Mitochondrial barcodes are diagnostic of shared refugia but not species in hybridizing oak gallwasps

JAMES A. NICHOLLS,* RICHARD J. CHALLIS,†‡ SERAP MUTUN† and GRAHAM N. STONE*

*Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK, †Department of Biology, Abant Izzet Baysal University, 14280 Bolu, Turkey

Abstract

Mitochondrial DNA barcodes provide a simple taxonomic tool for systematic and ecological research, with particular benefit for poorly studied or species-rich taxa. Barcoding assumes genetic diversity follows species boundaries; however, many processes disrupt species-level monophyly of barcodes leading to incorrect classifications. Spatial population structure, particularly when shared across closely related and potentially hybridizing taxa, can invalidate barcoding approaches yet few data exist to examine its impacts. We test how shared population structure across hybridizing species impacts upon mitochondrial barcodes by sequencing the cytochrome b gene for 518 individuals of four well-delimited Western Palaearctic gallwasp species within the Andricus quercuscalicis species group. Mitochondrial barcodes clustered individuals into mixed-species clades corresponding to refugia, with no difference in within- and between-species divergence. Four nuclear genes were also sequenced from 4 to 11 individuals per refugial population of each species. Multi-locus analyses of these data supported established species, with no support for the refugial clustering across species seen in mitochondrial barcodes. This pattern is consistent with mitochondrial introgression among populations of species sharing the same glacial refugium, such that mitochondrial barcodes identify a shared history of population structure rather than species. Many taxa show phylogeographic structure across glacial refugia, suggesting that mitochondrial barcoding may fail when applied to other sets of co-distributed, closely related species. Robust barcoding approaches must sample extensively across population structure to disentangle spatial from species-level variation. Methods incorporating multiple unlinked loci are also essential to accommodate coalescent variation among genes and provide power to resolve recently diverged species.

Keywords: DNA barcoding, introgression, multi-locus, spatial population structure

Received 06 February 2012; revision received 15 May 2012; accepted 23 May 2012.

Introduction

Modern molecular methods have added a suite of characters to assist with the delimitation and identification of species. Foremost among these are DNA barcodes - short DNA sequences derived from a standardized section of the genome, in animals part of the mitochondrial cytochrome c oxidase I gene (COI) (Hebert et al. 2003b). Barcoding is particularly useful for determining species status and limits in animal groups that are poorly studied or where traditional morphological characters are scarce (Smith et al. 2007; Nicholls et al. 2010b). The improved taxonomy derived through barcoding can have broader benefits for the understanding of ecological processes (Janzen et al. 2005; Kaartinen et al. 2010) or disease control (Bickford et al. 2007) as well as having direct commercial or quarantine applications (Armstrong & Ball 2005).

Species delimitation using DNA barcodes assumes that patterns of genetic diversity track species boundaries;
sequences from different species form reciprocally monophyletic groups and intra-species variation is less than inter-specific variation, providing the ‘barcode gap’ (for example Ács et al. 2010). Boundaries between species reflect divergence thresholds (absolute values such as 2% divergence (Hebert et al. 2003b) or a relative value scaled by intra-species diversity—the 10× rule (Hebert et al. 2004)) or transition thresholds from coalescent processes within species to birth-death processes among species (Pons et al. 2006; Papadopoulou et al. 2008). However, conventional DNA barcoding has a number of problems. The ubiquity of the barcode gap, required for any divergence threshold, has been contested both on theoretical (Hickerson et al. 2006) and empirical grounds (Moritz & Cicero 2004; Meier et al. 2006; Wiemers & Fiedler 2007). Processes including hybridization and incomplete sorting of ancestral variation can disrupt reciprocal monophyly among species, especially closely related species (for example Gompert et al. 2006; Haider et al. 2012); the resulting para- or polyphyly is estimated to affect almost a quarter of animal taxa (Funk & Omland 2003). Standard mitochondrial barcoding practice also disregards coalescent variation among genes by utilizing only a single gene, despite evidence for conflict between gene trees and the true species tree (Maddison 1997; Hudson & Turelli 2003; Edwards 2009). Endosymbiotic-associated mitochondrial selection (Ballard & Whitlock 2004; Hurst & Jiggins 2005; Whitworth et al. 2007) and the relative ease with which maternal genomes cross species boundaries (Chan & Levin 2005) can exacerbate this conflict in animal barcoding. Therefore, as with evolutionary and phylogeographic studies (Edwards 2009), robust molecular barcoding practices need to assess patterns across multiple unlinked loci. Not only does this provide additional power to delimit species (Monaghan et al. 2005; Elias et al. 2007; Yang & Rannala 2010), it also reveals mechanisms such as introgression or incomplete sorting that can lead to incorrect barcode classifications.

