

Longitudinal range expansion and cryptic eastern species in the western Palaearctic oak gallwasp, *Andricus coriarius*

RICHARD J. CHALLIS,* SERAP MUTUN,† JOSE-LUIS NIEVES-ALDREY,‡ SONJA PREUSS,* ANTONIS ROKAS,§ ALEXANDRE AEBI,* EBRAHIM SADEGHI,¶ MAJID TAVAKOLI** and GRAHAM N. STONE*

*Institute of Evolutionary Biology, University of Edinburgh, School of Biological Sciences, The King's Buildings, West Mains Road, Edinburgh, EH9 3JT, UK, †Faculty of Science & Arts, Department of Biology, Abant Izzet Baysal University, 14280 Bolu, Turkey, ‡Departamento de Biodiversidad y Biología Evolutiva Museo Nacional de Ciencias Naturales (CSIC) c/José Gutiérrez Abascal, 2. Madrid. E-28006, Spain, §The Broad Institute of MIT & Harvard, 320 Charles Street, Cambridge, MA 02141 USA, ¶Research Institute of Forest & Rangelands, PO Box: 13185–116, Teheran, Iran, **Lorestan Agricultural and Natural Resources Research Centre, Khorramabad, Iran PO Box: 348

Abstract

The oak gallwasp *Andricus coriarius* is distributed across the Western Palaearctic from Morocco to Iran. It belongs to a clade of host-alternating *Andricus* species that requires host oaks in two sections of *Quercus* subgenus *Quercus* to complete its lifecycle, a requirement that has restricted the historic distribution and dispersal of members of this clade. Here we present nuclear and mitochondrial sequence evidence from the entire geographic range of *A. coriarius* to investigate the genetic legacy of longitudinal range expansion. We show *A. coriarius* as currently understood to be para- or polyphyletic, with three evolutionarily independent (but partially sympatric) lineages that diverged c. 10 million years ago (mya). The similarities in gall structure that have justified recognition of single species to date thus represent either strong conservation of an ancestral state or striking convergence. All three lineages originated in areas to the east of Europe, underlining the significance of Turkey, Iran and the Levant as 'cradles' of gallwasp evolution. One of the three lineages gave rise to all European populations, and range expansion from a putative Eastern origin to the present distribution is predicted to have occurred around 1.6 mya.

Keywords: *Andricus*, cryptic speciation, gallwasp, glacial refugia, range expansion

Received 30 July 2006; revision accepted 23 October 2006

Introduction

The phylogeography of most western Palaearctic species is dominated by the influence of Quaternary glacial cycles (Hewitt 1996, 1999, 2004; Taberlet *et al.* 1998). During glacial periods, western Palaearctic species lacking tolerance to low temperatures were confined to a longitudinal spread of southern refugia spanning Europe, Turkey, the Levant, Iran and the Caucasus (Hewitt 1999, 2004). Current distributions of western Palaearctic taxa result from two linked processes: longitudinal dispersal associated with

occupation of new refugia and exchange between them, and changes in latitudinal distributions associated with the advance and retreat of each ice age. Many species show genetic differentiation between refugial populations, consistent with prolonged reproductive isolation, and refuge-specific polymorphism has been used to reconstruct recent postglacial range expansion routes in many taxa (Hewitt 1996, 1999; Taberlet *et al.* 1998). Fewer studies, however, have considered the longitudinal colonization of multiple southern refugia by widespread species (Rokas *et al.* 2003a), a process usually inferred to predate the most recent glacial period (c. 115 000–15 000 years before present, BP; e.g. Tzedakis 1993; Ferris *et al.* 1998; Taberlet *et al.* 1998). What are the origins of such widespread taxa, and when did longitudinal range expansion occur? Answering such questions requires sampling of species throughout their longitudinal range, which for many 'European' taxa extends

Correspondence: Richard Challis, Fax: 0131 650 6564; E-mail: r.j.challis@sms.ed.ac.uk

or

Graham Stone, Fax: 1131 650 6564; E-mail: graham.stone@ed.ac.uk

eastwards across Anatolia to the Caucasus and beyond. This is true, for example, of the keystone western Palaearctic oaks *Quercus robur*, *Q. pubescens* and *Q. petraea* (Govaerts & Frodin 1998; Petit *et al.* 2002), and for at least some of the insects associated with these trees (e.g. Stone *et al.* 2001; Rokas *et al.* 2003a). A recent extra-European origin is frequently revealed in pest species of commercial crops. The olive fly, *Bactrocera oleae*, has an African origin (Nardi *et al.* 2005) while the chestnut gallwasp, *Dryocosmus kuriphilus*, was introduced to Europe from the Far East (Aebi *et al.* 2006). It is only relatively recently that extra-European areas have been fully considered in the phylogeography of 'European' organisms (Hewitt 2004). Recent studies have shown areas of the western Palaearctic east of Europe to be the geographic origin of cryptic species of long-eared bats, *Plecotus* (Juste *et al.* 2004), of the house mouse subspecies, *Mus musculus domesticus* (Gunduz *et al.* 2005), and European lineages of butterflies (Schmitt *et al.* 2005) and fish (e.g. Kotlik *et al.* 2004; Culling *et al.* 2006).

Oak gallwasps (Hymenoptera: Cynipidae) are obligate plant parasites that develop within galls induced on specific plant tissues (Stone *et al.* 2002). Most oak gallwasps have a cyclically parthenogenetic lifecycle (Atkinson *et al.* 2002, 2003) with a sexual generation in the spring and a parthenogenetic generation in the summer/autumn (Stone *et al.* 2002). *Andricus coriarius* belongs to a monophyletic clade of western Palaearctic *Andricus* species that also show obligate host-alternation (Stone & Cook 1998; Cook *et al.* 2002). For all members of this clade whose lifecycles have been studied, the parthenogenetic generation gall is induced on section *Quercus* oaks (e.g. English oak *Quercus robur* and sessile oak, *Q. petraea*) while the sexual generation gall forms on section *Cerris* oaks (e.g. Turkey oak, *Q. cerris*, and cork oak, *Q. suber*) (Stone *et al.* 2001; Cook *et al.* 2002; Csóka *et al.* 2005). At least 10 members of this clade have extremely wide longitudinal ranges, extending from Iberia and Morocco to Iran and the Caucasus, and with populations in all of the major European oak refugia (the Iberian Peninsula, Italy, and the Balkans) (Atkinson 2000; Melika *et al.* 2000; Nieves-Aldrey 2001; Cook *et al.* 2002; Rokas *et al.* 2003a, b). A feature of these widespread gallwasps is that while their parthenogenetic generation oak hosts are found over much of this longitudinal range, the sexual generation of oak changes with longitude, from *Q. suber* in the Iberian Peninsula and northwestern Africa, to *Q. cerris* from Italy eastwards to Turkey and the Levant, and then to *Q. libani* and *Q. brantii* (Lebanese and Persian oaks) in Iran. The identification of such widespread species raise three general questions: (i) which region within these wide distributions was the 'cradle' for these species; (ii) what is the timescale and mode (continuous or iterated) of longitudinal range expansion; and (iii) what is the frequency and direction of shifts between alternative oak hosts? Phylogeographic analysis of another widespread host-

alternating *Andricus* species, *A. quercustozae* (Rokas *et al.* 2003a), supported an eastern origin, leading to the proposal of an 'Out of Anatolia' hypothesis for host-alternating *Andricus*. This paper tests this hypothesis for *A. coriarius*.

