Turkish Journal of Zoology

Volume 39 | Number 5 Article 1

1-1-2015

Investigation of genetic variation among Turkish populations ofAndricus lignicola using mitochondrial cytochrome b gene sequence data

SERAP MUTUN

HÜLYA KARAGÖZOĞLU

TÜBİTAK

Turkish Journal of Zoology

http://journals.tubitak.gov.tr/zoology/

Research Article

Turk J Zool
(2015) 39: 721-733
© TÜBİTAK
doi:10.3906/zoo-1408-60

Investigation of genetic variation among Turkish populations of *Andricus lignicola* using mitochondrial cytochrome b gene sequence data

Serap MUTUN*, Hülya KARAGÖZOĞLU
Department of Biology, Faculty of Science and Arts, Abant İzzet Baysal University, Bolu, Turkey

Received: 25.08.2014 • **Accepted/Published Online:** 03.02.2015 • **Printed:** 30.09.2015

Abstract: Genetic diversity and diversification of the Anatolian populations of *Andricus lignicola* (Hartig, 1840) (Hymenoptera: Cynipidae) were investigated using 433 base pairs of the mitochondrial cytochrome b gene. Eighteen distinct haplotypes from 15 populations were determined. Analyses indicated average haplotype and nucleotide diversity as 0.325 and 0.008, respectively. Phylogenetic analyses conducted through the application of maximum likelihood and maximum parsimony produced similar topologies with two major clade structures supported by high bootstrap values. Bayesian analysis produced more polytomies without major basal groupings of the haplotypes. Parsimony network analysis revealed four haplogroups, and the largest group comprised half of the haplotypes. Our preliminary ABGD analysis implied the presence of four hidden lineages with the possibility of a cryptic species complex within *A. lignicola*. Hierarchical F-statistics (AMOVA) for detecting partitioning of molecular variation supported the presence of a conspicuous amount of genetic differentiation among these four primary groups.

Key words: *Andricus lignicola*, mtDNA, gall wasp, cyt b gene, Anatolia

1. Introduction

In the last several decades fascinating advances in the application of novel molecular tools have made it easy to investigate genetic variation at both interspecific and intraspecific levels (Loxdale and Lushai, 1998; Arias et al., 2006). Revealing the genetic diversity and phylogeographic structure of populations of Turkish species through the applications of molecular markers has been of great interest because Anatolia is a center of genetic diversity and origin of many European species (Rokas et al., 2003a). The high number of species and genetic diversity in Turkey is closely associated with the presence of different topographical zones and the geologic history of the area together with its connection with three continents (Demirsoy, 2008; Kekeçoğlu and Soysal, 2010). Moreover, periodic glacial and interglacial periods such as Pleistocene cycles are important factors impacting the animals and plants distributed in Turkey (Çıplak, 2008).

One of the molecular markers proven to be a robust route for investigating the effects of past and current factors that shape the contemporary genetic structure of species is mitochondrial DNA (mtDNA) (Avise, 2000). Large bodies of literature from different animal taxa have been accumulated through mtDNA (Hewitt, 2004; Hickerson et al., 2010) and have proven that it is a powerful tool among diverse groups of organisms including oak gall wasps (Stone et al., 2002; Rokas et al., 2003b). In particular, mitochondrial cytochrome b gene has been utilized for estimating genetic variability of sampled populations and population genetic structure (Cook et al., 2002). It has also been shown that multiple haplotypes of the cyt b gene have been ordered phylogenetically with respect to the geographic location of oak gall wasp populations (Rokas et al., 2003b).

Andricus lignicola (Hartig, 1840) is one of the oak gall wasp species from the family Cynipidae, which is represented by 1400 species. It is distributed throughout southern Europe from northeastern Spain to the Balkans and Turkey and produces two generations per year (Melika, 2006). Spherical and unilocular (with a single larval chamber), asexual galls are induced on the shoots of white oaks, especially *Quercus petraea, Q. robur*, and *Q. pubescens*. Four to five galls may be found on the same shoot of a tree. In the present study we investigated intraspecific genetic variation of the asexual generation of Anatolian *A. lignicola* populations, and examined the phylogeographic structure of the species in Turkey.

* Correspondence: smutun@ibu.edu.tr

2. Materials and methods

2.1. Sampling, DNA isolation, and gene amplification

A total of 117 *A. lignicola* individuals representing 15 populations were collected during the fall from 2010 and 2012 (Table 1). We collected our galls randomly in the field and tried to use only a single collected gall per tree to reduce the possibility of studying progeny of the same female. Adults were used individually for DNA isolation using DNeasy Tissue Kit (QIAGEN). A 433-bp fragment of the mitochondrial cytochrome b (cyt b) gene was amplified using the CB1/CB2 primers following previously described protocols (Stone et al., 2007). Each PCR reaction contained 0.77 µL of dNTP (10 mM), 0.54 µL of CB1 and CB2 primers (20 mM each), 1.54 µL of DNA, 3.84 µL of 10X reaction buffer, 3.07 µL of $MgCl_2$ (25 mM), and 0.5 µL of Taq DNA polymerase enzyme. The appropriate amount of double-distilled sterile H_2O was used to complete the volume of samples to 40 µL. PCR reactions were carried out in a PTC-200 DNA engine (MJ Research) thermal cycler, and amplification was performed under the following conditions: 3 min at 94 °C, 35 cycles of 30 s at 94 °C, 1 min at 50 °C, 2 min at 72 °C, and a final extension of 10 min at 72 °C. PCR products were visualized on 1% agarose gel buffered

Table 1. Sampled populations of *Andricus lignicola* with their abbreviations, coordinates, and population sizes.

Population	Abbreviation	Coordinates	Sample size
Düzce	DUZ	N 40°541.47′ E 31°103.79′	8
Denizli	DEN	N 38°340.63′ E 29°740.94′	7
Çanakkale	CAN	N 40°240.05′ E 26°560.24′	10
Manisa	MAN	N 39°164.14′ E 28°631.54′	4
İstanbul	IST	N 41°094.02′ E 29°757.22′	4
Balıkesir	BAL	N 39°557.15′ E 27°199.89′	3
Eskişehir	ESK	N 39°353.67′ E 30°660.88′	9
Afyon	AFY	N 39°058.83′ E 30°545.17′	6
Antalya	ANT	N 36°413.42′ E 29°692.70′	7
Konya	KON	N 37°455.63′ E 31°541.86′	10
Kayseri	KAY	N 38°370.63′ E 36°231.43′	9
Kırıkkale	KIR	N 40°101.99′ E 33°718.34′	10
Kahramanmaraş	KAH	N 38°179.73′ E 36°779.15′	10
Uşak	USK	N 38°733.75′ E 29°217.66′	10
Kütahya	KUT	N 38°957.59′ E 29°329.71′	10

with tris-boric acid-EDTA (TBE) with a 1 kb DNA ladder. Both strands were sequenced to minimize PCR artifacts, ambiguities, and base-calling errors. Obtained haplotype sequences were deposited in GenBank (accession no.: KM211549-KM211566).