Spatial population structure can introduce further complications for barcoding. It is widely recognized that COI sequence variation can define phylogeographic groups within species (Hebert et al. 2003a), and refugial genetic structure is widespread in temperate (Taberlet et al. 1998; Hewitt 2000) and tropical ecosystems (Caraval et al. 2009). Theoretical work has shown that with limited spatial sampling, as is typical for most barcode data sets (Funk & Omland 2003; Zhang et al. 2010), within-species phylogeographic structure can mimic the pattern of low intra-cluster and high inter-cluster variation expected for an ideal species-delimiting barcode data set (Lohse 2009). This in itself would produce overly strict ‘species’ classifications under either divergence or process transition thresholds. However, many species often share refugia and hence have similar within-species genetic structure (Taberlet et al. 1998; Hewitt 2000). Even rare hybridization is enough to allow extensive introgression across species boundaries (Chan & Levin 2005), so should it occur among co-refugial populations of related species, then barcode data could reflect refugia rather than species. Barcoding can be misleading when the sampling across taxa, in particular sister species, is incomplete (Meyer & Paulay 2005); similarly it may also be misleading when sampling within species is not geographically complete (Trevick 2008; Zhang et al. 2010). Despite this concern, few studies have tested the robustness of molecular barcodes over spatial scales that encompass phylogeographic structure. Hebert et al. (2010) concluded that spatially widespread sampling had very limited impact on the efficacy of barcoding. However, their test was conducted using latitudinally separated individuals within eastern North America, a region within which many taxa show evidence for recent northwards range expansion from a single refugium resulting in little or no latitudinal genetic structure (Hewitt 2000, 2004). This reduces the generality of conclusions that can be drawn from Hebert et al.’s (2010) data, and leaves the issue of population structure in regions with more complex phylogeographic histories open to debate.

Here we test the efficacy of mitochondrial barcoding to delimit species showing shared phylogeographic structure using multi-locus data from a group of oak gallwasp (Hymenoptera: Cynipidae). We focus on four closely related species within the quercuscalicis species group, a well-supported clade within the genus Andricus (Cook et al. 2002; Stone et al. 2009). These taxa are reliably distinguished using adult morphology (Melika 2006) and the structure of the galls they induce (Stone et al. 2002), allowing strong a priori assignment of individuals among the four focal species. Phylogeographic structure within the Western Palaearctic has previously been shown in one of the focal species (Rokas et al. 2003a), with three major refugial populations found in the Iberian peninsula (including north-western Africa and far south-western France), the remainder of southern Europe (incorporating the Italian and Balkan peninsulas), and Asia Minor/north-western Iran. This phylogeographic structure is shared by many other species within the oak gall community, both gallwasp and their parasitoids (Rokas et al. 2001; Challis et al. 2007; Stone et al. 2007, 2012; Lohse et al. 2010; Nicholls et al. 2010a,b), and parallels refugial patterns found in many Western Palaearctic taxa (Taberlet et al. 1998; Hewitt 2000). Previous work on gallwasp involving only limited spatial sampling has shown that on occasion mitochondrial barcodes do not correspond to species limits owing to hybridization or incomplete sorting of
ancestral polymorphism (Rokas et al. 2003b); in addition, gallwaspse have been shown experimentally to hybridize (Folliot 1964). We use thorough sampling of closely related species across their shared refugial distributions to examine in greater detail how the interaction of hybridization, incomplete sorting and refugial population structure can disrupt the monophyly of mitochondrial species barcodes.