Here we address these issues through analysis of sequence data from 80 individuals of *A. coriarius* for a 433 base-pair (bp) fragment of the mitochondrial cytochrome *b* (*cytb*) gene. We detect cryptic lineages, and assess their phylogenetic status with additional sequence data for the D2 region of the nuclear 28S region. We extend the longitudinal range sampled in previous studies of oak gallwasps to Lebanon and Iran, and so include the full known range of host-alternating *Andricus* in the western Palaearctic. We also assess the demographic processes associated with longitudinal range expansion to investigate whether natural longitudinal range expansion has left a signature of rapid population growth, as suggested in other species (e.g. Michaux *et al.* 2003; Culling *et al.* 2006).

Materials and methods

Sample collection

Parthenogenetic generation galls of *A. coriarius* were collected from locations across the species' range from Spain to Iran (Table 1; Fig. 1). All galls were morphologically similar: galls were globular, with a surface coating of stout pointed spines, contained many larval chambers, and developed on lateral and terminal buds of young shoots. Among recognized western Palaearctic species, this combination of traits is only present in *A. coriarius*. As far as possible, sample sizes were balanced across distribution regions, with additional sampling efforts in Iran after preliminary investigation revealing high haplotype diversity. The only sexual generation oak host available to the sampled Iranian populations was *Q. libani*, while both *Q. libani* and *Q. cerris* were available to Lebanese populations. All populations further west have sexual generations on *Q. cerris*, with the exception of Iberian populations, for which *Q. suber* is the only available sexual generation host.

The parthenogenetic females that emerge from a single gall are commonly the offspring of a single sexual female (Atkinson *et al.* 2002). Only a single individual was thus sequenced from each gall. Adult wasps were reared from their galls under quarantine in Edinburgh then stored in ethanol at -20 °C.

DNA extraction and sequencing

DNA was extracted using the DNeasy Tissue kit (QIAGEN cat. 69504), following the manufacturer's protocol for insect DNA extraction. To infer relationships between sampled individuals and selected outgroups we used sequence for the mitochondrial *cytb* gene, and for the D2 region of the

Table 1 Locations, sample sizes and haplotypes sampled for each population. Locations are given in decimal degrees for latitude followed by longitude, and are shown in Fig. 1. Haplotype numbers are as in Table 2. Where multiple copies of a haplotype were sampled from a population, numbers in parentheses after haplotype number indicate the number of individuals sharing that haplotype

	Population	Country	Location	Sample size	Haplotypes
1	El Escorial	Spain	40.58, -4.13	3	18(2), 32
2	Orusco	Spain	40.28, -3.22	1	18
3	Llerida	Spain	41.61, 0.63	1	18
4	Valpiana	Italy	43.02, 10.84	1	1
5	Anconella	Italy	43.76, 11.30	1	12
6	Abruzzo	Italy	42.49, 13.72	2	9, 1
7	Cupoli	Italy	42.46, 13.83	2	8, 9
8	Molize	Italy	41.68, 14.54	2	1(2)
9	Gargano	Italy	41.89, 16.13	2	10, 11
10	Istria	Croatia	45.21, 13.89	1	3
11	Köszeg	Hungary	47.39, 16.54	3	4, 8, 20
12	Várpalota	Hungary	47.20, 18.15	1	5
13	Tatabánya	Hungary	47.60, 18.41	1	1
14	Szoloske	Hungary	47.88, 19.01	2	6, 7
15	Gödöllő	Hungary	47.61, 19.36	1	16
16	Balaton	Hungary	48.10, 20.31	1	9
17	Eger	Hungary	47.92, 20.38	1	2
18	Plástovce	Slovakia	48.16, 18.98	1	5
19	Prespa	Greece	40.77, 21.09	1	9
20	Pisoderi	Greece	40.78, 21.25	4	5, 9, 13, 15
21	Florina	Greece	40.78, 21.41	1	14
22	Komnina	Greece	40.59, 21.77	1	17
23	Edessa	Greece	40.80, 22.05	1	5
24	Tarsus	Turkey	36.92, 34.90	1	23
25	Küllüce	Turkey	38.20, 34.60	1	16
26	Çekerek	Turkey	40.07, 35.49	1	24
27	Suluova	Turkey	40.84, 35.65	3	16(2), 21
28	Tokat	Turkey	40.32, 36.55	2	16, 19
29	Niksar	Turkey	40.59, 36.95	3	22, 25(2)
30	Deir al Zahari	Lebanon	33.43, 35.47	2	43, 44
31	Jezzine	Lebanon	33.54, 35.59	1	43
32	Ain Dara	Lebanon	33.78, 35.73	1	41
33	Ibrahim River	Lebanon	34.07, 35.88	1	42
34	Pyranshahr	Iran	36.69, 45.23	4	26(2), 27, 28
35	Baneh	Iran	35.99, 45.90	14	16(6), 29, 30, 31(5), 33
36	Marivan	Iran	35.52, 46.17	11	34(2), 35(4), 36, 37, 38, 39, 40

nuclear 28S ribosomal array. A 433 bp fragment of the mitochondrial *cytb* gene was amplified and sequenced for 77 individuals using the CB1/CB2 primer pair of Jermini & Crozier (1994): *cytb* forward primer CB1, 5'-TAT GTA CTA CCA TGA GGA CAA ATA TC-3' and *cytb* reverse primer CB2, 5'-ATT ACA CCT CCT AAT TTA TTA GGA AT-3'. A subset of individuals was also sequenced for the D2 region of the 28S ribosomal array using the following primers: D2 forward primer, 5'-CGT GTT GCT TGA TAG TGC AGC-3' and D2 reverse primer, 5'-TCA AGA CGG GTC CTG AAA GT-3' (Hancock *et al.* 1988). PCRs were carried out in a PTC-200 DNA Engine (MJ Research). 25 µL PCRs were performed using 1 U *Taq* polymerase (Invitrogen or Promega), 2.5 µL 10× *Taq* buffer, 1.5 µL MgCl₂ (25 mM), 0.5 µL dNTPs (10 mM), 0.35 µL primers (20 pmol), 1.0 µL

template DNA and 18.85 µL dH₂O. PCR products were purified using shrimp alkaline phosphatase (USB Corporation, USA). Sequencing was carried out on an automated ABI Prism 3100 Genetic Analyser machine using ABI BigDye version 3.1 Terminator Sequencing chemistry. Chromatogram output was checked by eye. PCR products for each specimen were sequenced in both directions to minimize PCR artefacts, ambiguities and base-calling errors. The resulting traces were analysed using SEQUENCHER 4.1 (Gene Codes). All sequences are deposited in GenBank (accession numbers EF029996-EF030036, *cytb*, and EF030037-EF030039, 28SD2). Three previously published *A. coriarius* *cytb* sequences were also included: Gödöllő, Hungary (AJ228458); Edessa, Greece (AF539556); and El Escorial, Spain (AF539557) (Rokas *et al.* 2003b).