2.2. Data analysis

2.2.1. Sequence alignment, genetic variation, and population structure

Chromatograms were visually reviewed, and sequences were transferred to the program BioEdit 6.2 for haplotype determination (Hall, 1999). Nucleotide and amino acid composition, population genetic statistics, the number of polymorphic sites (S), nucleotide (π) and haplotype (h) diversity, and the number of pairwise nucleotide differences (k) were calculated using the program DnaSP 4.0 (Librado and Rozas, 2009). Since our sampling size is uneven among localities, the Chao-1 (S^*_1) estimator (Chao, 1984) was calculated to compare haplotype richness across collection sites through the application of rarefaction analysis using the rarefaction calculator (www2.biology.ualberta.ca/jbrzusto/rarefact.php#ColCod1994) (Gebiola et al., 2014). Transition and transversion ratios (ti/tv) were estimated by MEGA 5.0 software (Tamura et al., 2007). Mismatch distribution analyses were done using the program DnaSP 4.0 (Librado and Rozas, 2009). To test the deviation from neutrality we calculated Tajima's D (Tajima, 1989). Although Tajima's D statistics are very commonly used for assessing population demographic patterns, they are less powerful than some alternatives because they do not consider departures from neutrality, and other methods can be more sensitive indicators of population expansion than Tajima's D. Thus, for demographic patterns of populations we calculated Tajima's D (Tajima, 1989), Fu's Fs, and Li's D (Fu and Li, 1993) based on the infinite site model without recombination.

2.2.2. Phylogenetic and phylogeographic analysis

Multiple methods have been employed to infer phylogenetic relationships among *A. lignicola* haplotypes. Maximum parsimony (MP) and maximum likelihood (ML) analyses were performed in PAUP* 4.0b10 (Swofford, 2002). Bootstrap support for the MP trees was calculated using 10,000 replicates. For ML analyses a model of evolution was determined through the Akaike information criterion (AIC) implemented in jModeltest 3.7 (Posada and Crandall, 1998). ML analysis was performed using the PAUP* software with a heuristic search, and bootstrap support was calculated using 100 replicates (heuristic search, stepwise addition, 10 random-sequence addition replicates, and TBR branch swapping). Bayesian inference (BI) was performed to employ a metropolis-coupled, Markov chain Monte Carlo (MCMC) sampling approach using the software MrBayes 3.1 (Huelsenbeck and Ronquist, 2001). In all phylogenetic analyses sequences of *Andricus kollari* and *A. caliciformis* obtained from GenBank were used to root phylogenetic trees and give a direction to the evolutionary changes in the in-group species.

The TCS computer program was employed to further check the evolutionary relationships among *A. lignicola* haplotypes by means of a statistical parsimony network analysis (Clement et al., 2000). Haplotype networks are often better suited than phylogenetic trees to depict relationships within species, as gene flow may lead to a reticulate, rather than a hierarchical or treelike, structure between populations. Moreover, in networks, internal nodes have clear biological meaning as persistent ancestral haplotypes (Posada and Crandall, 2001).

Since there were unusually high numbers of pairwise nucleotide differences among some *A. lignicola* haplotypes, and the resulting network supported four major haplogroups that could not be connected under the 95% connection limit, we sought cryptic lineages. Although the COI gene is traditionally used for searching for hidden species, we performed this search as a preliminary analysis using our current dataset. For this purpose, we submitted our data as FASTA sequences to the Automatic Barcode Gap Discovery (ABGD) website (http://wwwabi.snv.jussieu.fr/public/abgd/abgdweb.html) (Puillandre et al., 2012). The ABGD approach does not require threshold distances to be set since it automatically finds the distance where the barcoding gap is present. We set P (prior intraspecific divergence) from 0.001 to 0.1 and steps to 10, X (minimum relative gap width) to 1.5, number of Nb bins (for distance distribution) to 20, and we selected the Kimura (K80) model with ti/tv set to 1.5, as we calculated this ratio for our dataset.

To further test whether genetic variation is geographically and hierarchically structured, populations were assembled into different groupings, and an analysis of molecular variance (AMOVA) was performed based on the Euclidian distances using the program Arlequin 3.0 (Excoffier et al., 2005). Two different grouping schemes were tested; in the first trial all populations were accepted as a single group, and in the second trial all populations were divided into four groups that reflected the groupings after haplotype network and ABGD analysis results. The significance of the variance components was calculated using 1000 random permutations of the data set.

3. Results

3.1. Genetic variation in *Andricus lignicola* populations

Amplification results of 117 individuals yielded 18 haplotypes with no indels, nonsense mutations, or stop codons. In the sequences of 433 bp of cyt b gene segment, 375 sites were constant and 58 characters were variable. Of the total variable sites, 21 characters were parsimony

uninformative and 37 characters were parsimony informative. Nucleotide frequencies in the haplotypes were 34%, 11%, 9%, and 44% for A, C, G, and T, respectively. There were multiple hits at eight sites (127, 142, 146, 272, 304, 376, 415, and 421). Among other variable sites, there were 30 transitions and 20 transversions (ti/tv = 1.5). The protein coding region contained 16 amino acid replacements without any indels or nonsense mutations in the translated protein sequences.

When haplotypes and their frequencies were examined we found that the most abundant haplotype was H1, which was detected in 36 individuals representing 8 populations (Table 2). Three haplotypes (H4, H5, and H6) were found as private haplotypes. Interestingly, the Çanakkale population had two of the private haplotypes (H5 and H6). When the populations were examined with respect to their haplotype number the Kütahya population had the highest number of haplotypes (n = 4) followed by Çanakkale, Kahramanmaraş, Uşak, and Kayseri (n = 3). The Balıkesir, İstanbul, Konya, and Manisa populations each had two distinct haplotypes. The remaining populations (Afyon, Antalya, Denizli, Düzce, Eskişehir, and Kırıkkale) had a single type of haplotype in each population. In our study, haplotype richness was uncorrelated with the sampling effort (Spearman correlation: $r_S = 0.161$, n = 117, P = 0.56). Rarefaction analysis results showed that the Chao-1 estimator would not be calculated for 10 populations, in which there were no doubleton haplotypes ($S^*_1 = -1$). However, Kütahya ($S^*_1 = 4.5 \pm 1.12$), Uşak ($S^*_1 = 3.0 \pm 0.01$), Kayseri ($S^*_1 = 3.0 \pm 0.01$), Balıkesir ($S^*_1 = 2.5 \pm 1.12$), and İstanbul ($S^*_1 = 2.0 \pm 0.01$) displayed a range of Choa-1 estimator values.