Specifically, we use multi-species, multi-regional data to test whether mitochondrial barcode clusters correspond to species limits, and assess concordance between mitochondrial and nuclear sequence data. If hybridization within refugia is a frequent occurrence, we expect mitochondrial barcodes to reconstruct mixed-species refugial groupings. Recent hybridization would result in haplotypes being shared among species; older hybridization within refugia followed by within-species diversification would result in species-related structure nested within refugial clades. However, in both cases barcodes from a single species will not be monophyletic. An alternative source of nonmonophyly in species barcodes is incomplete sorting of ancestral polymorphism among daughter species. In contrast to hybridization, this process predicts no consistent pattern of haplotype clustering with respect to geography. Assuming barcode sequence variation is neutral, we would not expect the same haplotypes to be retained in sympatric populations of multiple species during their independent colonizations of refugia. We also use nuclear sequence data from four genes to test the underlying assumption that gall morphology, the major character used to identify species, is a true indicator of species status; if so, nuclear sequences sampled across refugia should identify taxa concordant with recognized morphospecies. Examination of spatial patterning in nuclear alleles across species allows discrimination between nuclear introgression and polymorphism sorting processes. The comparison of signals in mitochondrial and nuclear genomes also provides insight into the processes influencing mitochondrial history in this species group and hence the utility of standard mitochondrial barcodes.

Methods

Mitochondrial barcode sampling

We obtained mitochondrial sequence data for 518 individuals from the four species A. quercuscalicis (69 individuals), A. dentimitratus (37 individuals), A. caputmedusae (198 individuals) and A. quercustozae (214 individuals) for a 433 base pair (bp) fragment of the cytochrome b gene (cyt b), using procedures described in Stone et al. (2009) (see Table S1, Supporting information). Samples were obtained from all the major refugial regions that each species occurs in (Asia Minor/Iran and central Europe for A. caputmedusae, Asia Minor/Iran, central Europe and Iberia for both A. dentimitratus and A. quercustozae, and central Europe for A. quercuscalicis; Table S2, Supporting information). Some northern European samples were also included in the central European sample for A. quercuscalicis as this species has recently invaded northern Europe from the Balkans (Stone & Sunnucks 1993). The cyt b gene (which shows very similar levels of sequence diversity to COI in gallwasp, Stone et al. 2012) was used to allow inclusion of existing gallwasp data sets (Rokas et al. 2003a,b); however, as the mitochondrion is a single linkage group the patterns obtained will be representative of standard COI barcodes. Seven outgroup sequences of oak gallwasp were also included (GenBank accessions AJ228454, AJ228464, AJ228481, AF481706, DQ286803, EF031412, EU552436).

Mitochondrial barcode analyses

Phylogenetic relationships among the 285 unique cyt b haplotypes were first assessed using a neighbour-joining analysis based on K2P distances in the software PAUP* v4.0b10 (Swofford 1998), with branch support assessed using 1000 bootstrap replicates. Species memberships and the geographic sources of individuals were mapped on to the resulting haplotype tree. Although phenetic clustering methods have been criticized (Will & Rubinoff 2004), this simple analysis method is typical of barcode studies (Hebert et al. 2003a, 2004).

As a comparison, we also conducted a more sophisticated Bayesian analysis that allowed for independent models of sequence evolution at each codon position and phylogenetic uncertainty. We first used a subset of 50 quercuscalicis-group haplotypes plus outgroups to determine the substitution model at each codon position and the applicability of a molecular clock following Nicholls et al. (2010b). The best substitution model was GTR + I + G for both first and third codon positions, GTR + I for second positions, with a clock supported. This model was implemented in the software BEAST v1.5.4 (Drummond & Rambaut 2007) to determine relationships within the full haplotype set. Alternative coalescent and clock models were evaluated with the best model utilizing a relaxed clock and an exponential growth coalescent model. Analyses were run twice, each time for 40 million generations, sampled every 5000 generations with parameters estimated using the final 5 million generations. Evidence against monophyly of sequences from each species was assessed by comparing the unconstrained analysis with analyses in
which one of the four species was constrained to be monophyletic. Models were compared using Bayes factors, estimated as twice the difference in the log of the harmonic mean of model likelihoods ($2\Delta\ln\text{HML}$), following cut-off values in Kass & Raftery (1995).

