Q. suber* possess region-specific haplotypes but no Iberia-specific allozyme alleles. Analysis of haplotype sequence variation suggests two discrete clades in Europe (Iberia and Italy/Hungary), with at least one clearly distinct lineage in Turkey. Analyses of allozyme allele frequency data add further resolution to this picture. First, they reveal differentiation (though with gene flow) between Italy and Hungary that is not apparent in the haplotype data. Second, they support the existence of at least two genetically discrete sets of populations in Turkey — one including populations in central and northeastern Turkey, and one including populations in the southwest. Third, allele frequency data show gene flow from Turkish sources to more western populations in Greece and Italy.

In general, regional variation in genetic diversity can be attributed to three processes. Variation could be (i) an artefact associated with limited sampling effort. Alternatively, if the pattern is real, it could be due to either (ii) genetic

drift, during and subsequent to longitudinal range expansion, in the frequencies of haplotypes and alleles originally present in an ancestral population or (iii) the generation of true refuge-specific polymorphism through the accumulation of mutations in the descendents of founding colonists. In the absence of genetic exchange between refugial areas, these alternatives can be distinguished for sequence data by the phylogenetic relationship among refuge-specific haplotypes. While the first and second explanations involve random loss of ancestral variants (and so have no expectation that haplotypes sampled in the same refugium should be similar in sequence), the third predicts that refuge-specific haplotypes should constitute one or more monophyletic groups (Templeton 1998). The pattern in *A. quercustozae* (Fig. 2a) is highly nonrandom, and is consistent with the third hypothesis. The extent of sequence divergence among clades further suggests that lineages in Iberia, Italy and the Balkans and the two Turkish clades diverged before the recent Pleistocene glaciations, and so represent discrete refugial populations.

The pattern and dating of refugial differentiation for *A. quercustozae* in Europe parallels that seen in many other taxa (Taberlet *et al*. 1998; Hewitt 1999), including other oak gallwasps (Stone & Sunnucks 1993; Atkinson 2000; Rokas *et al*. 2001; Stone *et al*. 2001). The extent of divergence between Italy and the Balkans is more pronounced than in other host-alternating gallwasps [*A. quercuscalicis* (Stone & Sunnucks 1993); *A. kollari* (Stone *et al*. 2001)], although less pronounced than is seen in their oak hosts, for which these two regions represent clearly differentiated refugia (Cooper *et al*. 1995; Dumolin-Lapegue *et al*. 1997; Ferris *et al*. 1998). The lesser divergence seen in the gallwasps may indicate high levels of gene flow, or genetic divergence that is so recent that refuge-specific haplotypes and alleles have yet to evolve. The ability of gallwasps to disperse rapidly over large distances (Stone & Sunnucks 1993; Sunnucks & Stone 1996; Csóka *et al*. 1998) makes gene flow between these two neighbouring regions highly likely. Two major features of population structure in *A. quercustozae* — the high genetic diversity of Turkish populations and the genetic divide at the boundary between the oaks *Q. suber* and *Q. cerris*- are now discussed in detail.

The significance of Turkey for A. quercustozae

Turkey represents not only a genetically discrete part of the range of *A. quercustozae*, it also represents the centre of genetic diversity for the species — whether measured in terms of haplotype nucleotide diversity, the number of region-specific private alleles or allelic diversity and heterozygosity measured over multiple enzyme loci. Turkish sites also share several rare allozyme alleles found otherwise in only one of Iberia, Italy or the Balkans. If we assume that the demonstrated distributions of these shared alleles are real and not sampling artefacts, the distributions could have been generated by three general processes: (a) regional variation in the retention of ancestral polymorphism. Alleles shared by Turkey and any one other refugium then represent ancestral polymorphisms that have been lost from other refugia. (b) Gene flow, such that alleles common to pairs of refugia but absent from the others are the result of long range migration. (c) Homoplasy — independent evolution of allozyme alleles with the same electrophoretic mobility in two different lineages. While gene flow is suggested by individual-based modelling of population structure, and is a plausible explanation for those alleles that are shared by neighbouring pairs of refugia (Fig. 3), it is less plausible for alleles, such as AK allele 4, which are shared by Turkish and Spanish populations over 5200 km apart and absent from all populations in between. The clear geographical differentiation in haplotypes discussed above also argues against frequent long-range gene flow. The AK4 allele is thus either homoplasious in Iberia and Turkey or (we suggest more probably) represents the retention in both regions of ancestral polymorphism.

The high genetic diversity in ancestral polymorphism present in Turkey could mean either that this region is the source of colonists founding populations further west, or that for some reason (for example, a lower frequency of population bottlenecks) populations there have retained a higher proportion of ancestral genetic diversity originally present in an alternative, unsampled origin. While the second interpretation is possible, there is no evidence for an ancient centre of oak and gallwasp diversity that could have given rise, independently, to populations in all the regions sampled here. Instead, several lines of evidence support the inference that Turkey represents the source for pre-Pleistocene colonization of Europe by *A. quercustozae*. First, east–west migration is supported by those assignment probabilities for individual gallwasps in Greece, Hungary and Italy that suggest a Turkish origin, and there was no evidence of west–east migration. Second, the longitudinal patterns in genetic diversity match closely those generated during range expansion in many taxa, including plants (Lagercrantz & Ryman 1990; Demesure *et al*. 1996; Dumolin-Lapegue *et al*. 1997; King & Ferris 1998), insects (Cooper *et al*. 1995; Armbruster *et al*. 1998; Wilcock *et al*. 2001) and vertebrates (Sage & Wolff 1986; Merilå *et al*. 1997). In particular, similar declines in diversity with distance from a known source have been generated during recent invasions of northern and western Europe by other host-alternating *Andricus* gallwasps. In these cases, the invaders are known to have colonized Europe from the Balkans over the last 500 years following human introduction of *Q. cerris* (Stone & Sunnucks 1993; Csóka *et al*. 1998; Stone *et al*. 2001). In at least three replicate gallwasp examples, heterozygosity and allelic diversity decrease with distance from the invasion origin. Populations further

