MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2013) doi: 10.1111/1755-0998.12158

Conflicting patterns of DNA barcoding and taxonomy in the cicada genus *Tettigettalna* from southern Europe (Hemiptera: Cicadidae)

VERA L. NUNES,* RAQUEL MENDES,* EDUARDO MARABUTO,* BRUNO M. NOVAIS,* THOMAS HERTACH,† JOSÉ A. QUARTAU,* SOFIA G. SEABRA,* OCTÁVIO S. PAULO* and PAULA C. SIMÕES*

**Computational Biology and Population Genomics Group, Centro de Biologia Ambiental, DBA/FCUL, Lisboa 1749-016, Portugal,*
†*Institute of Biogeography, University of Basel, St. Johanns-Vorstadt 10, Basel 4056, Switzerland*

Abstract

DNA barcodes have great potential to assist in species identification, especially when high taxonomical expertise is required. We investigated the utility of the 5′ mitochondrial cytochrome *c* oxidase I (COI) region to discriminate between 13 European cicada species. These included all nine species currently recognized under the genus *Tettigettalna*, from which seven are endemic to the southern Iberian Peninsula. These cicadas have species-specific male calling songs but are morphologically very similar. Mean COI divergence between congeners ranged from 0.4% to 10.6%, but this gene was proven insufficient to determine species limits within genus *Tettigettalna* because a barcoding gap was absent for several of its species, that is, the highest intraspecific distance exceeded the lowest interspecific distance. The genetic data conflicted with current taxonomic classification for *T. argentata* and *T. mariae*. Neighbour-joining and Bayesian analyses revealed that *T. argentata* is geographically structured (clades North and South) and might constitute a species complex together with *T. aneabi* and *T. mariae*. The latter diverges very little from the southern clade of *T. argentata* and shares with it its most common haplotype. *T. mariae* is often in sympatry with *T. argentata* but it remains unclear whether introgression or incomplete lineage sorting may be responsible for the sharing of haplotypes. *T. helianthemi* and *T. defauti* also show high intraspecific variation that might signal hidden cryptic diversity. These taxonomic conflicts must be re-evaluated with further studies using additional genes and extensive morphological and acoustic analyses.

Keywords: barcoding gap, cicadas, cytochrome *c* oxidase I, DNA barcoding, Southern Europe, *Tettigettalna*

Received 8 April 2013; revision received 16 July 2013; accepted 31 July 2013

Introduction

The DNA barcoding initiative was proposed as a fast and cost-effective system to identify the species of any specimen with the information provided by a short and standardized DNA fragment, avoiding the need for a high level of taxonomical expertise (Hebert *et al.* 2003a). DNA barcoding has thus a great potential for biodiversity assessment and ecological applications, such as the association of different life stages, the study of trophic relationships through molecular identification of gut content or the detection and control of disease vectors, invasive species and illegal trade of protected species (Valentini *et al.* 2008; Jinbo *et al.* 2011). In order to be effective, species-level assignments require a database with reference sequences representative of all known species (Virgilio *et al.* 2010). The mitochondrial gene cytochrome *c* oxidase I (COI) showed a high success rate in early barcode tests of diagnosing animal species (Hebert *et al.* 2003b, 2004a,b; Ward *et al.* 2005), and since then, the accumulation of reference sequences from animals has grown exponentially. However, the limited power of DNA barcoding to discriminate between closely related species when intra- and interspecific distances overlap has been criticized by several authors (Moritz & Cicero 2004; Meyer & Paulay 2005; Meier *et al.* 2006; Elias *et al.* 2007; Wiemers & Fiedler 2007). Moreover, the ongoing effort to complete the barcode reference database is still far behind in some animal groups, especially in those that are more diverse and less studied, such as most insect orders (Virgilio *et al.* 2010; Jinbo *et al.* 2011).

Here, we explore the utility of COI to barcode a group of closely related and morphologically similar cicadas from southern Europe that can be discriminated on the

Correspondence: Vera L. Nunes, Fax: +351 217500028;
E-mail: vlnunes@fc.ul.pt

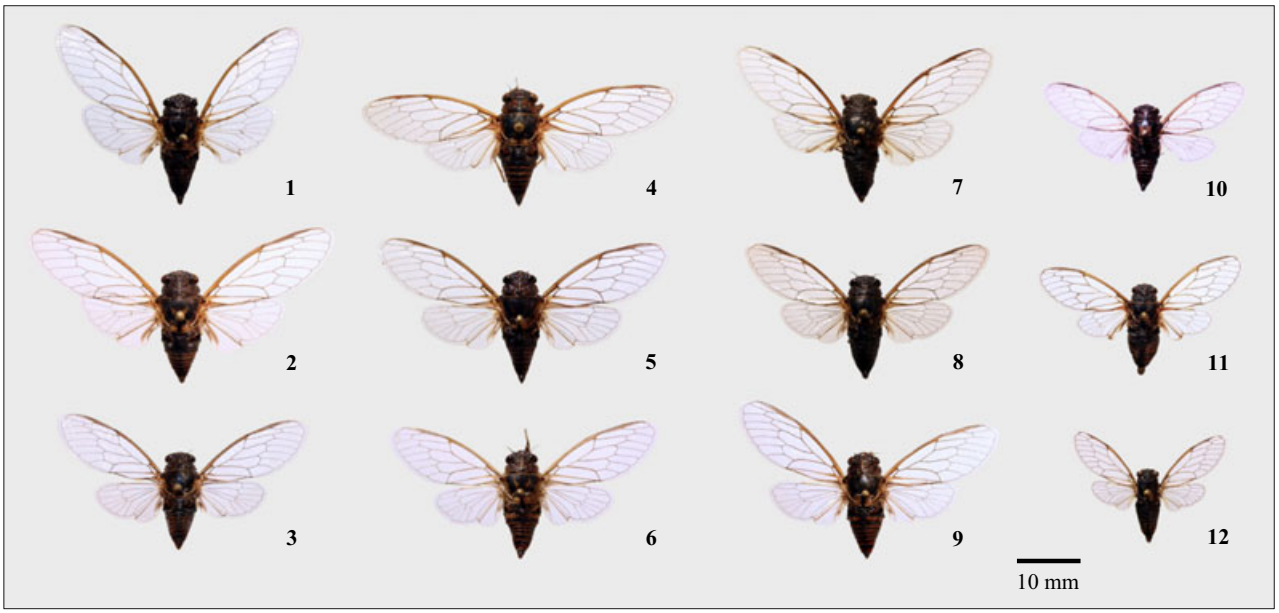
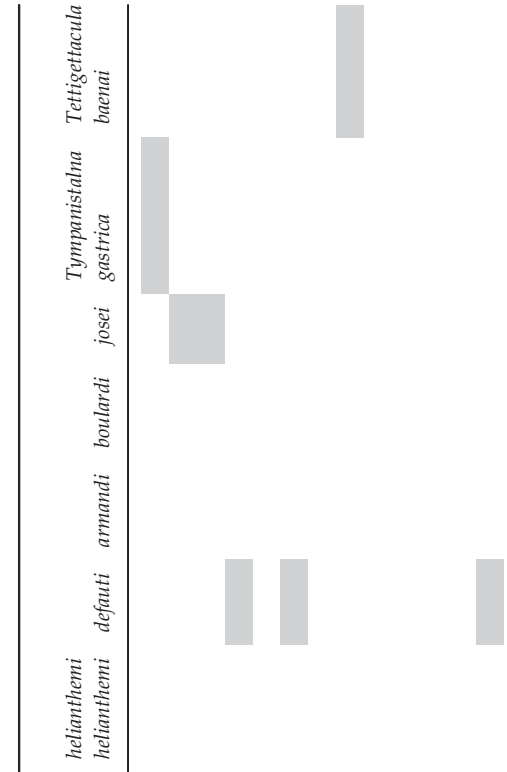