As an alternative to tree-based delimitation of sequence clusters, we used the software jMOTU v1.0.6 (Jones et al. 2011) to cluster sequences into groups that differed by a pre-defined number of base pairs; this analysis was repeated for cut-off values in the range 1–11 bp. If a barcoding gap exists, the number of groups should remain constant over the range of cut-off values corresponding to distances greater than intraspec
taxon diversity but less than inter-taxon diversity (Meyer & Paulay 2005; Ács et al. 2010). We then assessed whether groups containing more than one individual were sampled from either multiple species or multiple refugial regions. In an ideal barcoding data set multi-species groups should be absent (up to the cut-off point beyond the barcoding gap at which species are clustered into a common group); similarly a data set showing a clear pattern of refugial structure should contain no multi-regional groups. Some of our samples came from locations in potential phylogeographic suture zones (regions such as NE Spain/SW France or the areas adjacent to the Dardanelles/Bosphorus at the junction of Europe and Asia Minor in NW Turkey) in which populations could be derived through range expansion from multiple refugial sources. To assess whether the barcode data contained a strong signal of spatial structure among core refugial regions we repeated the jMOTU analyses excluding these samples.

Nuclear sequence generation and analysis

For each species we obtained nuclear sequence data for between four and 11 individuals from each of the major geographic regions in which each species occurred (Table S2, Supporting information). Four nuclear loci were screened: ITS2 (Campbell et al. 1993) and intron regions within the long-wavelength opsin gene (Stone et al. 2009) and ribosomal proteins RpL15 and RpL37 (Lohse et al. 2010). All individuals were females and hence diploid, so each contributed two alleles per locus to the analyses. Although sample sizes were typically smaller than those used for mitochondrial screening (between eight and 62 gene copies per locus per species), in the multi-locus analyses used here the power to estimate the species (or population) tree comes primarily from the number of included genes, not the number of terminals per species (Heled & Drummond 2010; Yang & Rannala 2010).

The ability of these nuclear data to discriminate species was assessed with a multi-locus species delimitation analysis using the software BPP v2.1 (Yang & Rannala 2010). This methodology assesses support for species-level divergence on a phylogeny while accounting for gene tree uncertainty and incorporating lineage sorting through a multi-species coalescent model. The first step in this analysis is the production of a guide tree, which was generated using the *BEAST algorithm (Heled & Drummond 2010) implemented within BEAST v1.5.4. The *BEAST algorithm estimates substitution models, tree topologies and molecular clock rates for each gene, while using a multi-species coalescent model to nest the gene trees within the concurrently estimated shared species tree topology (Heled & Drummond 2010). Substitution models were selected as mentioned previously, with the final analysis using independent GTR + I models, strict clocks for ITS2, opsin and RpL15 and a relaxed lognormal clock for RpL37. The species tree was estimated following a Yule process and constant population size model and was rooted using sequences from the gallwasp Dryocosmus kuriphilus. Analyses were run for 200 million generations, sampled every 25 000 generations, with parameters estimated over the final 50 million generations. Each different model was run twice to confirm convergence. Individuals were a priori assigned to species based upon their gall morphology. This guide tree was then entered into BPP, with equal prior probability given to each alternative rooted species tree compatible with it. The two alternative rjMCMC algorithms and different fine-tune parameters gave the same results; final analyses were conducted using algorithm 0 with fine-tuning parameter $\alpha = 20$. Rates were allowed to vary among loci following the Dirichlet distribution with $\alpha = 20$. Three different combinations of priors were used for population sizes and ancestral root age, although they had no effect on species delimitation: priors of $G(1,40)$ for the $\theta$s and $G(1,800)$ for $\tau_0$ (corresponding to large population sizes and an older root); priors of $G(1,800)$ for the $\theta$s and $G(1,1600)$ for $\tau_0$ (corresponding to small populations and recent divergence); and priors of $G(1,40)$ for the $\theta$s and $G(1,1600)$ for $\tau_0$ (reflecting biologically more sensible expectations of large populations and recent divergence). Each model was run twice from a different starting seed and initial starting tree topology to confirm consistency between runs. All analyses were run for 500 000 generations (sampling interval = 5) with a burn-in of 20 000.