The genetic diversity estimates of *A. lignicola* populations showed that haplotype diversity varied from 0.00 to 0.777 with an average of 0.32519 (Table 3). The highest haplotype diversity ($h = 0.7778$) was detected in the Kütahya population, followed by the Uşak ($h = 0.6889$) and Kayseri, Balıkesir, and İstanbul populations ($h = 0.6667$). Six of the remaining populations showed no haplotype diversity due to the occurrence of only a single type of

Table 2. Haplotypes and their frequencies detected in each *A. lignicola* population. n: number of specific haplotype found in the population. Population abbreviations are given in Table 1.

Haplotype	AFY	ANT	BAL	CAN	DEN	DUZ	ESK	IST	KAH	KAY	KIR	KON	KUT	MAN	USK	n
H1	6				7	8	9	2	1				2	1		36
H2		7														7
H3			2													2
H4			1													1
H5				1												1
H6				1												1
H7				8												8
H8								2								2
H9									6	2			1			9
H10									3							3
H11										5						5
H12										2						2
H13											10					10
H14												4	3		5	12
H15												6				6
H16													4	3		7
H17															3	3
H18															2	2
Total	6	7	3	10	7	8	9	4	10	9	10	10	10	4	10	117

Table 3. Genetic diversity estimates of each population. n: number of samples per population, N_h: number of haplotypes, h: haplotype diversity, π: nucleotide diversity.

Population	n	N_h	h	π
Afyon	6	1	0.0000 ± 0.0000	0.000000 ± 0.000000
Antalya	7	1	0.0000 ± 0.0000	0.000000 ± 0.000000
Balıkesir	3	2	0.6667 ± 0.3143	0.001540 ± 0.001920
Çanakkale	10	3	0.3778 ± 0.1813	0.001283 ± 0.001291
Denizli	7	1	0.0000 ± 0.0000	0.000000 ± 0.000000
Düzce	8	1	0.0000 ± 0.0000	0.000000 ± 0.000000
Eskişehir	9	1	0.0000 ± 0.0000	0.000000 ± 0.000000
İstanbul	4	2	0.6667 ± 0.2041	0.001540 ± 0.001728
K.Maraş	10	3	0.6000 ± 0.1305	0.011855 ± 0.007097
Kayseri	9	3	0.6667 ± 0.1318	0.002181 ± 0.001861
Kırıkkale	10	1	0.0000 ± 0.0000	0.000000 ± 0.000000
Konya	10	2	0.5333 ± 0.0907	0.039415 ± 0.021692
Kütahya	10	4	0.7778 ± 0.0907	0.030793 ± 0.017133
Manisa	4	2	0.5000 ± 0.2652	0.001155 ± 0.001432
Uşak	10	3	0.6889 ± 0.1038	0.042494 ± 0.023320

haplotype in these populations. Nucleotide diversity ranged between 0.00 and 0.042494. The average nucleotide diversity was calculated as π = 0.0087955 (0.8%). The Uşak population displayed the highest nucleotide diversity estimate (π = 0.042494) followed by the Konya population (π = 0.039415). The mismatch distribution, including all samples, indicated a bimodal profile. The Harpending raggedness index was low (r = 0.0044) but not significant. Overall, Tajima's D neutrality test (Tajima's D = 0.47873, P > 0.10) and Fu and Li tests (D^* = 0.41064, P > 0.10; F^* = 0.50013, P > 0.10) were not significant.

A pairwise comparison among *A. lignicola* haplotypes was conducted to determine the sequence differences. Among all pairwise comparisons, the sequence divergence varied from 0.02% to 10.6% (1 to 46 bp, respectively) (Table 4). The most divergent haplotypes H5 (n = 1, Çanakkale population) and H18 (n = 2, Uşak population) were separated from each other by 46 nucleotides. The least divergence, with 1 bp difference, was determined between H3 (n = 2 Balıkesir) and H4 (n = 1 Balıkesir), H1 (common haplotype) and H8– (n = 2 İstanbul) H16 (n = 4 Kütahya, n = 3 Manisa), H6 (n = 1 Çanakkale) and H7 (n = 8 Çanakkale), H9 (n = 6 Kahramanmaraş, n = 2 Kayseri, n = 1 Kütahya) and H10 (n = 3 Kahramanmaraş)–H11 (n = 5 Kayseri)– H12 (n = 2 Kayseri), H14 (n = 4 Konya, n = 3 Kütahya, n = 5 Uşak), and H17 (n = 3 Uşak) haplotypes.

The pairwise F_{st} calculations showed significant genetic differentiation among some populations (Table 5). In particular, two locations showed complete differentiation from some other populations; Antalya was significantly different from Afyon, Denizli, Düzce, Eskişehir, and Kırıkkale (F_{st} = 1); and Kırıkkale was different from Afyon, Antalya, Denizli, Düzce, and Eskişehir (F_{st} = 1). Some of the populations had no genetic differentiation (F_{st} = 0) from each other (Afyon and Denizli, Düzce and Eskişehir populations) because these populations share a single haplotype (H1). Other populations displayed genetic differentiation values on a scale of 0 and 1, with statistically significant support (P < 0.001) indicating some degrees of differentiation from each other.

3.2. Phylogenetic relationships among *Andricus lignicola* haplotypes

Using PAUP* 4.0b for estimation of phylogenetic relationships among 18 *A. lignicola* haplotypes, MP and ML produced similar tree topologies with different bootstrap values; thus, only a single phylogenetic tree was shown in Figure 1. MP is a consensus tree of 123 shortest trees, which was produced with CI = 0.691 and 139-step length. For ML analysis, jModeltest was used to determine the mutational model that best approximated the sequence evolution of the data set and calculate the transition and transversion ratios; it identified HKY + I

Table 4. Pairwise comparisons among 18 haplotypes. Percent differences are shown in the lower part of the diagonal, and the net number of nucleotide differences is given in the upper part of the diagonal.