Fig. 1 European cicada species used to test DNA barcoding: 1- *Tettigettalna argentata* (Olivier, 1790); 2- *T. mariae* (Quartau & Boulard 1995); 3- *T. aneabi* (Boulard, 2000); 4- *T. estrellae* (Boulard 1982); 5- *T. helianthemi galantei* Puissant, 2010; 6- *T. helianthemi helianthemi* (Rambur, 1840); 7- *T. defauti* Puissant, 2010; 8- *T. armandi* Puissant, 2010; 9- *T. boulardi* Puissant, 2010; 10- *T. josei* (Boulard 1982); 11- *Tympanistalna gastrica* (Stål, 1854); 12- *Tettigettacula baenai* (Boulard, 2000).

basis of their species-specific male calling songs (Fig. 1). Cicadas have long larval stages spent underground (3–17 years) and a short winged adult stage that lasts only a few weeks (Boulard & Mondon 1995; Williams & Simon 1995). Adult males produce species-specific acoustic signals for sexual recognition by means of a tymbal mechanism, and variation in these signals proved to be an important taxonomic character in cicadas (Claridge 1985; Boulard 2006; Sueur 2006). The diversity of cicadas in the Iberian Peninsula was largely underestimated until the recent description and taxonomic revision of nine small-sized species under the genus *Tettigettalna* (Boulard 1982; Quartau & Boulard 1995; Puissant & Sueur 2010). The distribution of seven of these species is restricted so far to southern Iberian Peninsula (Sueur *et al.* 2004; Puissant & Sueur 2010): *T. josei* (Boulard 1982), *T. mariae* (Quartau & Boulard 1995), *T. aneabi* (Boulard, 2000), *T. helianthemi* (Rambur, 1840), *T. defauti* Puissant, 2010, *T. armandi* Puissant, 2010 and *T. boulardi* Puissant, 2010. The species *T. estrellae* (Boulard 1982) is restricted to the northwest of the Iberian Peninsula, and *T. argentata* (Olivier, 1790) has the broadest known distribution range of the genus, occurring nearly all over the Iberian Peninsula, south of France, Italy, the southern border of Switzerland and in the carstic region of Slovenia, close to the border with Italy (Nast 1972; Gogala & Gogala 1999; Sueur *et al.* 2004; Hertach 2008; Puissant & Sueur 2010). The accurate distribution limits of each cicada are not well known yet, but several of these species have partially overlapping distributions and are usually found in close parapatry or in sympatry (e.g. *T. argentata* and *T. mariae* in Algarve, Portugal, Table 1). Specimens collected in such areas can be easily misidentified if acoustic information is lacking or if the specimens are females, which are unable to produce sound. The correct diagnosis thus requires high taxonomical expertise if based on morphological characters only. An efficient DNA barcoding tool for these cicadas would be extremely useful to assess regional diversity and facilitate ecological studies, because the diagnosis of larval forms, females and males with no song records would be achieved more readily.

An earlier study obtained a few mitochondrial sequences from four *Tettigettalna* species (Fonseca *et al.* 2008), but here, we present the first comprehensive molecular data set for all European species within this genus, which is an important contribution for the barcode reference database, in which European cicadas are severely underrepresented. The 5′ region of COI has not been used in previous phylogenetic and phylogeographical studies on cicadas (e.g. Sueur *et al.* 2007; Marshall *et al.* 2008), and the first study reporting the use of the barcoding region of COI in cicadas (Marshall *et al.* 2012) did not test explicitly for its efficiency as a species assignment tool. The present study is the first to test the utility of COI for DNA barcoding in cicadas and it provides also the first overview on the genetic relatedness among *Tettigettalna* species.

Table 1 Cicadas detected in sympatry during the fieldwork in the Iberian Peninsula



Sympatry with large-bodied species (e.g. *Cicada* sp. or *Tibicina* sp.) was not considered here. Due to its wide distribution range in the Iberian Peninsula, northern populations from *Tettigettala argentata*, [North], were considered separately from the southern ones, [South].