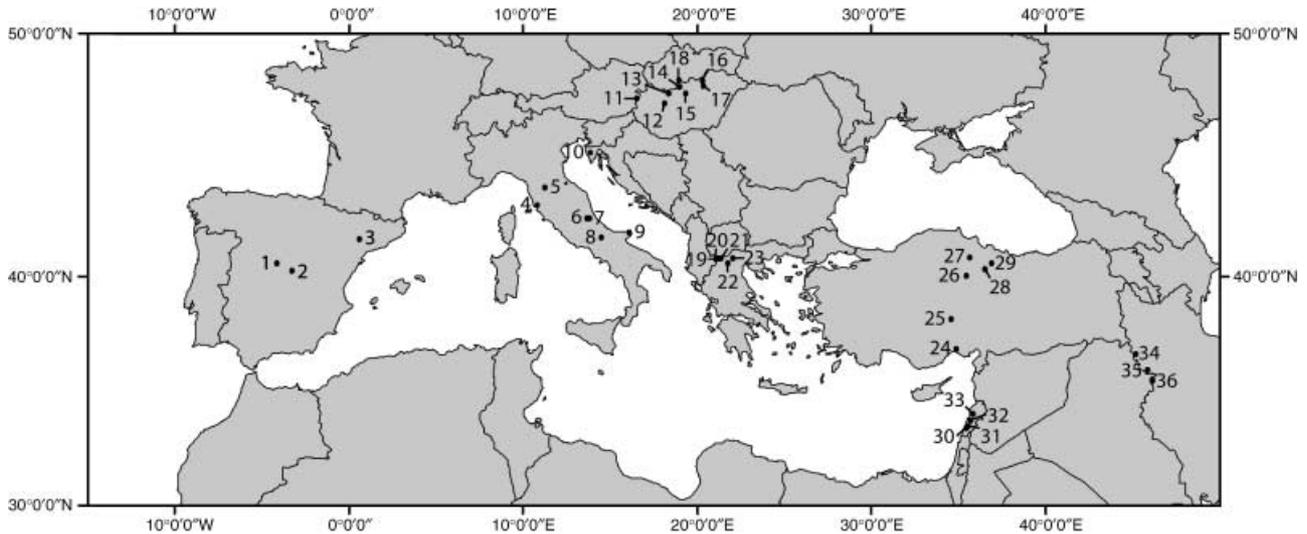


Fig. 1 Map showing the collection locations. Locations are numbered as in Table 1.

Testing the validity of a phylogenetic framework

In order to test the validity of using phylogenetic methods we applied the test for 'tree-likeness' of Huson & Bryant (2006). A phylogenetic network was generated for the two-gene sequence alignment in SPLITSTREE 4.4 (Huson & Bryant 2006) using the NeighbourNet (Bryant & Moulton 2004) distances transformation and equal angle splits transformation (Dress & Huson 2004). A 95% confidence network was produced from 1000 bootstrap replicates. The hypothesis that the data originated on a tree is accepted if this confidence network contains a tree and rejected if it does not. An important characteristic of this test is that, although it has low power, regions of tree-like and nontree-like evolution within a single network can be identified and appropriate analyses applied to each region.

The hypothesis of tree-like evolution was accepted for the deeper (interspecific) nodes of the two gene dataset network, i.e. the confidence intervals on edges in these regions of the confidence network were compatible with a single bifurcating phylogeny. However, intraspecific relationships between *A. coriarius* haplotypes were not compatible with any tree, and it is thus invalid to assume that the intraspecific data originated on a tree. Specifically, three edges in the network, all at the intraspecific level, could not be represented by bifurcating phylogenies. Phylogenetic methods have therefore only been used at the interspecific level and alternative network-based methods are adopted for intraspecific analyses (see below).

Phylogenetic analysis

Phylogenetic reconstruction was performed using MRBAYES 3.1 (Ronquist & Huelsenbeck 2003). The two-gene dataset

was partitioned by gene and the *cytb* partition was further divided into the three codon positions. Instead of a *a priori* model selection, a parameter-rich GTR + I + Γ model was applied to each partition, exploiting the efficiency of the Metropolis-coupled Markov chain Monte Carlo (MC³) algorithm within MRBAYES to optimize parameters. Two independent runs of four Markov chains over two-million generations were performed for each analysis with the temperature parameter set at 0.15. Convergence was assessed through examination of plots of chain parameters and comparison of the output from the two independent runs. Trees were sampled every 1000 generations after a burn in period of 200 000 generations. To assess the monophyly of *A. coriarius*, additional *cytb* and 28S D2 sequences were included in the phylogenetic analyses. All *cytb* outgroup sequences are previously published: *A. caputmedusae*, AF539553 and *A. conificus*, AF539555 (Rokas *et al.* 2003b); *A. conglomeratus*, AJ228468, *A. curvator*, AJ228453, *A. seckendorffi*, AJ228449, *A. solitarius*, AJ228475 and *Cynips quercus*, AJ228478 (Stone & Cook 1998); and *A. kollari*, AF242739 (Stone *et al.* 2001). Two previously published 28SD2 sequences were included: *A. curvator*, AF395155 and *A. kollari*, AF395156 (Rokas *et al.* 2002). 28SD2 sequences for *A. caputmedusae*, EF030040, *A. conglomeratus*, EF030041, *A. conificus*, EF030042, *A. seckendorffii*, EF030043, *A. solitarius*, EF030044, and *Cynips quercus*, EF030045, were provided by Antonio Hernandez-Lopez (unpublished data).