along the invasion route possess only a subset of the alleles seen nearer the invasion origin, and rare alleles are lost nearest the source. The pattern generated is compatible with stepping-stone colonization by small numbers of colonists, involving successive bottlenecking events and little or no long-range gene flow between the outer limits of the invaded range and the invasion source (Stone & Sunnucks 1993; Ibrahim *et al*. 1996; Luikart *et al*. 1998). If we extrapolate the results of these processes over longer timescales, with repeated genetic isolation of populations in glacial refugia, we expect patterns associated with the original range expansion process to be modified by the appearance of novel, refuge-specific polymorphism over time. The higher mutation rate of mitochondrial haplotypes, and their lower effective population size (1/4) relative to nuclear loci also predicts that refugial differentiation should be more marked in haplotypes than in nuclear allozyme alleles. The patterns seen in *A. quercustozae* are compatible with these predictions, with Turkey as an ancient invasion source. Iberia possesses only a subset of the ancestral allozyme polymorphism seen in populations further east, and in this hypothesis represents the western limit of the longitudinal expansion process. The colonization of Iberia must have happened long enough ago, however, for Iberia-specific haplotypes to have diverged from those found elsewhere.

Similar refugial differentiation and longitudinal trends in genetic diversity are seen in two other oak gallwasps with known or inferred host-alternating lifecycles: *A. kollari* (Stone *et al*. 2001) and *A. caputmedusae* (Atkinson 2000). The congruence in these patterns suggests to us that Asia Minor is not only a centre of intraspecific diversity for these gallwasps, but also the probable centre of origin for host-alternating *Andricus* species as a group. This conclusion is supported by the structure of the *Andricus* phylogeny: host-alternating *Andricus* gallwasps represent a single clade within the genus (Cook *et al*. 2002), and while lineages restricted to eastern Europe and Turkey appear throughout the *Andricus* tree, taxa restricted to Iberia represent only terminal branch tips within widespread species groups (Cook *et al*. 2002). This implies either that Iberia-specific taxa are derived from eastern lineages or (we suggest less parsimoniously) that many Iberian representatives of more deeply branching lineages have become extinct. Species richness within the host-alternating clade is also highest in the east (Melika *et al*. 2000; Nieves-Aldrey 2001). Taken together, these phylogenetic and population genetic patterns suggest to us that species now widespread in Europe diverged in Asia Minor during the radiation of the group, and expanded their distributions westwards prior to the Pleistocene glaciations. This hypothesis is compatible with phylogenetic evidence suggesting Asia Minor as the centre of radiation and diversity for Western Palaearctic lineages of both of the oak taxa required by host-alternating gallwasps (Govaerts & Frodin 1998; Manos *et al*. 1999). It also leads to the strong prediction that all the other widespread host-alternating oak gallwasps should show the same pattern seen in *A. quercustozae*.

Turkey, the Caucasus and Iran are known to be important centres of diversity for a range of taxa, including black alder *Alnus glutinosa* (King & Ferris 1998), the grasshopper *Chorthippus parallelus* (Cooper *et al*. 1995), the honeybee *Apis mellifera* (Franck *et al*. 2001) and the domestic mouse *Mus musculus* (Boursot *et al*. 1993). Nevertheless, little is known about the genetic diversity present in the easternmost distributions of many well-studied species. While this lack of data may reflect the dominant role of latitudinal patterns in the genetic diversity of many European taxa, such a bias may result in underestimation of the role of more easterly refugia in European biodiversity. Finally, Turkish populations of *A. quercustozae* are significant in the interpretation of the European distribution of genetic variation. Region-specific alleles have been used in estimation of gene flow (Slatkin 1987) and in the identification of genetically discrete taxa (e.g. Berlocher 2000). Inclusion of Turkish populations shows several rare alleles to represent ancestral polymorphism, rather than region-specific variants, and may represent a cautionary tale in the utility of this type of genetic data for other similarly widespread taxa.

Haplotype and allele frequency data both support the existence of two lineages of *A. quercustozae* in Turkey. Turkey represents the meeting point of three phytogeographical regions; the Irano-Turanian, the Euro-Siberian and the Mediterranean, each with a distinct climate and flora (Yaltirik 1982). The Irano-Turanian centre of endemism includes the majority of the interior of Turkey, extending from the western margin of the Anatolian Plateau eastwards into Iran. This Irano-Turanian region is in turn bisected by the Anatolian diagonal, a line of mountains running from the Anti-Taurus on the Mediterranean coast northeast to the Black Sea coast near Trabzon. The Anatolian divide represents an important floristic divide between a predominantly eastern Asiatic flora and a more western European flora, a divide that is also visible in the distribution of oak species in both of the oak sections required by host-alternating *Andricus* (Yaltirik 1982). The two clusters of Turkish *A. quercustozae* genotypes represent one from each side of the Anatolian diagonal. It is tempting to suggest that the divide between these lineages may represent a major division between two refuge populations of *Andricus quercustozae*. East of the divide, *Q. cerris* is replaced by *Q. libani* and *Q. persica*, and it is probable that the host-shift between *Q. suber* and *Q. cerris* discussed below is only one of several that have occurred during longitudinal range expansion in host-alternating gallwasps.