Materials and methods

Sampling

A total of 82 *Tettigettalna* spp. specimens were collected mostly in the Iberian Peninsula, but because of the wide distribution range of *T. argentata*, a few samples of this species were collected in the south of France and in the north of Italy as well, to cover its geographical variation (Fig 2). Table S1 (Supporting information) provides detailed information about the location of each collected specimen. All currently accepted European *Tettigettalna* species or subspecies were sampled (2–13 specimens from each taxon, Table 2). Whenever possible, more than one location was sampled for each species to survey intraspecific variation (Table 2). Two additional small-sized cicadas from Iberian Peninsula usually found in sympatry with *Tettigettalna* species were included in this study (Table 1, Table 2): three specimens from *Tettigettacula baenai* (Boulard, 2000) and three specimens from *Tympanistalna gastrica* (Stål, 1854). Two additional males from the genus *Cicada* were collected to be used as outgroup: one *C. orni* Linnaeus, 1758 and another *C. barbara* Stål, 1866. All collected specimens used in this study were males, and species identification was made in the field based on their species-specific calling songs. We made recordings of the calling song of some collected males for further confirmation of species assignment. Recordings were made in the collecting site using a Marantz PMD 661 Portable SD recorder (20 Hz–24 kHz) connected to a Telinga Pro 7 Dat-mic (60 Hz–18 kHz) microphone (Twin Science). Examples of the calling song of each species are available at the Songs of European Cicadas website designed by M. Gogala (www.cicada-song.eu) with song records provided by most authors that have been working with these species, such as J. Sueur, S. Puissant, M. Boulard or J.A. Quartau. The geographical location of collected specimens was also taken in consideration to confirm species identity following the information provided by Sueur *et al.* (2004) and Puissant and Sueur (2010) about each species' distribution range. Locations of sympatry were avoided to minimize misidentifications. All specimens were captured by hand or by means of a sweeping net. They were then pinned and stored dry at the Department of Animal Biology in the University of Lisbon. A front leg from each specimen was preserved in 100% ethanol for DNA isolation.

DNA isolation and COI amplification

Whole-genome DNA was isolated with the E.Z.N.A.® Tissue DNA Isolation kit (Omega Bio-Tek). Primers LepF (5′-ATTCAACCAATCATAAAGATATTGG-3′) and LepR

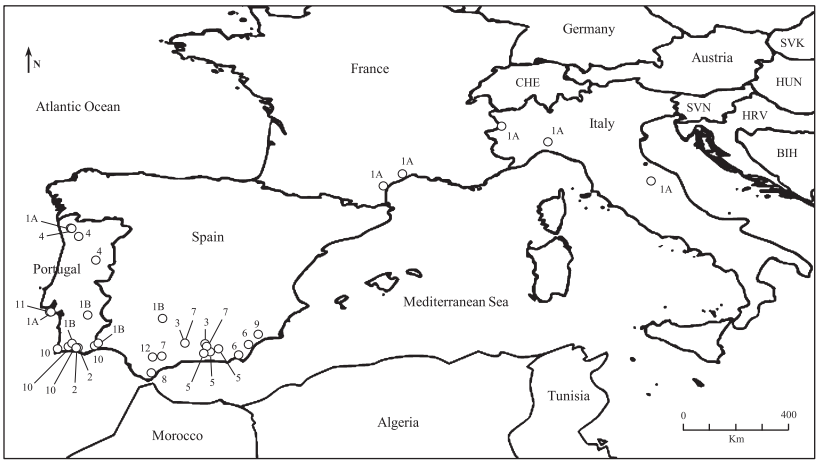


Fig. 2 Map of southern Europe with sampled locations for each cicada species (numbers as in Fig. 1): 1- *Tettigettalna argentata* (1A=[North] and 1B=[South]); 2- *T. mariae*; 3- *T. aneabi*; 4- *T. estrellae*; 5- *T. helianthemi galantei*; 6- *T. helianthemi helianthemi*; 7- *T. defauti*; 8- *T. armandi*; 9- *T. boulardi*; 10- *T. josei*; 11- *Tympanistalna gastrica*; 12- *Tettigettacula baenai*. Country codes: CHE, Switzerland; SVN, Slovenia; HRV, Croatia; SVK, Slovakia; HUN, Hungary; and BIH, Bosnia and Herzegowina.

Table 2 Sequence diversity measures from cytochrome *c* oxidase I (COI) are presented for each cicada species/subspecies: number of haplotypes and variable sites, mean GC content and haplotype (*H*) and nucleotide (π) diversity

					Mean GC content (%) (Standard deviation)					
Species/ subspecies	Sampled locations	No. of specimens	No. of haplotypes	Variable sites	Codon pos. 1	Codon pos. 2	Codon pos. 3	Total	*H*	π
T. argentata	11	20	13	27	44.0 (0.354)	43.0 (0.264)	14.6 (0.493)	33.9 (0.232)	0.942	0.018
T. mariae	2	10	4	6	44.3 (0.163)	42.8 (0.000)	14.9 (0.224)	34.0 (0.089)	0.644	0.002
T. aneabi	2	5	1	0	44.3 (0.000)	42.8 (0.000)	14.0 (0.000)	33.7 (0.000)	0	0
T. estrellae	3	8	5	7	43.7 (0.223)	43.7 (0.223)	11.6 (0.361)	33.0 (0.110)	0.857	0.004
T. helianthemi galantei	3	13	9	31	42.3 (0.186)	43.3 (0.137)	8.8 (0.760)	31.5 (0.285)	0.936	0.019
T. helianthemi helianthemi	2	7	3	10	42.8 (0.000)	42.8 (0.000)	8.9 (1.123)	31.5 (0.403)	0.667	0.009
T. defauti	3	7	4	23	43.2 (0.361)	42.8 (0.000)	9.6 (0.913)	31.9 (0.441)	0.810	0.019
T. armandi	1	2	1	0	43.8 (0.000)	42.8 (0.000)	9.3 (0.000)	32.0 (0.000)	0	0
T. boulardi	1	2	2	3	44.3 (0.729)	42.8 (0.000)	11.7 (0.366)	33.0 (0.365)	1	0.005
T. josei	4	8	5	5	41.3 (0.170)	42.3 (0.000)	9.1 (0.259)	30.9 (0.089)	0.786	0.002
Tympanistalna gastrica	1	3	2	2	42.9 (0.128)	43.3 (0.000)	12.6 (1.213)	33.2 (0.209)	0.667	0.002
Tettigettacula baenai	1	3	1	0	44.3 (0.000)	43.3 (0.000)	7.8 (0.000)	31.8 (0.000)	0	0