	H1	H2	H3	H4	H5	H6	H7	H8	H9	H10	H11	H12	H13	H14	H15	H16	H17	H18
H1		25	7	6	18	5	4	1	24	23	25	25	6	34	35	1	35	36
H2	0.057		26	25	33	22	23	26	18	17	19	19	23	30	31	26	31	32
H3	0.016	0.060		1	17	4	3	8	23	22	24	24	5	32	33	7	33	34
H4	0.013	0.057	0.002		16	3	2	7	22	21	23	23	4	31	32	6	32	33
H5	0.041	0.076	0.039	0.036		13	14	19	35	34	36	36	16	44	45	19	45	46
H6	0.011	0.050	0.009	0.006	0.030		1	6	23	22	24	24	3	33	34	6	34	35
H7	0.009	0.053	0.006	0.004	0.032	0.002		5	22	21	23	23	2	32	33	5	33	34
H8	0.002	0.060	0.018	0.016	0.043	0.013	0.011		25	24	26	26	7	33	34	2	34	35
H9	0.055	0.041	.053	0.050	0.080	0.053	0.050	0.057		1	1	1	22	31	32	25	32	33
H10	0.053	0.039	0.050	0.048	0.078	0.050	0.048	0.055	0.002		2	2	21	30	31	24	31	32
H11	0.057	0.043	0.055	0.053	0.083	0.055	0.053	0.060	0.002	0.004		2	23	32	33	26	33	34
H12	0.057	0.004	0.055	0.053	0.083	0.055	0.053	0.060	0.002	0.004	0.004		23	32	33	26	33	34
H13	0.013	0.053	0.011	0.009	0.036	0.006	0.004	0.016	0.050	0.048	0.053	0.053		33	34	7	34	35
H14	0.078	0.069	0.073	0.071	0.101	0.076	0.073	0.076	0.071	0.069	0.073	0.073	0.076		2	33	1	2
H15	0.080	0.071	0.076	0.073	0.103	0.078	0.076	0.078	0.073	0.071	0.076	0.076	0.078	0.004		34	3	4
H16	0.002	0.060	0.016	0.013	0.043	0.013	0.011	0.004	0.057	0.055	0.060	0.060	0.016	0.076	0.078		34	35
H17	0.080	0.071	0.076	0.073	0.103	0.078	0.076	0.078	0.073	0.071	0.076	0.076	0.078	0.002	0.006	0.078		3
H18	0.083	0.073	0.078	0.078	0.076	0.106	0.080	0.078	0.080	0.076	0.073	0.078	0.078	0.080	0.004	0.009	0.006	

Table 5. F_{st} values calculated in a pairwise manner to compare 15 *A. lignicola* populations. Population abbreviations are presented in Table 1.

	AFY	ANT	BAL	CAN	DEN	DUZ	ESK	IST	KAH	KAY	KIR	KON	KUT	MAN
AFY														
ANT	1.000													
BAL	0.971	0.993												
CAN	0.314	0.985	0.804											
DEN	0.000	1.000	0.974	0.919										
DUZ	0.000	1.000	0.977	0.924	0.000									
ESK	0.000	1.000	0.979	0.928	0.000	0.000								
IST	0.441	0.991	0.906	0.872	0.481	0.515	0.544							
KAH	0.848	0.836	0.800	0.859	0.858	0.866	0.873	0.821						
KAY	0.976	0.971	0.962	0.967	0.978	0.979	0.980	0.965	0.115					
KIR	1.000	1.000	0.973	0.879	1.000	1.000	1.000	0.973	0.872	0.980				
KON	0.661	0.622	0.557	0.695	0.679	0.694	0.708	0.610	0.452	0.534	0.707			
KUT	0.250	0.661	0.287	0.410	0.274	0.296	0.315	0.195	0.352	0.527	0.479	0.414		
MAN	0.736	0.993	0.914	0.888	0.760	0.780	0.979	0.533	0.825	0.967	0.981	0.611	0.177	
USK	0.626	0.577	0.514	0.661	0.645	0.661	0.676	0.572	0.329	0.414	0.673	0.037	0.347	0.574

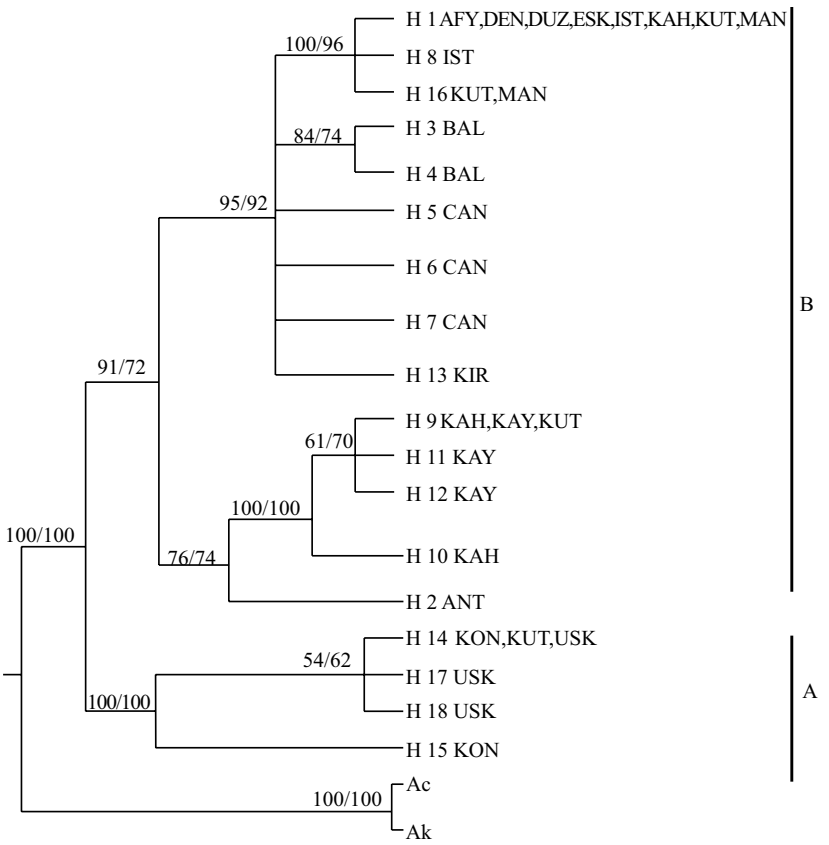

Figure 1. The consensus tree of both MP/ML analyses for the cyt b gene region of *A. lignicola*. Bootstrap values are shown on the branches for both MP and ML, respectively. Outgroup haplotypes: Ac (*Andricus caliciformis*) and Ak (*Andricus kollari*).