Host shifts between Q. cerris *and* Q. suber

Data for *A. quercustozae* show that populations in Iberia and Morocco contain only a subset of allozyme alleles found further east, show lower genetic diversity than populations further east, but represent an entirely distinct set of mitochondrial haplotypes representing a single clade. Allozyme data clusters Iberian and Moroccan populations in general with Hungary, although some individuals show affinity to Italian populations. As discussed above, these results support the conclusion that Iberia was colonized from the east, and thus that *Q. suber* represents the derived host and *Q. cerris* the ancestral host for this species. Similar patterns, although less highly resolved, have been shown for *A. kollari* (Stone *et al*. 2001) and the hypothesis developed above predicts similar patterns in the other widespread host-alternating *Andricus* species. In both *A. kollari* and *A. quercustozae,* Iberian populations belong to a single lineage emerging from a basal polytomy in the species' haplotype tree. This can be interpreted in two ways. Taken at face value, the pattern implies that the host shift to *Q. suber* occurred at about the same time as the occupation of other parts of southern Europe prior to the Pleistocene, and involved a single lineage. This interpretation is compatible with the rarity of host shifts in gallwasps (Abe 1988, 1991; Cook *et al*. 2002), the population bottlenecks seen during host-shifts in other plant-feeding insects (Brown *et al*. 1996; Groman & Pellmyr 2000) and limited gene flow between host-adapted ecotypes (Via 1999; Berlocher 2000; Nice & Shapiro 2001; Dres & Mallet 2002). An alternative interpretation is that many lineages in *A. quercustozae* and *A. kollari* may have made the shift to *Q. suber*, but higher extinction rates in Iberian lineages have resulted in the pattern we now see. The possibility of higher rates of extinction in Iberian lineages is supported by the fact that in both *A. quercustozae* and *A. kollari*, there is lower nucleotide diversity in Iberia than in any other region sampled, suggesting a more recent common ancestor of extant haplotypes in Iberia. Combined with the basal divergence of the Iberian lineage within the species haplotype trees, this implies ancient divergence but recent bottlenecks in the Iberian populations of both gallwasp species. Because *A. kollari* and *A. quercustozae* are similar in many aspects of their biology, either hypothesis could explain the congruence of patterns across these two species. More extensive sampling of haplotype sequence variation in Iberia is necessary to discriminate between these two alternatives.

Phenotypic diversity in A. quercustozae

The two gall phenotypes induced by *A. quercustozae* are separated geographically into an eastern form (the bright red, resinous *insana* form found in Turkey and Iran, the Levant and in Europe in southern parts of the Balkans and Greece) and a western form (the nonsticky *quercustozae* form) (Dalla Torre & Kieffer 1910; Chodjai 1980). There is abundant evidence that gallwasp genes control the major aspects of gall phenotype (Stone & Cook 1998; Stone *et al*. 2002), and it is thus possible that these two forms are induced by discrete monophyletic lineages within *A. quercustozae*. However, our limited mitochondrial sequence data do not support this hypothesis. It remains possible that this hypothesis is in fact true, and that the cytochrome *b* tree does not reflect accurately the phylogeny of genes involved directly in gall phenotypes. An alternative is that the differences in gall morphology result from the impact of host oak genes on gall phenotype (Stone & Schönrogge 2003) and reflect geographical variation in the asexual generation oak host. Loss of the *insana* gall form coincides with the western limit of *Quercus infectoria*, the most widespread (and inferred ancestral) host of *Andricus quercustozae* in Asia Minor and Iran, and it is possible that range expansion involved colonization of hosts in which generation of the *insana* phenotype is impossible. This is significant because the sticky outer surface of the *insana* form is of potential adaptive significance as a defence against natural enemies of the gall wasp (Stone & Cook 1998; Stone *et al*. 2002), and is a trait whose persistence through natural selection is thus predicted.

Conclusions

The Quaternary ice ages had a profound impact on the genetic structure of the European fauna and flora. Accumulating evidence from studies of European oak gallwasps (Stone & Sunnucks 1993; Atkinson 2000; this study; Rokas *et al*. 2001; Stone *et al*. 2001) and their oak hosts (Ferris *et al*. 1993; Dumolin-Lapegue *et al*. 1997; Ferris *et al*. 1998; Toumi & Lumaret 1998) highlights the genetic distinctness of southern European refugia, and the role of Asia Minor both as a centre of contemporary diversity and the probable origin of European lineages.

Analysis of patterns of intraspecific genetic diversity can be a useful tool in the identification of populations or regions of high conservation importance. This approach is most appropriate for gradients in genetic diversity resulting from very recent demographic processes in which the impacts on neutral and selectively important variation are most likely to be similar. Applicability of the approach to gradients associated with longer timescales may miss important phenotypic differences (such as host-associated ecotypes) that are not apparent in selectively neutral (or nearly neutral) markers. Generation of more extensive data sets for a range of species using a suite of available molecular markers and techniques is now necessary for better description of the processes in action. In the oak gallwasp system, knowledge of species-specific ecological

constraints, such as availability of hosts for different stages of the lifecycle, can be used in generating testable null hypotheses (Stone *et al*. 2001; this study) and in understanding how different species react to common environmental changes.

Acknowledgements

We would like to thank José-Luis Nieves-Aldrey, Alison Atkinson, Gyuri Csóka, Sara Lourie, Gil McVean, John Pannell, Alex Rowe and Yan Wong for their help in collecting samples and Rob Smallshire for help with the 3D tetrahedron. The paper was improved by constructive criticism from Paul Sunnucks and three anonymous referees. A.R. was partially funded by a NERC studentship and is currently a Human Frontier Science Program Long-Term Fellow. R.J.A. was funded by a BBSRC studentship, with additional funding for fieldwork by the Henrietta Hutton Memorial Fund and the John Ray Society. This work was supported by grants to G.N.S. from the British Ecological Society, the Royal Society (grant no. 574006) and NERC (GR/12847 and the EDGE thematic programme GR9/03553).