(5′-TAAACTTCTGGATGTCCAAAAAATCA-3′) designed for butterflies and moths by Hajibabaei *et al.* (2006) to amplify 658 bp of the 5′ region of the cytochrome *c* oxidase I (COI) gene were used in this study. PCRs were performed in 15 μl of total volume with 0.8 μL (circa 10–20 ng) of DNA template, 1 x PCR buffer (Promega), 0.6 U GoTaq® Flexi DNA polymerase (Promega), 1.0 mM $MgCl_2$, 0.10 mM dNTPs and 0.4 μM of each primer. The cycling conditions used were 94 °C for 3 min, 35 x (94 °C for 30 s, 45 °C for 45 s and 72 °C for 45 s) and 72 °C for 10 min. PCR products were purified with SureClean (Bioline) and sequenced in both directions with standard protocols using the BigDye® Terminator 3.1 cycle sequencing kit (Applied Biosystems) on an automated sequencer ABI PRISM 310 (Applied Biosystems).

Sequence analysis

Forward and reverse sequence strands were assembled and edited in Sequencher 4.0.5 (Gene Codes Co.), to correct noisy and ambiguous base calling. Consensus sequences were then aligned with Clustal W (Thompson *et al.* 1994) as implemented in BioEdit 7.0.9.0 (Hall 1999) and trimmed to the same length. Sequences were translated with the mitochondrial invertebrate genetic code in DnaSP 5.10 (Librado & Rozas 2009) to check for stop codons. The complete mitochondrial genome sequence

from *Philaenus spumarius* (GenBank: AY630340) was used as a reference to position the reading frame in cicada sequences. The number of haplotypes and variable sites, haplotype diversity (H) and nucleotide diversity (π) were determined for each species and subspecies with DnaSP, following the currently accepted taxonomic classification from Puissant and Sueur (2010). Intra- and interspecific genetic distances were calculated with the Kimura 2-parameter (K2P) model, the standard model used in DNA barcoding studies. Genetic distances, nucleotide composition and CG content were computed in Mega 5 (Tamura *et al.* 2011). A neighbour-joining (NJ) tree was constructed in Mega 5 using the K2P model, and node support was assessed with 1000 bootstrap replicates. The NJ tree was compared with a Bayesian phylogenetic tree generated by MrBayes 3.2.1 (Ronquist *et al.* 2012). The best model of sequence evolution (HKY + I + G) for the Bayesian analysis was selected under the Akaike information criterion (AIC), as implemented in MrModeltest 2.3 (Nylander 2004). The Metropolis-coupled Markov chain Monte Carlo analysis was carried out with four chains and two independent runs. The posterior probabilities for each node were generated from 1×10^7 generations, sampling at every 1000th iteration. The burn-in was set to the first 1000 trees, and the remaining trees were used to generate a consensus tree by the 50% majority rule. Conversion of the input file from fasta to nexus was performed with Concatenator 1.1.0 (Pina-Martins & Paulo 2008). *Cicada orni* and *Cicada barbara* were included in NJ and Bayesian inference analyses as outgroup taxa. We followed the tree-based identification criteria proposed by Meier *et al.* (2006): specimens were considered successfully identified when clustering with all conspecifics or forming a polytomy with conspecifics only; specimens were considered misidentified when conspecifics sequences occurred in multiple clusters or clustered with allospecifics, whereas the identification of specimens forming a sister group with conspecifics was considered as ambiguous.

SpeciesIdentifier 1.7.7 (Meier *et al.* 2006) was used to assess the proportion of correctly identified specimens based on K2P genetic distances according to Best Match (BM), Best Close Match (BCM) and All Species Barcodes (ASB) criteria. BM assigns the query to the species with the smallest distance sequence, whereas BCM only identifies the query when the closest sequence is within a distance threshold. The analysis was conducted with a distance threshold of 1.0%. The identification error associated with the 1.0% threshold was 0.02, which is < 0.05 as recommended by Virgilio *et al.* (2012). The same threshold was applied to ASB, which is the most stringent identification criterion, assuming a query as correctly identified only when all conspecific sequences are topping the list of best matches (Meier *et al.* 2006).

Results

Cytochrome *c* oxidase I (COI) was successfully sequenced in all individuals used in this study using the same PCR conditions, except for *Tympanistalna gastrica*. Most sequences from this species recurrently showed noisy and ambiguous base calling at 1.7% of the nucleotides. A range of alternative annealing temperatures was tested (48–55 °C) with little or no improvements in the quality of sequence chromatograms. Therefore, primers LepF and LepR produce suboptimal results for *T. gastrica* specimens, and the coamplification of nuclear copies of COI (pseudogenes) or even the occurrence of heteroplasmy in this species cannot be entirely excluded. Ambiguous bases in these chromatograms were considered as missing data (N) for downstream analyses. The final aligned data set contained 90 sequences of 581 base pairs without gaps or stop codons, corresponding to nucleotide positions 88–668 of the complete COI gene from *Philaenus spumarius*. Sequences were deposited in the GenBank (Accession nos KC807229-KC807318, Table S1, Supporting information).