(Hasegawa et al., 1985) as the best fit model to the data set (I = 0.3778). Thus, this substitution model was utilized in ML and BI analysis. ML and MP trees produced two well-supported clades: clade A is composed of a basally located haplotype (H15) from the Konya population and a small polytomous group of three haplotypes (H14, H17, and H18). However, clade B is larger and composed of two further subclades. Evolutionary relationships were not clearly resolved in small subclades; however, in clade B in the large polytomous part, in addition to the presence of a commonly shared haplotype (H1), all other haplotypes are geographically restricted to the western populations. In spite of a monophyletic grouping of two Balıkesir haplotypes (H3 and H4), the relationships of all other haplotypes showed polytomy that could be due to insufficient time elapsed since divergence between lineages or incomplete lineage sorting (Avise, 2000). Some of the haplotypes representing westerly populations such as H14, H17, H18, and H15 (from Konya, Kütahya, and Uşak) seem to be well-separated from other haplotypes.

The tree resulting from the Bayesian analysis (BI) is given in Figure 2 with posterior probabilities on the branches. In the Bayesian tree there is polytomy at the basal part, which comprises H16 (from Kütahya and Manisa), H8 (from İstanbul), H1 (shared common haplotype from Afyon, Denizli, Düzce, Eskişehir, İstanbul, Kahramanmaraş, Kütahya, and Manisa), and a large clade that covers the rest of the haplotypes. In the large clade there is also polytomy composed of H13 (KIR) and H7 (CAN), a monophyletic group including H5 and H6 from Çanakkale population, another monophyletic group including H3 and H4 from Balıkesir, and the third lineage-making polytomy that includes the remaining haplotypes. Within this lineage two haplogroups are observed to be monophyletic. Of these, the first subclade is composed of a small polyphyletic group, which includes H9 (from Kahramanmaraş, Kayseri, and Kütahya), H10 (from Kahramanmaraş), and H11 and H12 (both from Kayseri). Another polytomic group has a basally located haplotype H2 from Antalya and the small polytomic group composed of H14 (Konya, Kütahya, and Uşak) and H15, H16, and H17 (Uşak).

The haplotype network analysis shown in Figure 3 produced three main haplogroups with an additional

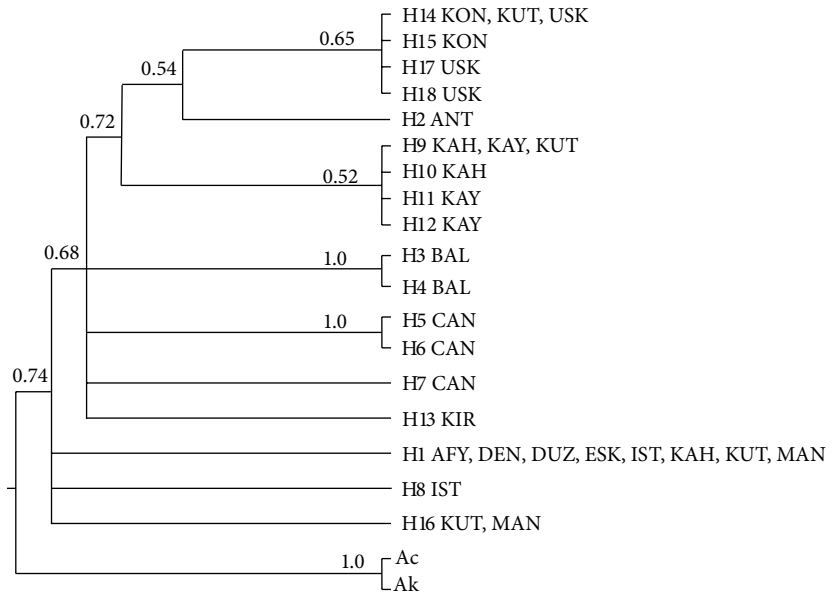

Figure 2. Bayesian analysis tree. Posterior probability values are given on the branches. Outgroup haplotypes: Ac (*Andricus caliciformis*) and Ak (*Andricus kollari*).

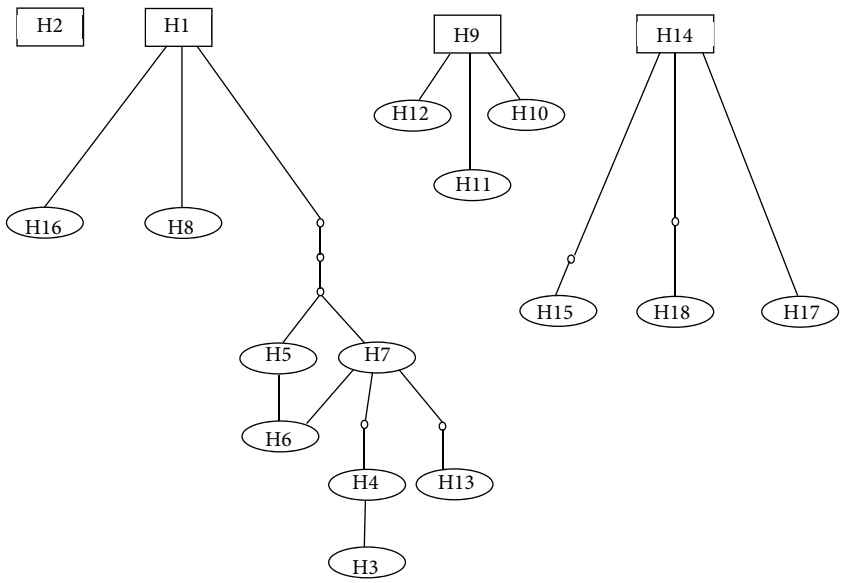

Figure 3. Parsimony network analysis of *A. lignicola* haplotypes. The haplogroups cannot be connected under the 95% limit. Haplotype numbers are shown in circles. Line length and circles are proportional to the number of mutational steps and frequency of the haplotype, respectively. The central haplotypes with the highest outgroup weights are depicted in squares. Small circles represent steps between observed haplotypes.

single haplotype (H2 from Antalya) that could not be connected under the 95% connection limit. H2, with the uppermost substitution value, could not be grouped with other haplotypes. In the obtained TCS network haplotypes of the largest cluster seemed to be derived from H1, which was the most common and, geographically, the most widely distributed haplotype. In the large cluster, H6 and H8 did not connect to any haplotypes other than H1. However, H1 was connected to three hypothetical haplotypes that provided a connection to several other haplotypes, such as H5 and H7. In addition to H5, H6, and H7 haplotypes, H4, H3, and H13 are clustered in the large

group of haplotypes of the network (Figure 2). The third haplotype cluster, which formed a small grouping with the uppermost substitution value, belonged to H9. Haplotype 9 seemed to be connected to H12, H10, and H11. The last haplogroup contains H14 with the uppermost value, which was connected to two hypothetical haplotypes with three haplotypes (H15, H17, and H18).