Alignments from all four nuclear genes were included in a second *BEAST analysis to assess the relationships of refugial populations among species. The strong geographic structure within the mitochondrial data implies little gene flow among regions, so in this analysis individuals were a priori assigned to one of the three major regional populations within each
species; hence, our units within the ‘species’ tree were populations within species rather than species. All *BEAST* model settings were otherwise the same as for the species-level analysis. This analysis tested the prediction that genetic variation at nuclear loci tracks gall structure and species status, with the expectation that sequences and populations should cluster by species, not by geographic region.

Results

Mitochondrial barcodes

Bayesian model-based analysis of mitochondrial haplotypes clustered them into clades that showed poor correspondence with recognized morpho-species (Fig. 1a), limiting their use as barcode markers of species status. Haplotypes from *A. quercustozae* clustered into discrete well-supported clades that were not sister groups. *Andricus caputmedusae* and *A. dentimitratus* haplotypes were intermixed in multiple nonsister clades with occasional haplotypes in clades otherwise dominated by *A. quercuscalicis* or *A. quercustozae*. Although *A. quercuscalicis* haplotypes formed a well-supported clade, this clade still contained individuals of another species. Monophyly for each species was strongly rejected by *BEAST* model tests (*A. caputmedusae* \(2\Delta\ln HML = 310.70\); *A. dentimitratus* \(2\Delta\ln HML = 359.31\); *A. quercuscalicis* \(2\Delta\ln HML = 33.35\); *A. quercustozae* \(2\Delta\ln HML = 102.95\)).

In contrast, these mixed-species clades clustered individuals from the same major refugial region (Fig. 1b), consistent with extensive within-refugial introgression. Three clades contained eastern samples: all Iranian and most Turkish samples clustered into one large clade (comprising subclades 9 & 10 in Fig. 1b), while smaller numbers of haplotypes from western and central Turkey grouped into two other clades (clades 1 & 4). There were multiple clades sampled exclusively, or almost so, from Europe to the exclusion of Iberia (clades 2, 3, 6–8), and a single clade contained nearly all Iberian samples (clade 5). Some clades contained samples from two major regions (clades 1–4, 6, 7), although in these cases the majority of samples came from one region and the remainder came from immediately adjoining parts of the neighboring region (samples from the far northwest of Asia Minor in the European clades 3 & 7, samples from the European section of Turkey in clades 1 & 4 that were otherwise found only in Asia Minor, or samples from north-eastern Spain or the far south-west of France in otherwise European clades 2 & 6). The standard barcoding practice of neighbour-joining analysis using a Kimura 2-parameter genetic distance produced virtually identical results (Fig. S1, Supporting information).

There was no barcoding gap in the mitochondrial data, as the number of sequence groupings showed a constant decline with cut-off distance (Fig. 2); by 11bp divergence all sequences were contained within a single group. The absence of a gap was also highlighted by similarity in intra- and inter-species sequence divergence in the clade as a whole (Table 1). Mitochondrial barcodes failed to resolve discrete species at any level; at least 16% of multi-individual groupings at sequence differences of 1–11 bp contained multiple species (Fig. 3). Small numbers of haplotypes/clades in the full data set were shared over refugial regions (Fig. 3a). However, when only individuals from core refugial areas were used (excluding samples from areas on the periphery of refugia), groupings up to 3bp divergence became diagnostic of regions (i.e. no multi-individual groups were sampled from multiple regions; Fig. 3b), although with no corresponding improvement in their ability to diagnose species.