References

Abe Y (1988) Two host races in *Andricus mukaigawae* (Mukaigawa) (Hymenoptera: Cynipidae). *Applied Entomology and Zoology*, **23**, 382–387.

Abe Y (1991) Host race formation in the gall wasp *Andricus mukaigawae*. *Entomologia Experimentalis et Applicata*, **58**, 15–20.

Armbruster P, Bradshaw WE, Holzapfel CM (1998) Effects of postglacial range expansion on allozyme and quantitative genetic variation of the pitcher-plant mosquito, *Wyeomyia smithii*. *Evolution*, **52**, 1697–1704.

Atkinson RJ (2000) The genetic analysis of natural history, reproductive strategy and population structure in European oak gallwasps (Hymenoptera: Cynipidae). PhD Thesis, University of Oxford, Oxford, UK.

Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2000) GENETIX *4.02, Logiciel Sous Windows TM Pour la Génétique Des Populations*. Laboratoire Génome, Populations, Interactions, CNRS UPR 9060, Université de Montpellier II, Montpellier.

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

Bensasson D, Zhang DX, Hewitt GM (2000) Frequent assimilation of mitochondrial DNA by grasshopper nuclear genomes. *Molecular Biology and Evolution*, **17**, 406–415.

Berlocher SH (2000) Radiation and divergence in the *Rhagoletis pomonella* species group: inferences from allozymes. *Evolution*, **54**, 543–557.

Boursot P, Auffray J-C, Britton-Davidian J, Bonhomme F (1993) The evolution of house mice. *Annual Review of Ecology and Systematics*, **24**, 119–152.

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

Brown JM, Abrahamson WG, Way PA (1996) Mitochondrial DNA phylogeography of host races of the goldenrod ball gallmaker, *Eurosta solidaginis* (Diptera: Tephritidae). *Evolution*, **50**, 777–786.

Camus A (1936–1954) Les Chênes. Monographie du genre *Quercus* and monographie du genre *Lithocarpus*. *Encycl*. *Encyclopédie Économique De Sylviculture*, **6**, 7.

Castroviejo S (1986) *Flora Iberica: Plantas Vasculatres de la Península Ibérica El Islas Baleares*. CSIC, Madrid.

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

Chodjai M (1980) Study of the gall forming cynipid Hymenoptera of Iranian oak forests. *Journal of the Entomological Society of Iran*, Supplement **3**.

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

Cooper SJB, Ibrahim KM, Hewitt GM (1995) Postglacial expansion and genome subdivision in the European grasshopper *Chorthippus parallelus*. *Molecular Ecology*, **4**, 49–60.

Csóka G, Stone G, Atkinson R, Schönrogge K (1998) The population genetics of postglacial invasions of northern Europe by cynipid gall wasps (Hymenoptera: Cynipidae). In: *The Biology of Gall-Inducing Arthropods* (eds Csóka G, Mattson WJ, Stone GN, Price PW), pp. 280–294. US Department of Agriculture, Forest Service, North Central Research Station, St Paul, MN.

Dalla Torre KW, Kieffer JJ (1910) *Cynipidae*. Friedlander & Sohn, Berlin.

Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. *Evolution*, **50**, 2515–2520.

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

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

Felsenstein J (1993) PHYLIP *(Phylogeny Inference Package)*. Distributed by the Author. Department of Genetics. University of Washington, Seattle.

Ferris C, King RA, Vainola R, Hewitt GM (1998) Chloroplast DNA recognizes three refugial sources of European oaks and suggests independent eastern and western immigrations to Finland. *Heredity*, **80**, 584–593.

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

Ferris C, Oliver RP, Davy AJ, Hewitt GM (1995) Using chloroplast DNA to trace postglacial migration routes of oaks into Britain. *Molecular Ecology*, **4**, 731–738.

Franck P, Garnery L, Loiseau A *et al.* (2001) Genetic diversity of the honeybee in Africa: microsatellite and mitochondrial data. *Heredity*, **86**, 420–430.

Govaerts R, Frodin DG (1998) *World Checklist and Bibliography of Fagales*. Royal Botanical Gardens, Kew, UK.

Groman JD, Pellmyr O (2000) Rapid evolution and specialization following host colonization in a yucca moth. *Journal of Evolutionary Biology*, **13**, 223–236.

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

Hewitt GM (2000) The genetic legacy of the Quaternary ice ages. *Nature*, **405**, 907–913.

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, **17**, 754–755.

Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP (2001) Bayesian inference of phylogeny and its impact on evolutionary biology. *Science*, **294**, 2310–2314.

Ibrahim KM, Nichols RA, Hewitt GM (1996) Spatial patterns of genetic variation generated by different forms of dispersal during range expansion. *Heredity*, **77**, 282–291.

Jalas J, Suominen J (1987) *Atlas Florae Europaeae*. *Distribution of Vascular Plants in Europe*. Cambridge University Press, Cambridge.

King RA, Ferris C (1998) Chloroplast DNA phylogeography of *Alnus glutinosa* (L.) Gaertn. *Molecular Ecology*, **7**, 1151–1161.

Konarov VL (1936) *Flora of the USSR*. Izdatel'stva Akademii Nauk SSSR, Moscow/Leningrad.

Lagercrantz U, Ryman N (1990) Genetic structure of Norway spruce (*Picea abies*): concordance of morphological and allozymic variation. *Evolution*, **44**, 38–53.

Lewis PO (2001) Phylogenetic systematics turns over a new leaf. *Trends in Ecology and Evolution*, **16**, 30–37.