Estimating the time depth of nodes in the haplotype tree

The validity of the assumption of a molecular clock for the *cytb* data was tested using Bayes factors (Kass & Raftery 1995). A strict clock model assumes a constant mutation rate throughout the phylogeny while a nonclock model

allows an independent mutation rate for each branch. The validity of these hypotheses can be tested using Bayes factors as the use of a less appropriate model in Bayesian phylogenetic reconstruction is expected to result in an increase in the marginal likelihood of the resulting phylogenetic hypothesis. Marginal likelihood was approximated by calculating the harmonic mean, H , log-likelihood, lnL scores obtained from MC³ runs under a nonclock and three strict clock models (uniform, birth-death and coalescence), implemented in MRBAYES 3.1 (Ronquist & Huelsenbeck 2003). The differences in $HlnL$ for each model were interpreted using the values presented by Kass & Raftery (1995). Dates of most recent common ancestors (MRCAs) and most ancient common ancestors (MACAs; Hayward & Stone 2006) were estimated at all nodes in the Bayesian phylogeny using BEAST 1.2 (Drummond & Rambaut 2003; available at <http://evolve.zoo.ox.ac.uk/beast/>). We calculated both MRCAs and MACAs because these estimates involve differing assumptions about ancestral sequences for a clade and cover the full span of possible timescales (Hayward & Stone 2006). Divergence was calibrated using the widely applied approximation for mitochondrial DNA of 2.3% sequence divergence per million years (Brower 1994). Although the actual rate of sequence divergence within *Andricus* may have differed from this value, this approximation should give an indication of timescales involved and does not affect the accuracy of relative time depth of nodes. Rates were kept constant across all branches as this was supported by our test for clock-like evolution. Operators were optimized using BEAUTI. The monophyly of MRCAs was not constrained as all reconstructed nodes had 100% posterior probability. The MC³ chain was run for 10-million generations under the general time reversible (GTR) model with gamma-distributed rate heterogeneity and sampled every 1000 generations after a burn-in of one-million generations. All parameters and likelihood values were assessed using TRACER 1.2.1 to estimate 95% confidence intervals and ensure a sufficient effective sample size for each parameter.

Nested clade analysis

A nested clade analysis (NCA) was performed to reveal the ecologic and evolutionary factors that are likely to have lead to the present phylogeographic distribution (Templeton 1998). A haplotype network was used to define a hierarchical set of nesting clades according to the nesting rules of Templeton *et al.* (1987) and Templeton & Sing (1993). The haplotype network was generated using statistical parsimony in TCS 1.2.1 (Clement *et al.* 2000) with the connection limit set at 95%. Loops were excluded according to the coalescent criteria of Crandall & Templeton (1993) with the geographic criteria of Pfenninger & Posada (2002) applied where this was ambiguous. Clade distances (D_c) and nested clade distances (D_n), calculated from the frequency

and locations of haplotypes, were tested against the null hypothesis of random haplotype distribution using GEODIS 2.4 (Posada *et al.* 2000). Where the null hypothesis was rejected, the cause was inferred using the latest version of the key of Templeton *et al.* (1995; available at http://inbio.byu.edu/Faculty/kac/crandall_lab/geodis.htm).

Inference of population demographic history

Pairwise mismatch plots of substitutional differences between pairs of sequences were calculated and compared against Poisson models under assumptions of constant population size (Slatkin & Hudson 1991; Rogers & Harpending 1992) and population growth-decline (Rogers & Harpending 1992) using DNASP 4.10 (Rozas *et al.* 2003). To meet the assumptions of this test, it was applied only to a subset of haplotypes inferred by NCA to represent a single contiguous, expanding population. Expected mismatch distributions are a multimodal ('ragged') exponential decline in frequency under the constant population size model and a unimodal ('bell-shaped') curve under the population growth-decline model. Comparison of data under these models allows inference of population demographic histories (Emerson *et al.* 2001). The raggedness statistic, r (Harpending 1994), was calculated to assess the smoothness of the pairwise mismatch plots. Since r is known to have low statistical power, a further test statistic, Fu's FS (Fu 1997), was calculated. This has greater statistical power to detect departures from neutrality that are indicative of population growth.

Results

Cryptic polyphyly in A. coriarius

Single gene phylogenies for *cytb* and 28SD2 each supported the division of *A. coriarius* into three clades, containing haplotypes from: (i) across the sampled range; (ii) Iran only; and (iii) Lebanon only. Alone, the *cytb* data resolved relationships within clades, but rooting of each clade within the wider phylogeny of *Andricus* was unclear. The 28SD2 clade provided resolution at the intraspecific level but there was no sequence divergence within the three clades. The following analyses thus use the partitioned two-gene dataset, as described in the Methods.

Bayesian phylogeny reconstruction of the combined *cytb*/28SD2 dataset showed that rather than representing a single monophyletic lineage, the sampled *A. coriarius* haplotypes fall into three distinct lineages (Fig. 2). One lineage (the Lebanese Clade, Fig. 2) contains only Lebanese haplotypes (for four of the five Lebanese specimens, representing three of the four Lebanese sites). The second (the Iranian Clade, Fig. 2) contains only Iranian haplotypes (for 15 of the 29 Iranian specimens, representing two of the three sites). The third (the Main Clade, Fig. 2) contains all

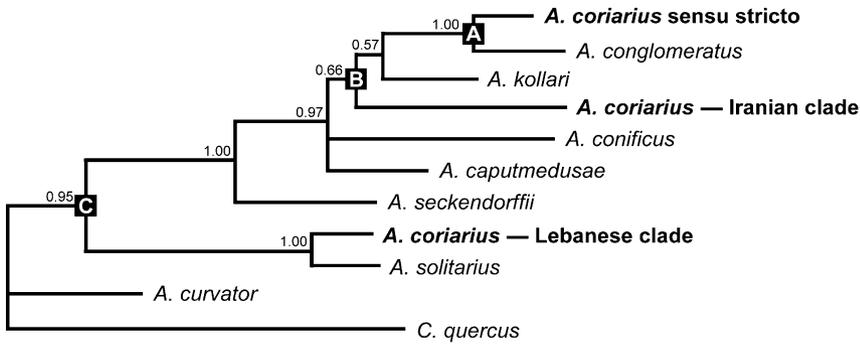


Fig. 2 Simplified two-gene Bayesian phylogeny showing the locations of the three clades of *A. coriarius* within *Andricus*. Data were partitioned by gene (28SD2 and cytochrome *b*) with the cytochrome *b* gene further partitioned by codon position. A GTR + I + Γ model of evolution was applied to each partition. Two independent MC³ runs of 2×10^6 generations were sampled every 1000 generations following a burn in period of 2×10^5 generations. Posterior probability support values are shown for each node. Labels A–C indicate nodes for which MACA dates have been estimated (Table 3).