Sequence variation

Sequences of COI from specimens belonging to genera *Tettigettalna*, *Tettigettacula* and *Tympanystalna* (N = 88) were AT rich (T = 41.6%, C = 14.7%, A = 25.7%, G = 18.0%). The proportion of GC content was the lowest at the third codon position (11.6%), whereas it was similar at the first (43.4%) and second (43.0%) positions. The overall proportion of GC varied among species (Table 2). The highest proportion was observed in *T. mariae* (34.0%), followed closely by *T. argentata* (33.9%) and *T. aneabi* (33.7%), while *T. josei* showed the lowest value (30.9%). The data set with all investigated species (excluding the two outgroup sequences from *Cicada* genus) contained 49 haplotypes and 159 variable sites, from which 145 were parsimony informative and 19 were nonsynonymous substitutions. The number of haplotypes observed in each species ranged from one to 13, and the highest haplotype or nucleotide diversity values were registered in *Tettigettalna argentata* (H = 0.942, π = 0.018), *T. helianthemi galantei* (H = 0.936, π = 0.019), *T. estrellae* (H = 0.857, π = 0.004) and *T. defauti* (H = 0.810, π = 0.019, Table 2).

Genetic distance and DNA barcoding

Mean genetic distance (K2P) between *Tettigettalna* and *Tettigettacula* (11%), *Tympanistalna* (13%) or *Cicada* (22%) was higher than distances between congeneric species. These ranged from 5% to 10% with the following exceptions (Table 3): *T. argentata*/*T. mariae*/*T. aneabi* (≤ 3%),

T. h. helianthemi/*T. h. galantei* (2.9%) and *T. defauti*/*T. armandi* (3.9%). Tree-based DNA barcoding identification could be achieved with success for eight of the 12 tested taxa, but specimens of *T. argentata*, *T. mariae*, *T. h. galantei* and *T. armandi* presented clustering patterns that lead to ambiguous or incorrect identifications.

According to the neighbour-joining (NJ) tree (Fig. 3), *T. argentata*, *T. mariae* and *T. aneabi* seem to compose a complex of species, in which specimens of *T. argentata* do not form a monophyletic group. Specimens of *T. argentata* collected in the south of the Iberian Peninsula (S. B. Messines, Ayamonte, Portel and Espiel, Fig. 2 and Table S1, Supporting information) clustered together with all sampled *T. mariae* specimens, but *T. argentata* individuals from northern locations (Italy, France and two localities from the Iberian Peninsula, Braga and Sesimbra, Fig. 2 and Table S1, Supporting information) are clustered in a distinct clade. The species *T. aneabi* represents the sister taxon of *T. mariae* + *T. argentata* (clade South) according to the NJ analysis. The overall maximum distance observed between *T. argentata* specimens was 3.7%, but this value was largely inflated by the distance between specimens allocated to clades North or South, which reached a mean value of 3.0% (Table 3). As for *T. mariae*, its most common haplotype is shared by one *T. argentata* specimen (Fig. 3, sample *Tar365*), which makes it impossible to unambiguously discriminate *T. mariae* specimens from *T. argentata* (clade South) on the basis of COI sequences only. The maximum intraspecific distance observed in both *T. argentata* (clade South) and *T. mariae* (0.9%) exceeds the minimum distance between them (0.0%, Table 3). As a result, the barcoding gap between intra- and interspecific distance is lacking in this sibling species pair (Fig. 4).

The barcoding gap is also absent in *T. h. galantei* and *T. defauti* due to high levels of intraspecific distance (4.3% and 3.7%, respectively), which overlaps or exceeds the minimum distance to their closest congener (Fig. 4, Table 3). The high intraspecific distance in *T. h. galantei* is caused by three divergent specimens that do not cluster with their conspecifics in the NJ tree and are placed as a sister taxa of *T. boulardi* instead (Fig. 3). High intraspecific distances within *T. defauti* are caused by geographical structure. Specimens collected in Sierra Nevada (Fig. 3, samples *Tde215* and *Tde218*) were nearly as divergent from their conspecifics collected in western locations as they were to their closest congener, *T. armandi* (3.7%, Table 3). The remaining *Tettigettalna* specimens clustered with their conspecifics as expected in the NJ tree (Fig. 3) and were associated with a clear barcoding gap that assures their correct identification with COI sequence data (Fig. 4). The Bayesian inference analysis (Fig. 5) corroborates the results from NJ method with slight differences in tree topology: *T. defauti* and *T. armandi* form an unresolved polytomy and *T. aneabi* is the sister taxon of the clade North from *T. argentata* and not clade South as in the NJ tree. The clade North is placed in the Bayesian tree as the most recently derived haplogroup within the *T. argentata*/*T. aneabi*/*T. mariae* complex, while *T. josei* is the most divergent species within the genus *Tettigettalna*.

Similar problems in DNA barcoding identification based on K2P genetic distances were detected using BM, BCM and ASB criteria (Table S2, Supporting information). With either BM or BCM, 79 sequences (89.8%) were correctly identified, while two (2.3%) were incorrect and eight were ambiguous (9.1%). All 88 sequences tested had a match closer than 1.0% and 59 of those had a match at 0.0% distance. However, seven of these sequences had

Table 3 Mean (lower diagonal) and minimum (upper diagonal) interspecific pairwise K2P distances (%)