Luikart G, Allendorf FW, Cornuet JM, Sherwin WB (1998) Distortion of allele frequency distributions provides a test for recent population bottlenecks. *Journal of Heredity*, **89**, 238–247.

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

Melika G, Csoka G, Pujade-Villar J (2000) Check-list of oak gall wasps of Hungary, with some taxonomic notes (Hymenoptera: Cynipidae, Cynipinae, Cynipini). *Annales Historico-Naturales Musei Nationalis Hungarici*, **92**, 265–296.

Merila J, Bjorklund M, Baker AJ (1997) Historical demography and present day population structure of the greenfinch, *Carduelis chloris* — an analysis of mtDNA control-region sequences. *Evolution*, **51**, 946–956.

Mirol PM, Mascheretti S, Searle JB (2000) Multiple nuclear pseudogenes of mitochondrial cytochrome b in *Ctenomys* (Caviomorpha, Rodentia) with either great similarity to or high divergence from the true mitochondrial sequence. *Heredity*, **84**, 538–547.

Nei M (1972) Genetic distance between populations. *American Naturalist*, **106**, 283–292.

Nice CC, Shapiro AM (2001) Population genetic evidence of restricted gene flow between host races in the butterfly genus *Mitoura* (Lepidoptera: Lycaenidae). *Annals of the Entomological Society of America*, **94**, 257–267.

Nieves-Aldrey J-L (2001) *Hymenoptera, Cynipidae*. Museo Nacional de Ciencias Naturales, CSIC, Madrid.

Petit RJ, Pineau E, Demesure B, Bacilieri R, Ducousso A, Kremer A (1997) Chloroplast DNA footprints of postglacial recolonization by oaks. *Proceedings of the National Academy of Sciences USA*, **94**, 9996–10001.

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

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

Raymond M, Rousset F (1995) GENEPOP (version-1.2) — population-genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

Richardson BJ, Baverstock PR, Adams M (1986) *Allozyme Electrophoresis: a Handbook for Animal Systematics and Population Studies*. Academic Press, San Diego.

Rokas A, Atkinson RJ, Brown GS, West SA, Stone GN (2001) Understanding patterns of genetic diversity in the oak gallwasp *Biorhiza pallida*: demographic history or a *Wolbachia* selective sweep? *Heredity*, **87**, 294–304.

Rokas A, Melika G, Abe Y, Nieves-Aldrey J-L, Cook JM, Stone GN (2003) 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.

Sage RD, Wolff JO (1986) Pleistocene glaciations, fluctuating ranges, and low genetic variability in a large mammal (*Ovis dalli*). *Evolution*, **40**, 1092–1095.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*, **236**, 787–792.

Sternlicht M (1968) The oak galls of Israel (*Quercus calliprinos* Webb and *Quercus ithaburensis* Decne). *Israel Journal of Entomology*, **3**, 17–58.

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

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

Stone GN, Schönrogge K (2003) The adaptive significance of insect gall morphology. *Trends in Ecology and Evolution*, in press.

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

Stone GN, Sunnucks PJ (1993) Genetic consequences of an invasion through a patchy environment — the cynipid gallwasp *Andricus quercuscalicis* (Hymenoptera: Cynipidae). *Molecular Ecology*, **2**, 251–268.

Sunnucks P, Stone GN (1996) Genetic structure of invading insects and the case of the knopper gallwasp. In: *Frontiers of Population Ecology*, pp. 485–496. CSIRO, Canberra, Australia.

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

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

Templeton AR (1998) Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. *Molecular Ecology*, **7**, 381–397.

Toumi L, Lumaret R (1998) Allozyme variation in cork oak (*Quercus suber* L.): the role of phylogeography and genetic introgression by other Mediterranean oak species and human activities. *Theoretical and Applied Genetics*, **97**, 647–656.

Townsend CC, Guest ER (1980) *Flora of Iraq*, 4. Part 1: *Cornaceae to Rubiaceae*. Ministry of Agriculture and Agrarian Research, Baghdad.

Turrill WB (1929) *The Plant Life of the Balkan Peninsula; a Phytogeographical Study*. Clarendon Press, Oxford.

Via S (1999) Reproductive isolation between sympatric races of pea aphids. I. Gene flow restriction and habitat choice. *Evolution*, **53**, 1446–1457.

Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.

Wilcock HR, Hildrew AG, Nichols RA (2001) Genetic differentiation of a European caddisfly: past and present gene flow among fragmented larval habitats. *Molecular Ecology*, **10**, 1821–1834.

Yaltirik F (1982) Fagaceae. In: *Flora of Turkey and the Eastern Aegean Islands*, Vol. 7 (ed. Davis DH), pp. Edinburgh University Press, Edinburgh, UK.

Zohary M (1966) *Flora Palaestina*, 1. *Equisetaeae to Moringaceae*. The Israel Academy of Sciences and Humanities, Jerusalem.

This project is part of a long-term research programme being carried out in Graham Stone's laboratory on the population biology, phylogeography and phylogeny of gallwasps. Antonis Rokas' main interest is the application of 'tree-thinking' to a variety of evolutionary ecological questions relating to gallwasps at various taxonomic levels. Rachel Atkinson is interested in the application of molecular population genetics to address ecological questions ranging from the oviposition behaviour of a given gallwasp species to its postglacial recolonization route into northern Europe. Lucy Webster is currently working on the effects of gene flow on local adaptation in a natural host–parasite system.