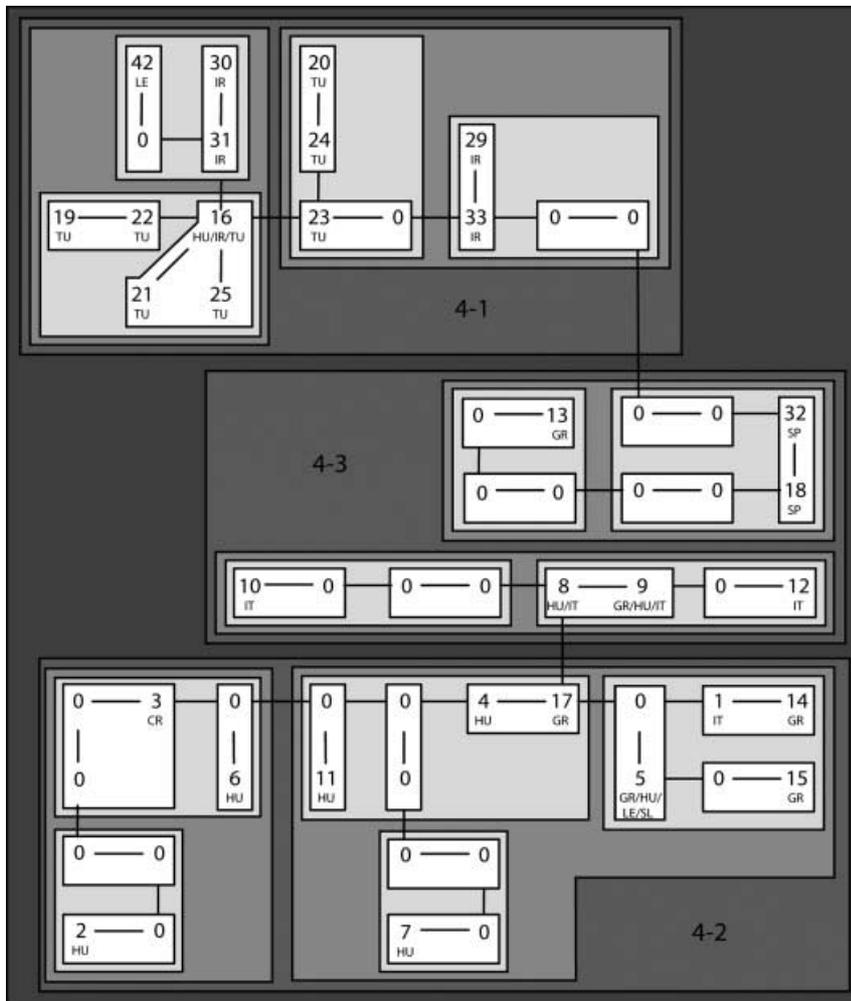


Fig. 3 Haplotype network for the *A. coriarius* sensu stricto clade estimated using statistical parsimony (95% connection level). Haplotype numbers are as in Table 1, and locations are indicated by letters in parentheses (CR, Croatia; GR, Greece; HU, Hungary; IR, Iran; IT, Italy; SL, Slovakia; SP, Spain; TU, Turkey). 0 indicates a missing haplotype.

remaining haplotypes, including haplotypes from one Lebanese individual, all 14 individuals from the Iranian site of Baneh, and all European sequences. No single site contained haplotypes in more than one of the three clades. Table 2 shows the polymorphic sites for the haplotypes in each of the three clades.

Intraspecific relationships within the Main Clade are shown in a statistical parsimony reconstruction of the haplotype network (Fig. 3). The divergent Lebanese and Iranian haplotype groups could not be connected to the rest of the network (Fig. 3) at either the 95% or the 90% confidence levels. The Main Clade forms a monophyletic

Table 3 Ages of MRCAs and MACAs calculated using BEAST. The MRCA for a given clade incorporates sequences only for that clade, while the age of the MACA is calculated by including the nearest sister taxon. Values in parentheses are 95% confidence intervals assuming a clock calibration of 2.3% pairwise divergence per million years

Node	Ancestor of	Estimated date (mya)	
		MRCA	MACA
A	Main <i>A. coriarius</i> clade	1.6 (1.0–2.2)	3.5 (2.5–4.8)
B	Iranian Clade	0.82 (0.25–1.5)	4.5 (3.1–6.1)
C	Lebanese Clade	0.30 (0.037–0.66)	8.9 (5.9–12.4)

Dating lineage divergence and longitudinal range expansion in *A. coriarius*

Bayes factor (*BF*) comparison with the nonclock model ($H \ln L = 1316.19$) for the *cytb* dataset provides strong support for the use of a coalescent clock ($H \ln L = 1310.00$; $BF = -12.38$) but does not support either the uniform or the birth-death models (BF : 23.92 and 77.60, respectively). Calculation of MRCA/MACA dates in BEAST was therefore performed under a coalescent model of population growth. Estimated ages of MRCAs and MACAs for key nodes are shown in Table 3.

The Main *A. coriarius* lineage is inferred to have diverged from the rest of the *kollari* clade between the mid-Pliocene (MACA 3.5 mya) and the Pleistocene (MRCA 1.6 mya). The Iranian and Lebanese Clades are inferred to have diverged from the main *A. coriarius* lineage long before the Pleistocene, around 4.5 mya (Pliocene) and 8.9 mya (late Miocene), respectively. Of the outgroups we used, the Lebanese Clade clusters closest to *A. solitarius*; however, we suggest that this relationship is more likely to be a result of long-branch attraction than a true relationship. Placement of the Lebanese Clade within *Andricus* requires further analysis.

Although the divergence between lineages is inferred to be relatively ancient, sequence divergence and nucleotide diversity within clades (Main Clade, 3.6%, $\pi = 0.01554$; Iranian Clade, 2.1%, $\pi = 0.00529$; and Lebanese Clade, 1.0%, $\pi = 0.00644$) are much lower than sequence divergence and nucleotide diversity between clades (Main vs. Iranian Clade, 8.3%, $\pi = 0.03427$; Main vs. Lebanese Clade, 11.3%, $\pi = 0.02884$; and Iranian vs. Lebanese Clade, 11.5%, $\pi = 0.04742$). This contrast within all three species is compatible with population bottlenecks since species divergence.