Multi-locus nuclear analyses

The four nuclear markers typically showed divergence levels an order of magnitude smaller than the mitochondrial data with some species sharing a few alleles at each locus, although allelic diversity was comparable (Table S3, Supporting information). These data showed a contrasting pattern to the mitochondrial barcodes that correlated well with morpho-species definitions, supporting our assignment of individuals to species. The multi-locus species delimitation analysis provided very strong support for the distinctiveness of the four morpho-species, with posterior probabilities of one for every speciation event in the species tree (Fig. 4a). At a population level, the three *A. quercustozae* refugial populations formed a monophyletic clade with posterior probability close to one (Fig. 4b), providing very good evidence for the distinctiveness of this species. Population samples from the other species did not group by taxonomy; however, in contrast to mitochondrial relationships there was no statistical support for clustering by geographic origin, consistent with incomplete sorting of ancestral polymorphism within these three closely related species. Similarly, there was no observed statistical support for a pattern of spatial clustering of alleles sampled from different species within any of the nuclear gene trees (Fig. S2, Supporting information).

Discussion

Our data clearly demonstrate that in this system mitochondrial barcodes do not correspond to species limits at any level of divergence. In contrast, the multi-locus nuclear data (although sampled from fewer individuals
BARCODES REFLECT REFUGIA NOT SPECIES

than the mitochondrial barcodes) do delimit our four species, indicating our prior understanding of species limits based on gall and adult morphology is correct. The nuclear data cluster refugial populations of one species, although the remaining species show no consistent pattern of clustering with respect to either species or geography, consistent with incomplete sorting of ancestral polymorphism among three recently diverged species. Against this nuclear background of discrete but only partially sorted species-level gene pools, the mitochondrial barcodes show a strong signal of mixed-species clades corresponding to geographic regions, as predicted by a process of introgression within shared refugia. This highlights two major issues that have broad implications for the general utility of DNA barcoding: appropriate spatial sampling and dependence on a single gene, particularly when that gene is of mitochondrial origin.

Sampling over spatial refugial structure is critical

The data presented here provide a clear empirical example of the pattern predicted by recent theory (Lohse 2009) that spatial population structure can mimic the process producing species-level clustering in barcode data, resulting in discrete regional clades. All geographic mismatches (i.e. haplotypes sampled from the ‘wrong’ refugial region) can be attributed to dispersal from a neighbouring refugium, as demonstrated by the lack of any multi-regional groupings at 1–3 bp cut-off values when individuals from the periphery of refugia were excluded. Previous work on the gallwasp system indicates that many of these regional lineages are relatively old, dating from the pre- or early Pleistocene, but still represent divergence within species as individuals derived from different refugia can interbreed (Rokas et al. 2003a; Challis et al. 2007; Stone et al. 2007, 2012; Nicholls et al. 2010a). By itself, intra-species diversification into regional clades is not a problem for barcoding if the regional clades are sister groups. However, mitochondrial introgression among species in sympathy breaks down this sister group relationship of intra-specific lineages, and could be expected whenever closely related species share refugia. By sampling over a spatial scale encompassing multiple refugial populations we reveal just how important shared population structure can be for disrupting species-level signal in barcodes. In contrast, sampling over the same spatial scale but from younger populations recently expanded from a single refugium, as done by Hebert et al. (2010) or exemplified by our A. quercuscalicis data (Stone & Sunnucks 1993), returns the result that barcodes cluster by species irrespective of geographic distance among samples. So although spatial population genetic structure is welcomed by phylogeographers, molecular barcoding practices must also account for this source of variation.

This highlights two important issues that extensive spatial sampling will circumvent. First, it will help resolve the status of outlying divergent sequences within a sample, to clarify whether they are regional variants or represent different species. This is of particular importance for samples taken from areas that have been colonized by multiple source populations. Accounting for such diversity is fundamental to accurate estimation of species limits within barcode data based on either distance metrics (Hebert et al. 2003b, 2004) or transitions between speciation and coalescent processes (Lohse 2009). Second, it highlights the need for extensive sampling to avoid confounding species with geographic structure (Zhang et al. 2010). As an extreme example, imagine our data were much more limited with A. quercustozae only sampled from Iberia, A. dentimitratus only from Hungary and A. caputmedusae only from the Middle East. Barcode data from new unidentified specimens would group with the respective refugial reference sample, potentially leading to spurious species assignments and thus defeating the purpose of a barcode-based taxonomy.