Appendix I

Locations, allozyme sample sizes and allele frequencies for all *Andricus quercustozae* populations sampled

Location	Azrou	Madrid	Prado del Rey	Bombiana	Greve in Chianti	Jelsi	Massa Marittimo	Moio	Monte Sant'Angelo	Poppi	Rieti	San Venanzo	Sopron
Country	Morocco	Spain	Spain	Italy	Italy	Italy	Italy	Italy	Italy	Italy	Italy	Italy	Hungary
° latitude	33.45	40.42	36.80	44.20	43.58	41.53	44.63	40.15	41.72	43.72	42.40	42.87	47.67
° longitude	–5.23	–3.72	–5.55	10.95	11.32	14.80	9.70	15.17	15.97	11.77	12.85	12.27	16.58
N	39	9	17	16	39	28	27	28	40	39	22	27	40
GPD1													
1	0	0	0	0	0.039	0.071	0.074	0.036	0.113	0	0.046	0.093	0
2	1	1	1	1	0.962	0.911	0.926	0.964	0.863	1	0.932	0.907	1
3	0	0	0	0	0	0.018	0	0	0.025	0	0.023	0	0
GPD2													
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0.013	0	0	0	0	0	0	0	0
3	1	1	1	1	0.987	1	1	0.982	1	1	1	1	0.975
4	0	0	0	0	0	0	0	0.018	0	0	0	0	0.025
5	0	0	0	0	0	0	0	0	0	0	0	0	0
GOTs													
1	0	0	0	0	0	0	0	0.054	0	0	0	0	0.025
2	1	0.944	1	1	1	1	1	0.929	0.963	0.974	0.886	1	0.975
3	0	0.056	0	0	0	0	0	0.018	0.038	0.026	0.114	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0
GOTm													
1	0	0	0	0	0	0	0	0	0.175	0	0	0	0
2	1	1	1	1	1	1	1	1	0.825	1	1	1	1
3	0	0	0	0	0	0	0	0	0	0	0	0	0
GPI													
1	0	0	0	0.125	0.167	0.125	0.130	0.196	0.175	0.167	0.205	0.167	0
2	0.923	1	0.971	0.844	0.821	0.875	0.870	0.786	0.813	0.833	0.796	0.833	1
3	0.077	0	0.029	0.031	0.013	0	0	0.018	0.013	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0
HK													
1	0	0	0	0.031	0	0	0	0	0	0.013	0	0	0
2	1	1	1	0.969	0.987	1	1	1	1	0.987	0.977	1	0.975
3	0	0	0	0	0.013	0	0	0	0	0	0.023	0	0.025

Appendix I *Continued*

Location	Azrou	Madrid	Prado del Rey	Bombiana	Greve in Chianti	Jelsi	Massa Marittima	Moio	Monte Sant'Angelo	Poppi	Rieti	San Venanzo	Sopron
Country	Morocco	Spain	Spain	Italy	Italy	Italy	Italy	Italy	Italy	Italy	Italy	Italy	Hungary
° latitude	33.45	40.42	36.80	44.20	43.58	41.53	44.63	40.15	41.72	43.72	42.40	42.87	47.67
° longitude	−5.23	−3.72	−5.55	10.95	11.32	14.80	9.70	15.17	15.97	11.77	12.85	12.27	16.58
N	39	9	17	16	39	28	27	28	40	39	22	27	40
MDHm													
1	0.026	0	0	0	0	0.054	0	0	0	0	0	0	0
2	0.577	0.556	0.647	0.375	0.192	0.357	0.296	0.268	0.288	0.346	0.318	0.315	0.988
3	0.397	0.444	0.353	0.625	0.808	0.589	0.704	0.732	0.713	0.654	0.682	0.685	0.013
4	0	0	0	0	0	0	0	0	0	0	0	0	0
ME													
1	1	1	1	1	0.974	1	1	1	1	1	1	1	1
2	0	0	0	0	0.026	0	0	0	0	0	0	0	0
PEPb													
1	0	0	0	0	0	0.036	0	0	0	0	0	0	0
2	0	0	0	0.344	0.397	0.304	0.389	0.179	0.325	0.346	0.318	0.333	0.100
3	1	1	0.971	0.656	0.590	0.661	0.611	0.804	0.675	0.641	0.682	0.667	0.825
4	0	0	0.029	0	0.013	0	0	0.018	0	0.013	0	0	0.075
5	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
PGM													
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0.077	0.056	0.059	0.031	0	0.036	0	0	0	0.013	0	0	0
3	0.923	0.944	0.912	0.906	1	0.946	1	1	1	0.962	1	0.963	1
4	0	0	0.029	0.063	0	0.018	0	0	0	0.026	0	0.037	0
AK													
1	0	0	0	0	0	0	0.019	0	0	0	0.023	0	0.013
2	0.026	0	0.029	0	0.026	0	0	0	0	0.013	0	0	0.125
3	0.974	1	0.941	1	0.974	1	0.982	1	1	0.987	0.977	1	0.863
4	0	0	0.029	0	0	0	0	0	0	0	0	0	0
6PGD													
1	0	0	0	0	0.013	0.036	0	0.018	0	0	0.046	0	0
2	1	1	1	1	0.987	0.964	1	0.929	1	1	0.955	1	1
3	0	0	0	0	0	0	0	0.054	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix I *Continued*