	1a	1b	2	3	4	5	6	7	8	9	10	11	12
1a. *T. argentata* [North]	**1.2**	2.3	2.3	2.1	7.5	6.5	6.4	7.6	6.4	7.7	10.0	13.9	11.4
1b. *T. argentata* [South]	3.0	**0.9**	0.0	0.9	7.1	5.0	5.4	6.4	6.4	7.1	9.6	13.1	10.8
2. *T. mariae*	3.0	0.4	**0.9**	1.0	7.5	5.6	6.0	6.6	6.4	7.3	9.2	13.1	11.4
3. *T. aneabi*	2.7	1.1	1.1	**0.0**	8.1	6.3	6.2	7.4	6.4	7.7	9.2	13.7	11.2
4. *T. estrellae*	9.0	8.0	8.1	8.6	**1.0**	4.7	5.0	4.5	5.2	5.4	9.7	12.4	11.2
5. *T. h. galantei*	7.5	6.1	6.3	6.8	5.5	**4.3**	1.7	3.4	5.0	3.0	8.6	10.9	8.6
6. *T. h. helianthemi*	7.4	6.2	6.4	6.3	6.1	2.9	**1.6**	4.3	4.9	5.4	7.8	11.2	9.4
7. *T. defauti*	8.6	7.3	7.5	7.8	5.3	5.0	5.8	**3.7**	3.7	4.5	8.6	10.4	9.4
8. *T. armandi*	7.0	6.6	6.5	6.4	6.0	5.2	5.5	3.9	**0.0**	5.6	8.3	12.2	9.6
9. *T. boulardi*	8.5	7.5	7.6	7.8	5.7	5.1	5.9	5.0	5.7	**0.5**	9.5	12.5	10.0
10. *T. josei*	10.6	9.9	10.0	9.4	10.0	9.1	8.5	9.3	8.4	10.0	**0.7**	11.0	11.6
11. *Tympanistalna gastrica*	15.0	13.9	14.1	14.2	13.1	11.8	12.0	10.2	12.8	13.3	11.7	**0.4**	9.8
12. *Tettigettacula baenai*	11.8	11.3	11.5	11.2	11.3	9.2	9.7	9.7	9.6	10.3	11.9	11.7	**0.0**

Values in the diagonal represent the maximum observed value of intraspecific pairwise distance. Grey shade highlights interspecific distance values that overlap with intraspecific distance.

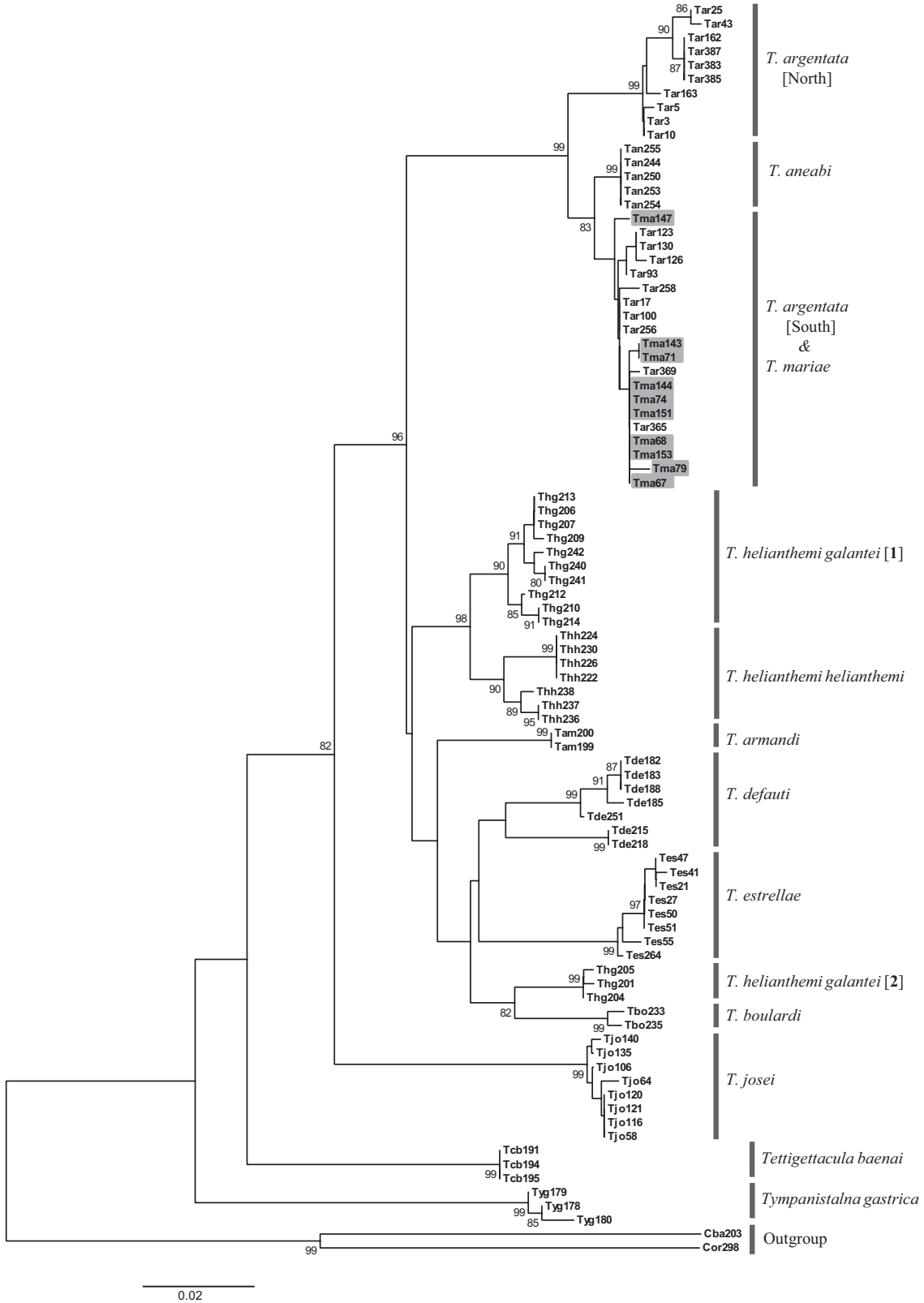


Fig. 3 A Kimura 2-parameter neighbour-joining tree constructed from 581-bp sequences from mitochondrial cytochrome *c* oxidase I (COI) gene. The tree includes 90 sequences from 13 European cicadas. Bootstrap values above 70% are indicated on tree branches. Grey shaded labelled specimens were assigned to *Tettigettalna mariae* according to acoustic data.

a match at 0.0% with an allospecific instead, thus preventing their correct identification. All incorrect or ambiguous identifications under BM or BCM criteria correspond to *T. mariae* or *T. argentata* (clade South) specimens (Table S2, Supporting information). When using the ASB criterion, the number of correct identifications dropped to 59 (58.0%), and the identification of the remaining 37 sequences (42.0%) was considered as ambiguous. These consisted in all *T. argentata* and *T. mariae* specimens, *T. armandi*, *T. boulardi* and the three *T. h. galantei* specimens that do not cluster with their conspecifics in neither NJ nor Bayesian trees (Table S2, Supporting information).