Mitochondrial barcodes can be misleading

As with any inference based upon a single gene, mitochondrial barcoding reconstructs the history of the
sampled gene, which may well be different from the evolutionary history of the sampled species or populations (Maddison 1997; Edwards 2009). As a result, mitochondrial barcodes can provide a misleading signal of species status. In the case of *quercuscalicis* -clade gallwasps the barcodes show strong spatial structuring rather than species-related clusters. This pattern is consistent with a process of occasional mitochondrial introgression among distinct species occurring in shared refugia, a pattern also observed in other taxa (Berthier et al. 2006; Schmidt & Sperling 2008). However, the lack of similar spatial structure in nuclear data suggests any hybridization is followed by extensive backcrossing to paternal parent species. This would lead to nuclear (and phenotypic) characters indicating species status despite the presence of a maternally inherited mitochondrion in the ‘wrong’ species. Although rare, within-refuge hybridization must be a continual process in the history of this group. Older hybridization events within a refugium result in within-species diversification of mitochondrial lineages that form sister groups (such as the split within the Asia Minor/Iran clade into subclades 9 & 10 containing only *A. quercustozae* and *A. caputmedusa*/dentimitratus samples, respectively), while more recent hybridization is evidenced by shared haplotypes among species.

This continual hybridization will have contributed to the lack of any barcoding gap. Mean intra-species divergence for mitochondrial DNA is of similar magnitude to inter-species divergence (see Table 1), and if we use the smallest inter-specific distance measure as the species delimiter rather than the mean (as suggested by Meier et al. (2008)), all samples should be considered the same species in contradiction to other available evidence. Hence any general rule involving a standardized percentage cut-off (for example 2%) or a relative distance such as the 10· rule of Hebert et al. (2004) will not work in this instance. In addition, the species delimitation approach of Pons et al. (2006) is unlikely to work for these data as the presence of introgression breaks down the distinctiveness of any transition from within-to between-species processes.

These data highlight the need for a multi-locus approach to molecular barcoding, in the same way that inference based upon multiple unlinked loci is now integral to phylogeography and population genetics (Edwards 2009). In this particular case the mismatch of mitochondrial barcodes with species is owing to mitochondrial introgression. More generally, potential for barcode/species mismatches owing to coalescent processes will be high where speciation is recent and effective population sizes are large, both of which apply to many invertebrate groups (Seehausen 2004; Elias et al. 2007; Edwards 2009). Gall morphology acts as an extended phenotype of gallwasp genes to highlight the

---

**Table 1** Intra- and inter-species Kimura 2-parameter distance measures for *quercuscalicis*-clade cytochrome *b* sequences

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Number of individuals</th>
<th>Number of haplotypes</th>
<th>Within species</th>
<th>Between species</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Mean</td>
<td>Range</td>
</tr>
<tr>
<td><em>A. caputmedusa</em></td>
<td>198</td>
<td>88</td>
<td>0.029</td>
<td>0.002–0.060</td>
</tr>
<tr>
<td><em>A. dentimitratus</em></td>
<td>37</td>
<td>23</td>
<td>0.032</td>
<td>0.002–0.058</td>
</tr>
<tr>
<td><em>A. quercuscalicis</em></td>
<td>69</td>
<td>49</td>
<td>0.016</td>
<td>0.002–0.041</td>
</tr>
<tr>
<td><em>A. quercustozae</em></td>
<td>214</td>
<td>134</td>
<td>0.041</td>
<td>0.002–0.087</td>
</tr>
<tr>
<td>Whole clade</td>
<td>518</td>
<td>285</td>
<td>0.040</td>
<td>0–0.087</td>
</tr>
</tbody>
</table>