Location	Szeghalom	Tatabanya	Veszprem	Arnissa	Ağlasun	Gezende	Küllüce	Lysandra	Madenli	Refahiye	Yeniyol	Tarbes	Bordeaux
Country	Hungary	Hungary	Hungary	Greece	Anatolia	Anatolia	Anatolia	Anatolia	Anatolia	Anatolia	Anatolia	France	France
° latitude	47.23	47.52	47.10	40.46	37.65	36.53	38.20	36.48	38.13	39.90	41.40	43.23	44.83
° longitude	16.70	18.42	17.90	21.56	30.53	33.15	34.60	30.05	31.02	38.75	41.63	0.08	–0.57
N	39	10	40	8	15	40	7	6	40	7	5	9	12
GPD1													
1	0	0	0.013	0	0	0.013	0	0	0	0	0	0	0
2	1	1	0.988	1	1	0.975	1	1	1	1	1	1	1
3	0	0	0	0	0	0.013	0	0	0	0	0	0	0
GPD2													
1	0	0	0	0	0	0	0	0	0.013	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0.987	1	0.988	1	1	1	1	1	0.988	1	1	1	1
4	0.013	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0.013	0	0	0	0	0	0	0	0	0	0
GOTs													
1	0.077	0	0.075	0.063	0.200	0.038	0	0.167	0.225	0	0.200	0	0
2	0.923	1	0.925	0.938	0.400	0.513	1	0.500	0.388	0.857	0.800	1	1
3	0	0	0	0	0.333	0.425	0	0.333	0.388	0.143	0	0	0
4	0	0	0	0	0.067	0.025	0	0	0	0	0	0	0
GOTm													
1	0	0	0	0.875	0	0.013	0.929	0	0.050	1	1	0	0
2	1	1	1	0.125	1	0.988	0	1	0.950	0	0	1	1
3	0	0	0	0	0	0	0.071	0	0	0	0	0	0
GPI													
1	0.013	0	0.013	0.063	0	0.038	0.071	0	0	0.071	0	0	0
2	0.936	0.950	0.975	0.813	1	0.900	0.929	1	0.963	0.929	1	1	1
3	0.051	0.050	0.013	0.063	0	0.050	0	0	0.038	0	0	0	0
4	0	0	0	0.063	0	0.013	0	0	0	0	0	0	0
HK													
1	0.013	0	0	0	0	0.013	0	0	0.013	0	0	0	0
2	0.987	1	0.975	1	0.967	0.938	1	0.917	0.913	1	1	1	1
3	0	0	0.025	0	0.033	0.050	0	0.083	0.075	0	0	0	0

Appendix I *Continued*

Location	Szeghalom	Tatabanya	Veszprem	Arnissa	Ağlasun	Gezende	Küllüce	Lysandra	Madenli	Refahiye	Yeniyol	Tarbes	Bordeaux
Country	Hungary	Hungary	Hungary	Greece	Anatolia	Anatolia	Anatolia	Anatolia	Anatolia	Anatolia	Anatolia	France	France
° latitude	47.23	47.52	47.10	40.46	37.65	36.53	38.20	36.48	38.13	39.90	41.40	43.23	44.83
° longitude	16.70	18.42	17.90	21.56	30.53	33.15	34.60	30.05	31.02	38.75	41.63	0.08	–0.57
N	39	10	40	8	15	40	7	6	40	7	5	9	12
MDHm													
1	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0.974	1	0.913	0.875	0.433	0.500	0.429	0.833	0.325	0.429	0.600	0.833	0.792
3	0.026	0	0.088	0.125	0.567	0.500	0.571	0.167	0.663	0.571	0.400	0.167	0.208
4	0	0	0	0	0	0	0	0	0.013	0	0	0	0
ME													
1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	0	0	0	0	0	0	0	0	0	0	0	0	0
PEPb													
1	0	0	0	0	0	0.100	0.071	0.083	0.050	0	0.100	0	0
2	0.128	0.100	0.113	0.313	0.233	0.188	0.214	0.167	0.188	0.357	0.600	0	0
3	0.872	0.900	0.888	0.625	0.600	0.350	0.500	0.500	0.413	0.357	0.300	1	1
4	0	0	0	0.063	0.167	0.275	0.071	0.250	0.288	0	0	0	0
5	0	0	0	0	0	0.088	0.071	0	0.050	0.071	0	0	0
6	0	0	0	0	0	0	0	0	0.013	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0.214	0	0	0
8	0	0	0	0	0	0	0.071	0	0	0	0	0	0
PGM													
1	0	0	0	0	0	0.013	0	0.250	0.050	0	0	0	0
2	0	0	0.013	0	0.167	0.125	0.143	0.167	0.125	0.357	0.500	0	0
3	1	1	0.988	1	0.833	0.838	0.643	0.583	0.775	0.643	0.500	1	1
4	0	0	0	0	0	0.025	0.214	0	0.050	0	0	0	0
AK													
1	0	0	0	0	0.533	0.875	0.286	0.583	0.688	0.286	0.300	0	0
2	0	0	0	0	0.467	0.125	0	0.417	0.288	0.429	0.100	0	0
3	1	1	1	1	0	0	0	0	0.025	0.143	0.600	1	1
4	0	0	0	0	0	0	0.714	0	0	0.143	0	0	0
6PGD													
1	0	0.050	0	0	0	0.025	0.214	0	0.088	0	0.100	0	0
2	1	0.950	1	1	0.800	0.813	0.571	0.917	0.625	0.786	0.900	1	1
3	0	0	0	0	0.200	0.163	0	0.083	0.288	0.214	0	0	0
4	0	0	0	0	0	0	0.214	0	0	0	0	0	0

Appendix II

Pairwise F_{ST} values between all populations sampled in the study. Values in bold type are significant at $P < 0.05$, having adjusted significance levels using a Bonferroni correction for multiple tests