Discussion

Sequence variation

This study generated 90 sequences belonging to 13 European cicada species and reports for the first time the genetic relatedness among all nine cicada species described to date within genus *Tettigettalna*. Primers LepF and LepR (Hajibabaei *et al.* 2006) were efficient to obtain COI fragment sequences for DNA barcoding in all tested cicada species, except for *Tympanistalna gastrica*. The coamplification of nuclear copies of mitochondrial genes has been commonly reported and might be favoured when universal primers are used (Song *et al.* 2008; Moulton *et al.* 2010). Therefore, the redesign of more efficient primers is probably necessary to attain the reliable amplification of mitochondrial COI copies in *T. gastrica*.

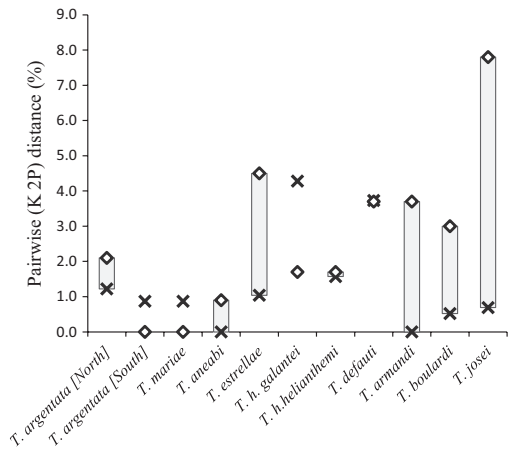


Fig. 4 Barcoding gap observed in genus *Tettigettalna* between maximum intraspecific (black crosses) and minimum interspecific (white diamonds) distance (K2P) to the closest congener. Bars denote the extent of the barcoding gap. These are absent when intraspecific distance overlaps or exceeds minimum congeneric distances.

The total GC content in COI sequences varied among *Tettigettalna* species (30.9–34.0%). These values are similar or slightly larger than the ones observed in other Hemiptera (*Philaenus spumarius*, GC = 29.7%, Seabra *et al.* 2010; *Ceroplastes* spp., GC = 20.4%, Deng *et al.* 2012) but are within the range reported for several other insect orders (27.7–39.5%, Hebert *et al.* 2003a). Although the sampling size is modest for most species here investigated, some showed high levels of haplotype diversity (*T. argentata* and *T. h. galantei*), while others show very little sequence variation (*T. aneabi*).

Genetic distance and DNA barcoding

The genetic distance based on COI sequence variation between cicada genera investigated here (*Tettigettalna*, *Tettigettacula*, *Tympanistalna* and *Cicada*) is clear (> 9.0%) and enables the use of this marker as an efficient DNA barcoding tool for genus-level assignment. However, the power of COI to clearly diagnose closely related species within the genus *Tettigettalna* is compromised by some overlap of intraspecific and congeneric genetic distances. Initial DNA barcoding tests with COI in animal groups suggested that correct species diagnosis would be assured when a clear gap exists between mean intra- and interspecific divergence and the later should exceed the former by at least an order of magnitude (Hebert *et al.* 2004a). This premise indeed holds true for most animal groups and lineages, but recent studies showed that within- and among-species divergence may overlap more often that initially expected. The absence of such a DNA barcoding gap is usually attributed to recent divergence, incomplete lineage sorting, introgression, genetic geographical structure and insufficient sampling of geographical variation (Meyer & Paulay 2005; Wiemers & Fiedler 2007; Linares *et al.* 2009; Bergsten *et al.* 2012).

According to the present study, most *Tettigettalna* species appear well differentiated but show limited genetic divergence, and their origin probably results from relatively recent speciation events. Similarly, some studies with other cicada genera suggested episodes of species radiation to explain diversification patterns (Buckley & Simon 2007; Sueur *et al.* 2007; Marshall *et al.* 2008). For this study, the barcoding gap was considered as the difference between maximum intraspecific distance and the smallest distance to the closest congener, because the use of mean interspecific distances can produce artificially inflated barcoding gaps (Meier *et al.* 2008). *T. josei* and *T. estrellae* are the only species in the genus that fully pass DNA barcoding tests: they are both monophyletic and there is a clear gap between intraspecific variation and genetic distance to every other *Tettigettalna* spp. that enables their correct identification. The species *T. boulardi* apparently passes most DNA barcoding tests too,