The same haplotype groupings were also produced by ABGD analysis and generated four groups, including recursive evaluation of group splitting with the initial partition with prior maximal distance P = 1.67e-03. The ABGD analysis united H1, H3–H8, H13, and H16 as group 1; H2 as group 2; H9 and H1–H12 as group 3; and H14–H15 and H17–H18 as group 4. The genetic distances were d = 0.0062, 0.021, and 0.0036 for the first, third, and fourth groups, respectively. The genetic distance could not be calculated for the second group due to the presence of a single haplotype in this group. The calculated pairwise genetic distances between each of the four detected groups were 0.037 (between groups 1 and 2), 0.0036 (between groups 1 and 3), and 0.0054 (between groups 1 and 4). Likewise, group 4 showed a distance of 0.049 from group 2 and 0.051 from group 3.

AMOVA for revealing population structuring was conducted through several trials in groupings of the populations. They were tested for significant clustering and partitioning of the genetic variation at two/three levels; however, only two of the trial schemes are included in Table 6. In the first trial scheme all populations were considered a distinct group, and it was determined that variation was partitioned among groups with 71.43%, and the remaining variation (28.57%) was at the "within population" level. These results were statistically significant (P < 0.001) (Table 6a). On the other hand, when all the populations were divided into four groups based on the groupings obtained after ABGD and haplotype network analyses, an even greater level of genetic partitioning value with statistically significant support was detected; 74.12% of variance components were recovered among groups, 16.03% among populations within groups, and 9.86% within population (Table 6b).

4. Discussion

4.1. Genetic variation of *Andricus lignicola* populations

The base pair composition in *A. lignicola* cyt b haplotypes indicates that the collected data are genuine mitochondrial DNA, since anti-G bias is a characteristic of mitochondrial DNA genes (Zhang and Hewitt, 1996). Neither pseudogenes nor heteroplasmy were detected when *A. lignicola* haplotypes were compared to the corresponding cyt b region of other insect species. In insect mitochondrial protein-coding gene segments, transitions are observed more often than transversions because of poor or deficient mtDNA repair mechanisms and tautomeric base pairing (Brown et al., 1979). In the sequenced region of *A. lignicola*, the rate of substitution was well within the range of transition and transversion ratios of other insect species (Jermin and Crozier, 1994).

Table 6. AMOVA results of *A. lignicola* populations obtained from the Arlequin program (d.f.: degrees of freedom).

a) One group: all populations are accepted as a distinct group.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	F-index	P-value
Among groups	14	702.614	6.16084	71.43	0.43080	<0.001
Within population	102	251.394	2.46465	28.57	0.17114	<0.001

b) Four groups: group 1: Afyon, Balıkesir, Çanakkale, Denizli, Düzce, Eskişehir, İstanbul, Kırıkkale, Manisa; group 2: Antalya; group 3: Kahramanmaraş, Kayseri, Kütahya; group 4: Konya, Uşak.

Source of variation	d.f.	Sum of squares	Variance components	Percentage of variation	F-index	P-value
Among groups	3	749.280	9.34722	74.12	0.74117	<0.001
Among populations within groups	11	180.601	2.02118	16.03	0.61918	<0.001
Within population	102	126.794	12.61148	9.86	0.90143	<0.001

A total of 18 haplotypes were detected out of the 117 individuals of *A. lignicola*. Haplotype 1 (H1) is the most common haplotype and is found in 36 individuals representing eight populations (Afyon, Denizli, Düzce, Eskişehir, İstanbul, Kahramanmaraş, Kütahya, and Manisa). The frequency and wide distribution area across populations of H1 may imply that this haplotype is older than other haplotypes detected in this species. Indeed, phylogeographic studies revealed that more common haplotypes might have had a longer time to disperse through the distribution area compared to more geographically restricted haplotypes that may be derived (Crandall and Templeton, 1993). Moreover, sharing haplotypes between/among populations may entail that *A. lignicola* had a widespread natural distribution in the region, as proposed by the distribution data of the species.

In *A. lignicola* average nucleotide diversity was 0.8%. In other gall wasp species genetic diversity shows wide variation. In *A. coriarius* nucleotide diversity was 0.5% in Iranian populations and 0.6% in Lebanese populations; however, in Turkey as it was placed in the main clade, nucleotide diversity was 1.5% (Challis et al., 2007). RFLP-based haplotype and nucleotide diversity were 0.4631 and 0.3204 for *A. caputmedusae* (Mutun, 2010), 0.8089 and 0.115542 for *A. lucidus* (Mutun, 2011), and 0.45 and 0.054 for *A. quercustozae* (Dinç and Mutun, 2011). Overall assessment of the findings confirms high genetic diversity in the Anatolian populations of *A. lignicola* that is higher than in studied European populations of oak gall wasp species, and well within the range of other gall wasp taxa examined so far from Turkey. On the other hand, with respect to haplotype diversity, the Kütahya population had the highest variation followed by the Uşak population (Table 3). Likewise, the highest nucleotide diversity was estimated for the Uşak population, followed by the Konya and Kütahya populations. However, in six populations (Afyon, Antalya, Denizli, Düzce, Eskişehir, and Kırıkkale) both haplotype and nucleotide diversity could not be observed (both values are 0) due to the detection of a single haplotype. Possible explanations for the low variation may be related to insufficient or different sampling sizes among localities. Haplotype richness is thought to be correlated with sampling size (Kalinowski, 2004); however, in our case there was no significant correlation between number of individuals sampled per population. This was further supported by the Chao-1 estimators in which only five populations displayed positive values. Alternately, species-specific demographic factors might have influenced these *A. lignicola* populations. However, observing only a single haplotype with balanced sampling across the distribution range may well be correlated with parasitoid attacks (Hayward and Stone, 2006), which are common in *A. lignicola* populations and it may be possible that some of the lineages have been swapped out of these localities. Galls of oak gall wasps can be parasitized by inquilines and parasitoids, leading some of the individuals to fail to develop inside the gall (Bailey et al., 2009). *Synophrus politus*, for example, infects galls of some cynipid species (Washburn and Cornell, 1981). Infected galls are quite similar in coloration pattern and other phenotypic features when compared with the noninfected galls of *A. lignicola*. Similar effects have been reported for *Andricus burgundus* gall wasps (Pujade-Villar et al., 2001); however, there is no report for *A. lignicola* infected by *S. politus* (except personal observations). In addition to inquiline and parasitoid attacks, there have been several reports of certain alphaproteobacteria, such as *Wolbachia* spp., and fungi infecting and causing high mortality in gall wasps during the developmental processes of larvae (Rokas et al., 2002). Parasitic attacks may even skew the sex ratio in oak gall wasps (Atkinson et al., 2003).