	Prado del Rey (Sp)	Madrid (Sp)	Tarbes (Fr)	Bordeaux (Fr)	Poppi (It)	Bombiana (It)	Rieti (It)	San Venanzo (It)	Greve in Chianti (It)	Massa Marittimo (It)	Jelsi (It)	Moio (It)
Azrou (Mor)	–0.011	–0.016	0.055	0.043	**0.146**	**0.129**	**0.148**	**0.156**	**0.227**	**0.186**	**0.113**	**0.129**
Prado del Rey (Sp)		–0.024	0.014	0.006	**0.144**	**0.124**	**0.142**	**0.154**	**0.229**	**0.186**	**0.105**	**0.135**
Madrid (Sp)			0.093	0.068	**0.111**	0.093	**0.095**	0.117	**0.186**	**0.148**	0.070	0.083
Tarbes (Fr)				–0.068	**0.228**	**0.223**	**0.217**	**0.246**	**0.322**	**0.282**	**0.178**	**0.229**
Bordeaux (Fr)					**0.219**	**0.220**	**0.214**	**0.239**	**0.314**	**0.274**	**0.173**	**0.221**
Poppi (It)						–0.017	–0.005	–0.007	0.009	–0.005	–0.005	**0.016**
Bombiana (It)							–0.007	–0.014	0.011	–0.009	–0.018	0.011
Rieti(It)								–0.008	0.007	–0.004	–0.005	–0.003
San Venanzo (It)									0.002	–0.014	–0.012	0.007
Greve in Chianti (It)										–0.005	0.020	**0.026**
Massa Marittimo (It)											–0.007	0.021
Jelsi (It)												0.013

	M. San'Angelo (It)	Tatabanya (Hu)	Sopron (Hu)	Szeghalom (Hu)	Veszprem (Hu)	Madenli (Tur)	Ağlasun (Tur)	Gezende (Tur)	Lysandra (Tur)	Arnissa (Gr)	Yeniyol (Tur)	Refahiye (Tur)	Küllüce (Tur)
Azrou (Mor)	**0.161**	**0.169**	**0.191**	**0.183**	**0.135**	**0.375**	**0.420**	**0.394**	**0.487**	**0.484**	**0.602**	**0.596**	**0.615**
Prado del Rey (Sp)	**0.154**	**0.140**	**0.145**	**0.150**	**0.095**	**0.330**	**0.364**	**0.352**	**0.428**	**0.457**	**0.562**	**0.544**	**0.562**
Madrid (Sp)	**0.112**	**0.259**	**0.220**	**0.224**	**0.152**	**0.299**	**0.323**	**0.330**	**0.405**	**0.466**	**0.539**	**0.508**	**0.529**
Tarbes (Fr)	**0.219**	0.054	**0.057**	**0.041**	0.004	**0.346**	**0.383**	**0.365**	**0.445**	**0.496**	**0.607**	**0.568**	**0.585**
Bordeaux (Fr)	**0.214**	**0.097**	**0.086**	**0.072**	0.026	**0.356**	**0.404**	**0.375**	**0.483**	**0.521**	**0.635**	**0.597**	**0.614**
Poppi (It)	**0.020**	**0.289**	**0.318**	**0.315**	**0.277**	**0.329**	**0.362**	**0.354**	**0.444**	**0.418**	**0.494**	**0.503**	**0.537**
Bombiana (It)	0.015	**0.287**	**0.322**	**0.325**	**0.278**	**0.289**	**0.317**	**0.318**	**0.391**	**0.393**	**0.445**	**0.452**	**0.484**
Rieti (It)	0.007	**0.278**	**0.331**	**0.329**	**0.288**	**0.276**	**0.297**	**0.305**	**0.377**	**0.380**	**0.440**	**0.443**	**0.476**
San Venanzo (It)	0.006	**0.312**	**0.348**	**0.347**	**0.304**	**0.319**	**0.355**	**0.348**	**0.442**	**0.422**	**0.488**	**0.492**	**0.522**
Greve in Chianti (It)	**0.022**	**0.383**	**0.413**	**0.416**	**0.378**	**0.336**	**0.378**	**0.370**	**0.480**	**0.461**	**0.510**	**0.511**	**0.545**
Massa Marittimo (It)	0.010	**0.344**	**0.371**	**0.375**	**0.332**	**0.317**	**0.357**	**0.346**	**0.454**	**0.440**	**0.497**	**0.500**	**0.533**
Jelsi (It)	0.014	**0.239**	**0.285**	**0.283**	**0.241**	**0.309**	**0.334**	**0.334**	**0.402**	**0.380**	**0.451**	**0.466**	**0.496**
Moio (It)	0.022	**0.311**	**0.357**	**0.352**	**0.308**	**0.310**	**0.344**	**0.349**	**0.441**	**0.430**	**0.502**	**0.495**	**0.521**
M. San'Angelo (It)		**0.278**	**0.328**	**0.321**	**0.286**	**0.323**	**0.345**	**0.350**	**0.418**	**0.314**	**0.398**	**0.415**	**0.454**
Tatabanya (Hu)			0.009	–0.018	–0.004	**0.365**	**0.409**	**0.377**	**0.447**	**0.479**	**0.608**	**0.582**	**0.599**
Sopron (Hu)				**0.028**	**0.030**	**0.417**	**0.469**	**0.420**	**0.478**	**0.494**	**0.647**	**0.647**	**0.670**
Szeghalom (Hu)					–0.004	**0.445**	**0.516**	**0.452**	**0.551**	**0.520**	**0.681**	**0.686**	**0.704**
Veszprem (Hu)						**0.431**	**0.495**	**0.441**	**0.530**	**0.500**	**0.659**	**0.666**	**0.685**
Madenli (Tur)							0.006	**0.025**	**0.056**	**0.390**	**0.303**	**0.251**	**0.299**
Ağlasun (Tur)								**0.042**	0.018	**0.430**	**0.333**	**0.270**	**0.337**
Gezende (Tur)									**0.049**	**0.416**	**0.346**	**0.303**	**0.343**
Lysandra (Tur)										**0.435**	**0.314**	**0.281**	**0.337**
Arnissa (Gr)											**0.186**	**0.268**	**0.328**
Yeniyol (Tur)												0.040	**0.153**
Refahiye (Tur)													**0.078**