Nested clade and pairwise mismatch analyses of the Main *A. coriarius* clade

Due to the high sequence divergence between the three lineages of *A. coriarius* and the associated difficulty of linking these in a single network, NCA was attempted only for the Main *A. coriarius* clade, using the *cytb* sequence

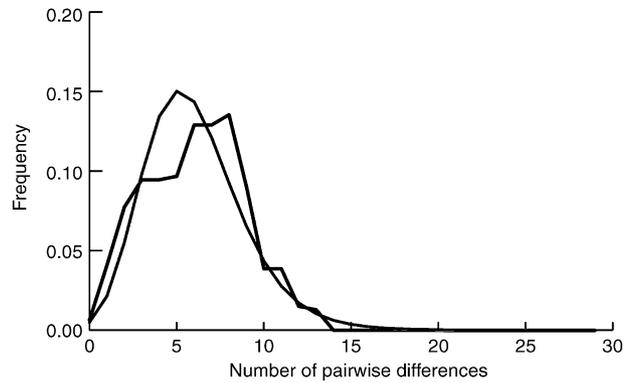


Fig. 4 Pairwise mismatch plot of the Main *A. coriarius* Clade. Black line indicate observed data. Grey line indicate the distribution expected under a population growth-decline model (θ initial = 1.414, θ final = 1000, $\tau = 4.664$).

data. The nesting structure for NCA is shown in Fig. 3. The highest root probability (11.1%) was assigned to haplotype 16, which was sampled from Hungary, Iran and Turkey. Considering all samples from each country, the greatest mean root probability (5.3%) was shared by Iran and Turkey, suggesting an eastern origin for this clade. All other countries had mean root probabilities below 3.9%. The NCA revealed statistically significant geographic associations at the fourth nesting level and at the total cladogram level. Clade 4-3 shows contiguous range expansion, while at the total cladogram level there is support for historic gradual range expansion with subsequent fragmentation or long-distance dispersal. Support for an eastern root suggests that range expansion was predominantly westwards. While we cannot distinguish between the two eastern oak species *Q. cerris* and *Q. libani* as ancestral sexual generation hosts, the association with the Iberian host *Q. suber* is probably derived.

A pairwise mismatch plot (Fig. 4) of the Main *A. coriarius* Clade has a single, smooth peak ($r = 0.0093$, $P < 0.05$), which indicates that westwards range expansion was associated with population growth by a single demographic entity. Further test statistics (Fu's $FS = -24.01$, $P < 0.05$) also indicate the departure from neutrality expected in an expanding population. The same distribution is supported in mismatch plots of the fourth nesting level subclades (not shown). Smoothly unimodal mismatch distributions are incompatible with significant population fragmentation, so it is likely that long-distance dispersal accounts for the substructure within the Main *A. coriarius* Clade.

Discussion

Cryptic lineages in *A. coriarius*

Our data show that insects reared from galls currently recognized as representing a single species in fact represent

three deeply divergent lineages. The Main *A. coriarius* Clade contains haplotypes from all sampled countries, and is either locally or regionally sympatric with divergent monophyletic lineages from each of Iran and Lebanon. The Lebanese, and possibly Iranian, Clades lie outside the monophyletic group formed by the Main *A. coriarius* Clade and its sister taxa in the *A. kollari* Clade, and *A. coriarius* (as represented by our sampled individuals) is thus paraphyletic.

Andricus coriarius was described from central Europe by Hartig in 1843 (Hartig 1843). All previous studies of gallwasps from Turkey (Dalla Torre & Kieffer 1910) eastwards through to Iran (Chodjai 1980) and the Caucasus (Maisuradze 1962) have identified insects emerging from the characteristic spiny and multichambered galls as members of this single species. However, the support for three separate lineages from nuclear and mitochondrial sequence data and the separation of two clades from the third by other recognized species suggest that there are in fact three cryptic species. Hartig's original specimens were collected in Central Europe, suggesting that the name *A. coriarius* sensu stricto rightly applies to what we have termed the Main Clade, while the Lebanese and Iranian Clades represent new species.

The discovery of these taxa and their placement within *Andricus* has implications for our understanding of the evolution of gall phenotypes. Gall morphology is controlled by gallwasp genes, and its adaptive significance remains a subject of debate (Stone & Schönrogge 2003). *Andricus coriarius* (as previously understood) represented the only member of the *A. kollari* group of gallwasps to show two significant gall characteristics: the presence of many larval chambers (multilocularity) rather than a single chamber, and the presence of many surface spines (rather than galls that are spineless). Both of these traits are thought to play a role in gall defence against natural enemies (Stone & Schönrogge 2003). A reconstruction of character evolution through the genus *Andricus* (Stone & Cook 1998; using a sequence for *A. coriarius* sensu stricto from Hungary) inferred that both multilocularity and spininess represent derived states that evolved in *A. coriarius* after its divergence from an ancestor, inducing single-chambered, spineless galls — character states present in the remaining members of the *A. kollari* group. The inference that two lineages inducing spiny, multichambered galls diverged basally to the remainder of the *A. kollari* group implies that either each represents an independent evolution of both spininess and multilocularity, or that the spineless, single-chambered galls characteristic of the *Andricus* species comprising the sister group to *A. coriarius* sensu stricto are themselves derived. Resolution of these alternatives awaits formal reanalysis of gall phenotype evolution through *Andricus* as a whole. Independent convergent evolution of both spininess and multilocularity would be of particular interest, because it has occurred independently elsewhere in *Andricus* (Stone

& Cook 1998) and in other gall-inducing insects (Stone & Schönrogge 2003). It has been proposed that natural selection has favoured the evolution of spines as a defence against natural enemies attracted to the greater concentration of prey resources within a multilocular gall (Stone & Schönrogge 2003).

The geographic origin of A. coriarius sensu stricto

Although intraspecific phylogenetic analyses were not valid, the haplotype root probabilities derived from the statistical parsimony network support a geographic origin for *A. coriarius* sensu stricto to the east of Europe. This parallels the pattern seen in *A. quercustozae* (Rokas *et al.* 2003a). There were probably multiple glacial refugia in Asia Minor and the Caucasus, and the existence of multiple mountain range barriers to gene flow is thought to underlie much of the diversity of the region (Davis 1965–85, 1971). A major faunistic and floristic divide in the region, the Anatolian Diagonal, runs from the Taurus Mountains in southeastern Turkey, northeastwards towards the Caucasus. Of the section *Cerris* oaks available for host-alternating *Andricus*, *Q. cerris* extends far to the west of this divide, while *Q. libani* is predominantly found to the south and east of it (Davis 1965–85, 1971; Yaltirik 1982). One possible explanation for the eastern restriction of the Lebanese and Iranian Clades is that they exploit only *Q. libani* as a sexual generation host, and are unable to exploit *Q. cerris*. If true, such a limit to range expansion would mirror the inability of Iberian host-alternating *Andricus* to escape their glacial refuge by making an equivalent eastwards host shift from *Q. suber* to *Q. cerris* (Stone *et al.* 2001; Hayward & Stone 2006). In contrast, the geographic distribution of haplotypes of *A. coriarius* sensu stricto implies that basal members of this lineage were able to exploit both *Q. libani* (the only host available to Iranian populations) and *Q. cerris*, and to make a subsequent host shift to *Q. suber* when they reached Iberia.