---

Fig. 3 The percentage of multi-individual mitochondrial sequence groupings that incorporate multiple species or refugia. Black bars indicate multi-species groups, white bars indicate groups sampled from multiple major refugia. (a) results incorporating all data and (b) results using only samples from core refugial areas. Numbers above bars indicate the total number of multi-individual groups at each base pair cut-off.
barcoding errors in this study. However, the lack of useful phenotypic characters in many invertebrate taxa means these potentially widespread barcoding errors would go undetected in exactly the taxa for which molecular barcoding is most necessary as a taxonomic tool. The use of multiple independent markers for species delimitation will uncover any single gene conflicts with species boundaries, irrespective of the mechanism generating the conflict, given adequate sample sizes across all screened genes (for example Hailer et al. 2012). In addition, if they are sufficiently variable then increasing the number of sampled loci provides the power to discriminate recently diverged species pairs (Edwards 2009; Yang & Rannala 2010). Exactly how many loci might be needed for reliable discrimination will be system-dependent and can be explored both empirically and theoretically (Knowles & Carstens 2007; Edwards 2009; Knowles 2009). Hence, mitochondrial barcodes by themselves, without corroboration from morphological or nuclear genetic data, need to be treated with some caution.

In conclusion, these data are a clear example of how the complexity of biological systems sometimes cannot be accommodated within the framework of mitochondrial barcoding. While barcoding can provide a useful complement to existing identification methods, it is not known how frequently patterns such as those demonstrated here will be encountered, although it has been suggested that at least 10% of animal species hybridize (Mallet 2005). In gallwasps, the failure of mitochondrial barcodes to reconstruct a taxonomy of species is apparent as specimens can be reliably identified using the extended phenotype of their gall structure. In many other animal groups such obvious phenotypic characters are lacking, yet these groups encompass radiations of species that have occurred over the same temporal and spatial scales (Taberlet et al. 1998; Hewitt 2000, 2004) so the hybridization and spatial structuring processes leading to sharing of mitochondrial barcodes among species may not be isolated to gallwasps. Therefore, more robust practices such as the use of multiple unlinked loci, sampling across the entire distribution of species and thorough screening of sister taxa need to be implemented to demonstrate the efficacy of molecular barcoding, even within taxa where we think mitochondrial barcoding works.

Acknowledgements

We would like to thank Jenny Sturgeon and the Edinburgh Genepool Sequencing Facility for their assistance in generating sequence data, and numerous collaborators across the Western Palaearctic for providing samples. This work was supported by the Natural Environment Research Council (grant numbers NE/B504406/1 to G.N.S. and NE/E014453/1 to G.N.S. and J.A.N.).

References


Fig. 4 Species and population trees based upon combined analysis of four nuclear loci. (a) Species relationships among the four sampled quercuscalicis-clade species; numbers at nodes indicate posterior support for the corresponding speciation event. (b) Maximum clade credibility population tree of quercuscalicis-clade species sampled from three glacial refugia. Species names are coloured as in Fig. 1a; symbols indicate refugia; numbers above branches indicate posterior probability support.


J.A.N. designed the study, generated some mitochondrial and all nuclear data, analysed the data and led the writing of the paper. R.J.C. and G.N.S. initiated the study, collected much of the mitochondrial data and contributed to writing and analysis. S.M. provided a comprehensive sample set from Turkey.

Data accessibility
DNA sequences deposited in GenBank, accessions AF481704, AF539593, AF539594, AF539561, AF539562, AF539567, AJ228450, AJ228456, AJ228459, AJ228467, AY157269-AY157273, AY157275, AY157277-AY157298, EU552433, JQ228863-JQ229433. Table S1 relates these GenBank numbers to individuals as well as providing sampling information.
Supporting Information

Additional supporting information may be found in the online version of this article.

**Figure S1** Neighbour-joining phylogram of 285 *quercuscalicis*-clade cytochrome *b* haplotypes using Kimura 2-parameter genetic distances.

**Figure S2** Maximum clade credibility cladograms of four nuclear loci sampled from regional populations of *quercuscalicis*-clade species.

**Table S1** Collection locality information for the 518 *quercuscalicis*-clade gallwasp individuals sequenced for cytochrome *b*. Haplotype number corresponds to terminals in Figure 1 and Figure S1; sample names are used for terminals in Figure S2.

**Table S2** Number of individuals screened from the four focal *quercuscalicis*-clade gallwasp species from each of three major Western Palaearctic refugia for the loci cytochrome *b*, ITS2, opsin, RpL15 and RpL37 respectively.

**Table S3** Levels of variation in the five genes sequenced from the four focal species of *quercuscalicis*-clade gallwasps.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.