Pairwise comparisons among eighteen A. lignicola haplotypes revealed that the highest number of base differences in mere counting was between haplotype 5 (Çanakkale population) and haplotype 18 (Uşak population) with 46 nucleotide differences. Morphologically indistinguishable species, or cryptic species, may lie within taxonomically defined species. One way to detect cryptic species is to use DNA barcoding; a higher level of sequence difference (barcoding gap) is observed between species and a lower level of genetic distance is observed within species (Leasi and Norenburg, 2014). Cryptic species may be more common than was once thought (Williams et al., 2012); a new cryptic oak gall wasp species from Turkey and Iran has been described recently that was previously classified under *A. coriarius* (Challis et al., 2007). The presence of a high sequence difference in our case may imply a cryptic species complex. Our preliminary ABGD analysis provided supporting results with an emphasis on the presence of four hidden lineages within *A. lignicola*. Since speciation is not always accompanied by morphological changes—in our case both gall wasp characters and overall gall characteristics do not show distinguishable differences—large genetic distances within this traditionally recognized species might be handled carefully and more deeply. Therefore, further research is necessary to identify the presence of a cryptic species complex within *A. lignicola*. Our larger-scale studies to test this hypothesis are ongoing.

4.2. Phylogenetic and phylogeographic structure of *Andricus lignicola*

Phylogenetic analysis of the *A. lignicola* haplotypes using both MP and ML analyses produced the same tree topology. In the MP and ML phylogenetic tree the formation of two major clades is obvious, and it was significantly supported with high bootstrap value (Figure 1). Evolutionary relationships

were not clearly resolved in small subclades in both clades A and B; however, in clade B in the large polytomous part, in addition to a commonly shared haplotype (H1), all other haplotypes are geographically restricted to the western populations. In spite of a monophyletic grouping of two Balıkesir haplotypes (H3 and H4), relationships of all other haplotypes showed polytomy possibly due to insufficient time since divergence between lineages or incomplete lineage sorting (Avise, 2000). Some of the haplotypes represent westerly populations such that H14, H17, H18, and H15 (from Konya, Kütahya, and Uşak) seem to be well-separated from other haplotypes. Unlike MP and ML trees, BI tree is composed of more polytomic groups at the basal part. H1 (common haplotype), H8 (from İstanbul), and H16 (Kütahya and Manisa) all are polytomous, and another polytomous lineage gives rise to the main clade structure. Some differences are observed between the trees of both MP-ML and BI. This may be due to the use of a single type of genomic data, and further additions should address the incongruity observed between analysis methods, because under certain conditions phylogenetic and phylogeographic inferences can be misleading by introgression and retention of the ancestral polymorphism seen in the taxon (Kyriazi et al., 2008).

In conjunction with previous phylogenetic analyses of *A. lignicola*, further work covering a network analysis indicated the presence of three major groupings with an additional single haplotype that could not be connected within the 95% confidence limit. One advantage of applying a network analysis is that hybridization and any other type of reticulation can be revealed by nested clade and network analysis (Templeton, 1998; Bandelt et al., 1999; Posada et al., 2006); as observed in the network result, an apparent reticulation event may further support the presence of a cryptic species and incomplete sorting of the haplotypes.

F-statistics (AMOVA) utilized for hierarchical partitioning of variance into researcher-predefined groups indicated four groups with the highest and statistically significant values (Table 6b). Of the aforementioned grouping in Table 6b, ~74% of variation was present among groups and ~10% within populations of each group. The highly significant differentiation detected among groups is quite interesting and provides further support for the presence of high differentiation among some sampled localities, which may indicate a physical barrier to gene flow or a cryptic species. Similar cases were reported for other species (Challis et al., 2007; Williams et al., 2012). Indeed, pairwise comparisons of populations through F_{st} values demonstrated that several populations were differentiated from each other with the uppermost value. In particular, a complete differentiation ($F_{st} = 1$) between Kırıkkale and Antalya, Denizli, Afyon, Düzce, and Eskişehir is conspicuous (Table 5). Moreover, Antalya is highly differentiated from Afyon, Denizli, Düzce, and Eskişehir. Because high F_{st} values indicate little or no migration between populations due to separation of the populations by large geographic distances (Nguyen et al., 2006), this can be applied to some of the *A. lignicola* populations. However, the lack of differentiation ($F_{st} = 0$) observed among Afyon, Denizli, Düzce, and Eskişehir is also quite noticeable. This is because these populations share the same haplotype. In this study, paradoxical results regarding differentiation between populations may be partially explained by swapped out lineages, resulting in some *A. lignicola* populations having a single type of haplotype. Such heavy attacks by parasitoid or other predators have been reported from gall wasp species *A. kollari* and *A. corruptrix* (Walker et al., 2002).

Further research is necessary in *A. lignicola* with more sampling from Turkey and the nearby area to detect whether a cryptic species complex is present within the distribution of the species. Furthermore, additional molecular markers can be used to examine the species in much greater detail. However, we think that the current study will shed some light on oak gall wasps in Turkey.

Acknowledgment

This work was financially supported by a grant from Abant İzzet Baysal University Scientific Research Project Unit (BAP 2011.03.01.463).

References

Arias CM, Brito MR, Francisco OF, Moretto G, De Oliveira FF, Silvestre D, Sheppard S (2006). Molecular markers as a tool for population and evolutionary studies of stingless bees. Apidologie 37: 259–274.

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

Avise JC (2000). Phylogeography: The History and Formation of Species. Cambridge, MA, USA: Harvard University Press.

Bailey R, Schönrogge K, Cook JM, Melika G, Csoka G, Thuroczy C, Stone GN (2009). Host niches and defensive extended phenotypes structure parasitoid wasp communities. PLoS Biology 7: 1–12.

Bandelt HJ, Forster P, Rohl A (1999). Median-joining for inferring intraspecific phylogenies. Mol Biol Evol 16: 37–48.

Brown WM, George MJR, Wilson AC (1979). Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA. 76: 1967–1971.

Challis RJ, Mutun S, Nieves-Aldrey J-L, Preuss S, Rokas A, Aebi A, Sadeghi E, Tavakoli M, Stone GN (2007). Longitudinal range expansion and cryptic eastern species in the western Palaearctic oak gallwasp *Andricus coriarius*. Mol Ecol 16: 2103–2114.

Chao A (1984). Nonparametric estimation of the number of classes in a population. Scand J Statist 11: 265–270.

Çıplak B (2008). The analogy between interglacial and global warming for the glacial relicts in a refugium: a biogeographic perspective for conservation of Anatolian Orthoptera. In: Fattorini S, editor. Insect Ecology and Conservation. Kerela, India: Research Sign Post, pp. 135–163.