Timescale and modes of dispersal

A. coriarius sensu stricto, in common with *A. kollari* and *A. quercustozae*, diverged around 3.5 mya, before the Pleistocene glacial cycles (Table 3, Stone *et al.* 2001; Rokas *et al.* 2003a). The Iranian and Lebanese Clades are inferred to have more ancient origins in the early Pliocene and late Miocene, respectively. To an order of magnitude (given the caveats associated with assuming a given rate of mitochondrial sequence evolution), the aspects of gall structure common to the three clades of *A. coriarius* have either been conserved in independently evolving lineages for almost 10 million years, or converged over a similar timescale. Vicariance patterns in oak floras in both sections *Cerris* and *Quercus* sensu stricto suggest that the three lineages currently known as *A. coriarius* diverged at the same time as

their host oaks diversified into a characteristic western Palaearctic flora (Manos & Stanford 2001). Although the Lebanese and Iranian Clades are both ancient lineages, each is inferred to have undergone a recent bottleneck, since diversity within these clades dates from the mid-Pleistocene (Table 3). A similar pattern is seen in Iberian populations of *A. kollari*, for which the MRCA is dated with the same methods at 0.4 mya (Hayward & Stone 2006).

Nested clade analysis of *A. coriarius* sensu stricto could not reject the null hypothesis of no geographic association of haplotypes for any clades at the first three nesting levels. Such a failure to detect significant associations can arise through insufficient sampling, panmixia, or lack of genetic variation (Templeton *et al.* 1995). Within the lower level clades, each of these explanations could be sufficient to account for the lack of significant associations. A benefit of the nested statistical design is that the statistical power of the NCA is pooled at higher levels (Templeton 2004) so geographic associations can be detected in higher nesting clades. Colonization of Spain (clade 4-3) can be explained by contiguous range expansion. The inference at the total cladogram level is past gradual range expansion with limited long-distance dispersal. The suggestions of contiguous range expansion are supported by the unimodal distribution of the pairwise mismatch plot of the main *A. coriarius* clade, which infers growth as a single demographic entity. Such a pattern has been associated with western Palaearctic postglacial range expansion in other organisms, including woodmice (Michaux *et al.* 2003) and the spined loach (Culling *et al.* 2006).

Acknowledgements

We sincerely thank Gyuri Csóka (Hungary, Croatia), Rachel Atkinson and Gil McVean (Turkey) for their help in collecting material contributing to this study and Antonio Hernandez-Lopez for providing 28SD2 sequences. We would particularly like to thank Dr Rida Y. Nuwayhid of the Faculty of Engineering and Architecture, American University of Beirut, for his invaluable help in Lebanon. Richard Challis was supported by a NERC studentship held in the Institute of Evolutionary Biology, Edinburgh University. This work and Sonja Preuss were supported by NERC grants NER/B/504406/1, NER/B/S2003/00856 to GNS. The research used material collected during NERC grants GST032035 and GR/12847.

References

Aebi A, Schönrogge K, Melika G *et al.* (2006) Parasitoid recruitment to the globally invasive chestnut gall wasp *Dryocosmus kuriphilus*. In: *Galling Arthropods and Their Associates; Ecology and Evolution* (eds Ozaki K, Yukawa J, Ohgushi T, Price PW), pp. 103–122. Springer, Tokyo.

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

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

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

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.

Bryant D, Moulton V (2004) NeighborNet: an agglomerative algorithm for the construction of planar phylogenetic networks. *Molecular Biology and Evolution*, **21**, 255–265.

Chodjai M (1980) L'étude des Hyménoptères cynipides et les Espèces Cécidogènes dans la Faune des Forêts du Chêne en Iran. *Journal of the Entomological Society of Iran*, Supplement **3**, 1–67.

Clement M, Posada D, Crandall KA (2000) rcs: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

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

Crandall KA, Templeton AR (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, **134**, 959–969.

Csóka G, Stone GN, Melika G (2005) Biology, ecology and evolution of gall inducing Cynipidae. In: *The Biology, Ecology and Evolution of Gall-Inducing Arthropods* (eds Raman A, Schaefer CW, Withers TM), pp. 569–636. Science Publishers Inc., Enfield, New Hampshire, USA.

Culling MA, Janko K, Borón A, Vasil'ev VP, Côté IM, Hewitt GM (2006) European colonization by the spined loach (*Cobitis taenia*) from Ponto-Caspian refugia based on mitochondrial DNA variation. *Molecular Ecology*, **15**, 173–190.

Dalla Torre KW, Kieffer JJ (1910) Cynipidae. *Das Tierreich*, **24**. Friedlander & Sohn, Berlin, Germany.

Davis PH (ed.) (1965–85) *Flora of Turkey and the east Aegean islands*, Vols 1–9. Edinburgh University Press, Edinburgh.

Davis PH (1971) Distribution patterns in Anatolia with particular reference to endemism. In: *Plant Life of Southwest Asia* (eds Davis PH, Harper PC, Hedge IC), pp. 15–27. Botanical Society of Edinburgh, UK.

Dress AWM, Huson DH (2004) Constructing splits graphs. *IEEE Transactions on Computational Biology and Bioinformatics*, **1**, 109–115.

Drummond AJ, Rambaut A (2003) BEAST v1.2, Available from <http://evolve.zoo.ox.ac.uk/beast/>.

Emerson B, Paradis E, Thébaud C (2001) Revealing the demographic histories of species using DNA sequences. *Trends in Ecology and Evolution*, **16**, 707–716.

Ferris C, King RA, Väinölä 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.

Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **141**, 915–925.

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

- Gunduz I, Rambau RV, Tez C, Searle JB (2005) Mitochondrial DNA variation in the western house mouse (*Mus musculus domesticus*) close to its site of origin: studies in Turkey. *Biological Journal of the Linnean Society*, **84**, 473–485.
- Hancock JM, Tautz D, Dover GA (1988) Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. *Molecular Biology and Evolution*, **5**, 393–414.
- Harpending H (1994) Signature of ancient population growth in a low resolution mitochondrial DNA mismatch distribution. *Human Biology*, **66**, 591–600.
- Hartig T (1843) Zweiter nachtrag zur naturgeschichte der Gallwespen. *Zeitschrift für Entomologie*, **4**, 395–422.
- Hayward A, Stone GN (2006) Comparative phylogeography across two trophic levels: the oak gall wasp *Andricus kollari* and its chalcid parasitoid *Megastigmus stigmatizans*. *Molecular Ecology*, **15**, 479–489.
- Hewitt GM (1996) Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*, **58**, 247–276.
- Hewitt GM (1999) Post-glacial re-colonization of European biota. *Biological Journal of the Linnean Society*, **68**, 87–112.
- Hewitt GM (2004) Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society of London, Series B*, **359**, 183–195.
- Huson DH, Bryant D (2006) Application of phylogenetic networks to evolutionary studies. *Molecular Biology and Evolution*, **23**, 254–267.
- Jermiin LS, Crozier RH (1994) The cytochrome *b* region in the mitochondrial DNA of the ant *Tetraponera rufoniger* sequence divergence in Hymenoptera may be associated with nucleotide content. *Journal of Molecular Evolution*, **38**, 282–294.
- Juste J, Ibanez C, Munoz J, Trujillo D, Benda P, Karatas A, Ruedi M (2004) Mitochondrial phylogeography of the long-eared bats (*Plecotus*) in the Mediterranean Palaeartic and Atlantic islands. *Molecular Phylogenetics and Evolution*, **31**, 1114–1126.
- Kass RE, Raftery AE (1995) Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.
- Kotlik P, Bogutskaya NG, Ekmekci FG (2004) Circum Black Sea phylogeography of *Barbus* freshwater fishes: divergence in the Pontic glacial refugium. *Molecular Ecology*, **13**, 87–95.
- Maisuradze NL (1962) Studies on the oak gall wasps in Great and Small Caucasus of Azarbaijan. *Scientific Notes of the Azarbaijan State University*, **2**, 49–59 [In Russian].
- Manos PS, Stanford AM (2001) The historical biogeography of Fagaceae: tracking the Tertiary history of temperate and subtropical forests of the Northern hemisphere. *International Journal of Plant Sciences*, **162** (Suppl.), 577–593.
- Melika G, Csóka G, Pujade-Villar J (2000) Checklist of oak gall wasps of Hungary, with some taxonomic notes (Hymenoptera: Cynipidae, Cynipinae, Cynipini). *Annales Historico-Naturalis Musei Nationalis Hungarici*, **92**, 265–296.
- Michaux JR, Magnanou E, Paradis E, Nieberding C, Libois R (2003) Mitochondrial phylogeography of the woodmouse (*Apodemus sylvaticus*) in the Western Palaeartic region. *Molecular Ecology*, **12**, 685–697.
- Nardi F, Carapelli A, Dallai R, Roderick GK, Frati F (2005) Population structure and colonization history of the olive fly, *Bactrocera oleae* (Diptera, Tephritidae). *Molecular Ecology*, **14**, 2729–2738.
- Nieves-Aldrey JL (2001) *Hymenoptera; Cynipidae. Fauna Ibérica*. Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas, Madrid, Spain.
- Petit RJ, Csaikl UM, Bordacs S *et al.* (2002) Chloroplast DNA variation in European white oaks — Phylogeography and patterns of diversity based on data from over 2600 populations. *Forest Ecology and Management*, **156**, 5–26.
- Pfenninger M, Posada D (2002) Phylogeographic history of the land snail *Candidula unifasciata* (Helicellinae, Stylommatophora): fragmentation, corridor migration, and secondary contact. *Evolution*, **56**, 1776–1788.
- Posada D, Crandall KA, Templeton AR (2000) GEODIS: a program for the cladistic nested analysis of the geographical distribution of genetic haplotypes. *Molecular Ecology*, **9**, 487–488.
- Rogers AR, Harpending H (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.
- Rokas A, Nylander JAA, Ronquist F, Stone GN (2002) A maximum likelihood analysis of eight phylogenetic markers in gallwasps (Hymenoptera: Cynipidae); implications for insect phylogenetic studies. *Molecular Phylogenetics and Evolution*, **22**, 206–219.
- Rokas A, Atkinson RJ, Webster L, Csóka G, Stone GN (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, Melika G, Abe Y, Nieves-Aldrey JL, Cook JM, Stone GN (2003b) Lifecycle closure, lineage sorting, and hybridisation revealed in a phylogenetic analysis of European oak gallwasps (Hymenoptera: Cynipidae: Cynipini) using mitochondrial sequence data. *Molecular Phylogenetics and Evolution*, **26**, 36–45.
- Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.
- Rozas J, Sanchez-Del Barrio JC, Messeguer X, Rozas R (2003) DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.
- Schmitt T, Röber S, Seitz A (2005) Is the last glaciation the only relevant event for the present genetic population structure of the meadow brown butterfly *Maniola jurtina* (Lepidoptera: Nymphalidae)? *Biological Journal of the Linnean Society*, **85**, 419–431.
- Slatkin M, Hudson RR (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.
- Stone GN, Cook JM (1998) The structure of cynipid oak galls: patterns in the evolution of an extended phenotype. *Proceedings of the Royal Society of London, Series B*, **265**, 979–988.
- Stone GN, Schönrogge K (2003) The adaptive significance of insect gall morphology. *Trends in Ecology and Evolution*, **18**, 512–522.
- Stone G, Atkinson R, Rokas A, Csóka G, Nieves-Aldrey JL (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, Atkinson RJ, Bellido D, Pujade-Villar J (2002) The population biology of oak gall wasps (Hymenoptera: Cynipidae). *Annual Review of Entomology*, **47**, 633–668.
- 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.
- Templeton AR (2004) Statistical phylogeography: methods of evaluating and minimising inference errors. *Molecular Ecology*, **13**, 789–809.

- Templeton AR, Sing CF (1993) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics*, **134**, 659–669.
- Templeton AR, Boerwinkle E, Sing CF (1987) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of alcohol dehydrogenase activity in *Drosophila*. *Genetics*, **117**, 343–351.
- Templeton AR, Routman E, Phillips C (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, *Ambystoma tigrinum*. *Genetics*, **140**, 767–782.
- Tzedakis PC (1993) Long-term tree populations in northwest Greece through multiple quaternary climatic cycles. *Nature*, **364**, 437–440.
- Yaltirik F (ed.) (1982) *Fagaceae. Flora of Turkey and the eastern Aegean Islands*. Edinburgh University Press, Edinburgh, UK.

This project is part of a long-term research programme being carried out in Graham Stone's lab on the population biology, phylogeography and phylogeny of insect-plant interactions, using oak gallwasps as a model system. Richard Challis' main interest is the application of 'tree-thinking' to a variety of evolutionary ecological questions relating to gallwasps at various taxonomic levels. Antonis Rokas' current interests are in comparative and functional genomics. Alexandre Aebi is currently working on patterns of transmission of *Wolbachia* and other symbionts in oak gallwasp communities. Serap Mutun works on organismal phylogenetics and molecular evolution, particularly of insects. She has recently initiated a research program on the phylogeographic impacts of the Anatolian Diagonal. Jose-Luis Nieves-Aldrey has a long-term interest in the ecology, evolution and taxonomy of insect-plant interactions. Ebrahim Sadeghi is head of a unit working on applied aspects of insect-plant interactions. Majid Tavakoli works on applied insect-plant issues and has a long-term interest in the biology of gall-inducing insects, particularly on oak.