Clement M, Posada D, Crandall KA (2000). TCS: a computer program to estimate gene genealogies. Mol Ecol 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: 18.

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

Demirsoy A (2008). Genel Zoocoğrafya ve Türkiye Zoocoğrafyası 'Hayvan Coğrafyası.' 7th ed. Ankara, Turkey: Meteksan (in Turkish).

Dinç S, Mutun S (2011). PCR-RFLP variation of the oak gallwasp, *Andricus quercustozae* (Bosc, 1792) (Hymenoptera: Cynipidae) from Turkey. Turk J Entom 35: 47–58.

Excoffier L, Laval G, Schneider S (2005). Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47–50.

Fu Y-X, Li W-H (1993). Statistical tests of neutrality of mutations. Genetics 133: 693–709.

Gebiola M, Lopez-Vaamonde C, Nappo AG, Bernardo U (2014). Did the parasitoid *Pnigalio mediterraneus* (Hymenoptera: Eulophidae) track the invasion of the horse chestnut leafminer? Biol Invasions 16: 843–857.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41: 95–98.

Hasegawa M, Kishino H, Yano T (1985). Dating of human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22: 160–174.

Hayward A, Stone GN (2006). Comparative phylogeography across two trophic levels: the oak gall wasp *Andricus kollari* and its chalcid parasitoid *Megastigmus stigmatizans*. Mol Ecol 15: 479–489.

Hewitt GM (2004). The structure of biodiversity – insights from molecular phylogeography. Front Zool 1: 1–16.

Hickerson MJ, Carstens BC, Cavender-Bares J, Crandall KA, Graham CH, Johnson JB, Rissler L, Victoriano PF, Yoder AD (2010). Phylogeography's past, present, and future: 10 years after Avise, 2000. Mol Phylogenet Evol 54: 291–301.

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

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

Kalinoswski ST (2004). Counting alleles with rarefaction: private alleles and hierarchical sampling designs. Conserv Genet 5: 539–543.

Kekeçoğlu M, Soysal İM (2010). Genetic diversity of bee ecotypes in Turkey and evidence for geographical differences. Rom Biotech Lett 15: 5646–5653.

Kyriazi P, Poulakakis N, Parmakelis A, Crochet PA, Moravec J, Rastegar-Pouyani N, Trsigenopoulos CS, Magoulas A, Mylonas M, Lymberakis P (2008). Mitochondrial DNA reveals the genealogical history of the snake-eyed lizards (*Ophisops elegans* and *O. occidentalis*) (Sauria: Lacertidae). Mol Phylogenet Evol 49: 795–805.

Leasi F, Norenburg JL (2014). The necessity of DNA taxonomy to reveal cryptic diversity and spatial distribution of meiofauna, with a focus on Nemertea. PLoS ONE 9: e104385.

Librado P, Rozas J (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25: 1451–1452.

Loxdale HD, Lushai G (1998). Molecular markers in entomology. B Entomol Res 88: 577–600.

Melika G (2006). Gall wasps of Ukraine: Cynipidae. Kyiv, Ukraine: Vestnik Zoologii.

Mutun S (2010). Intraspecific genetic variation and the phylogeography of the oak gallwasp *Andricus caputmedusae* (Hymenoptera: Cynipidae): effects of the Anatolian Diagonal. Acta Zool Acad Sci H 56: 153–172.

Mutun S (2011). Intraspecific genetic diversity and phylogeography of the oak gall wasp *Andricus lucidus* (Hymenoptera: Cynipidae) populations in Anatolia. Turk J Zool 35: 559–570.

Nguyen TTT, Ingram B, Sungan S, Gooley G, Sim SY, Tinggi D, De Silva SS (2006). Mitochondrial DNA diversity of brood stock of two indigenous mahseer species, *Tor tambroides* and *Tor dourenensis* (Cyprinidae) cultured in Sarawak. Aquaculture 253: 259–269.

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

Posada D, Crandall KA (2001). Intraspecific gene genealogies: trees grafting into networks. Trends Ecol Evol 16: 37–45.

Posada D, Crandall KA, Templeton AR (2006). Nested clade analysis statistics. Mol Ecol 6: 590–593.

Puillandre N, Lambert A, Brouillet S, Achaz G (2012). ABGD, automatic barcode gap discovery for primary species delimitation. Mol Ecol 21: 1864–1877.

Pujade-Villar J, Bellido D, Segú G, Melika G (2001). Current state of knowledge of heterogeny in Cynipidae (Hymenoptera, Cynipoidea). Sessió Conjunta d'Entomologia de l'Institució Catalana d'História Natural i la Societat Catalana de Lepidopterologia, 11: 97–107.

Rokas A, Atkinson JR, Nieves-Aldrey J-L, West A, Stone GN (2002). The incidence and diversity of *Wolbachia* in gallwasps (Hymenoptera; Cynipidae) on oak. Mol Ecol 11: 1815–1829.

Rokas A, Atkinson RJ, Webster LMI, Csöka G, Stone GN (2003a). Out of Anatolia: longitudinal gradients in genetic diversity support an eastern origin for a circum-Mediterranean oak gallwasp *Andricus quercustozae*. Mol Ecol 12: 2153–2174.

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

Stone GN, Atkinson RJ, Brown G, Rokas A (2002). The population genetic consequences of range expansion: a review of pattern and process, and the value of oak gallwasps as a model system. Biodiversity Science 10: 80–97.

Stone GN, Challis RJ, Atkinson RJ, Csöka G, Hayward A, Melika G, Mutun S, Preuss S, Rokas A, Sadeghi E et al. (2007). The phylogeographical clade trade: tracing the impact of human-mediated dispersal on the colonization of northern Europe by the oak gall wasp *Andricus kollari*. Mol Ecol 16: 2768–2781.

Swofford DL (2002). PAUP*. Phylogenetic Analysis Using Parsimony. Sunderland, MA, USA: Sinauer Associates.

Tajima F (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.

Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596–1599.

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

Walker P, Leather SR, Crawley MJ (2002). Differential rates of invasion in three related alien oak gall wasps (Cynipidae: Hymenoptera). Divers Distrib 8: 335–349.

Washburn JO, Cornell HV (1981). Parasitoids, patches, and phenology: their possible role in the local extinction of a cynipid gall wasp population. Ecology 62: 1597–1602.

Williams PH, An J, Brown MJF, Carolan JC, Goulson D, Jiaxing H, Masao I (2012). Cryptic bumblebee species: consequences for conservation and the trade in greenhouse pollinators. PLoS ONE 7: e32992.

Zhang XD, Hewitt MG (1996). Assessment of the universality and utility of a set of conserved mitochondrial COI primers in insects. Insect Mol Biol 6: 143–150.