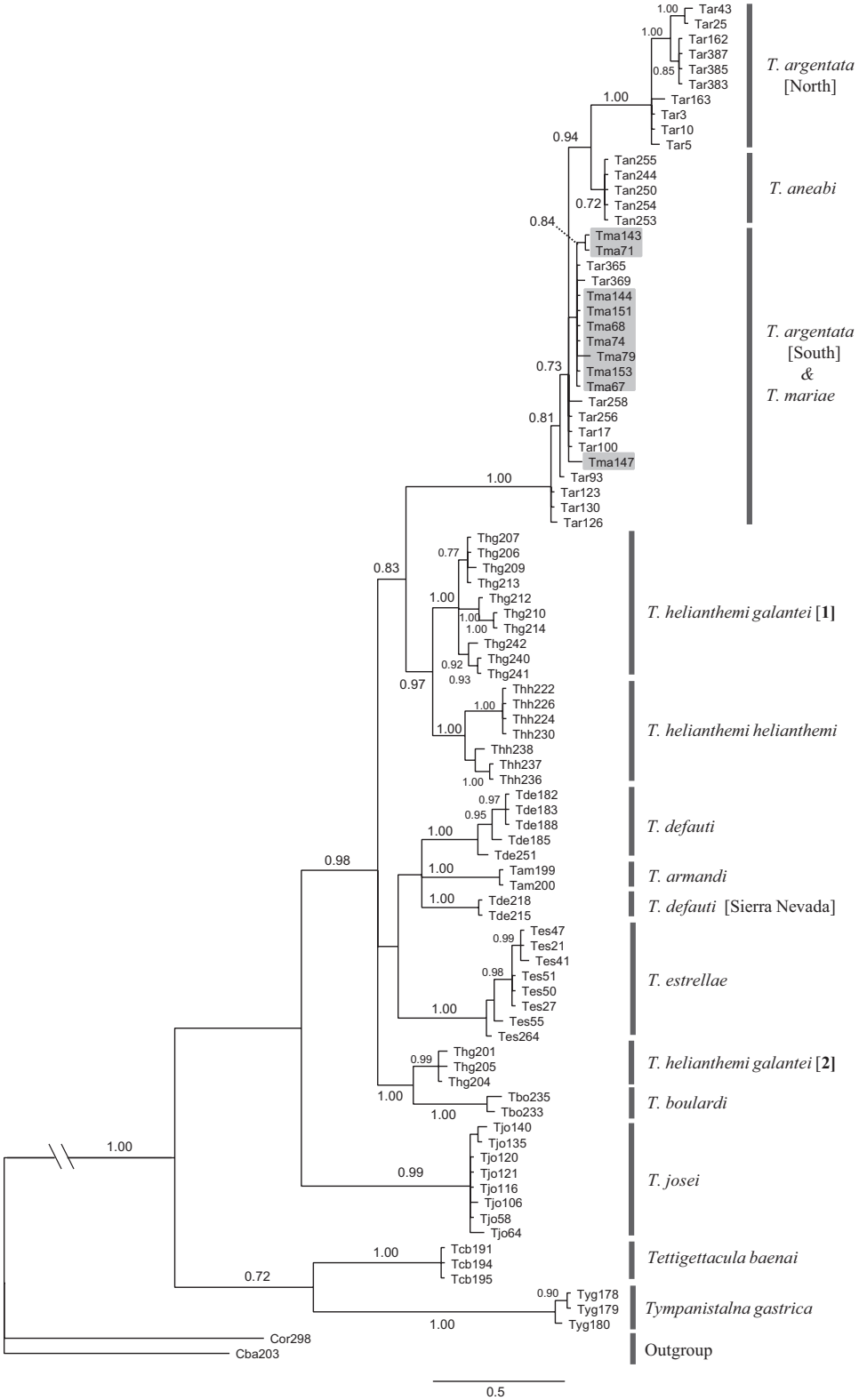


Fig. 5 Bayesian tree constructed from 581-bp sequences from mitochondrial cytochrome *c* oxidase I (COI) gene with the HKY + I + G model of sequence evolution. The tree includes 90 sequences from 13 European cicadas. Posterior probability values higher than 0.70 are indicated on tree branches. The scale bar represents the number of estimated changes per branch length. Grey shaded labelled specimens were assigned to *Tettigettalna mariae* according to acoustic data.

but only two specimens were analysed and intraspecific variation might have been underestimated.

Three *Tettigettalna* species show geographical genetic structure: *T. argentata*, *T. helianthemi* and *T. defauti*. Population genetic structure in *T. argentata* could be expected given its wide distribution range (Fig. 2). Although geographical variation in male calling songs has not been documented yet in this species, northern populations are genetically distinct from the southern ones. For *T. helianthemi* and *T. defauti*, population genetic structure is higher than expected from their restricted distribution range, and in both cases, cryptic divergence was detected in deep association with the Sierra Nevada mountain range (Fig. 2). These mountains are part of the Betic Cordillera in southern Spain, which is considered as a hotspot for Mediterranean biodiversity, harbouring many endemic species (Médail & Quézel 1999; Hewitt 2011). Changes in vegetation and climate are steep along the altitudinal cline, contrasting with the semi-arid lowlands. The two *T. helianthemi* subspecies recognized by Puissant and Sueur (2010) are morphologically distinct, have different calling songs and are geographically segregated. While *T. h. helianthemi* subspecies is associated with dry scrubland plains in the southeast of Sierra Nevada, *T. h. galantei* seems to be less thermophilous and can be found up the mountains till the top (circa 2000 m of altitude). The distinctiveness of these subspecies is clearly supported here by genetic divergence, but we also found that *T. h. galantei* is polyphyletic based on COI sequences. Three specimens collected in the southwestern slope of Sierra Nevada (near Lanjarón, Table S1, Supporting information) are divergent and were recovered as more related to *T. boulardi*, a species apparently limited to the eastern provinces of Murcia and Valencia. These results indicate that *T. h. galantei* from that area must be flagged for further taxonomic scrutiny, as these individuals might represent a cryptic species not yet described. The relationship between *T. defauti* and *T. armandi* also remains unclear as evidenced by the unresolved polytomy in the Bayesian analysis, although *T. armandi* is clearly undersampled and further investigation is needed to clarify this relationship.

Finally, *T. argentata*, *T. mariae* and *T. aneabi* show little COI differentiation and they seem to form a species complex. The cluster composed by these three species corresponds to the most derived haplotypes within *Tettigettalna*. Because these species have distinct male calling songs, acoustic behaviour probably plays an important role in reproductive isolation within this complex, but it remains unclear whether haplotype sharing between *T. mariae* and *T. argentata* is due to introgression or incomplete lineage sorting. The distribution of *T. mariae* is restricted to a small coastal region in the south of Portugal, and this species is often found in sympatry or close parapatry with *T. argentata* (VLN and RM, *personal observation*). Thus, acoustic behaviour isolation must be critical to prevent hybridization among this sibling species pair, and further studies involving mate preference tests are needed to address this question.

Despite the success of COI in DNA barcoding for several animal groups, the use of a single DNA fragment is often not enough to diagnose closely related species, as it happens in several examples among insects (e.g. Kaila & Ståhls 2006; Meier *et al.* 2006; Wiemers & Fiedler 2007; Langhoff *et al.* 2009; Žurovcová *et al.* 2010). Some authors advocate the use of a multigene approach for animals, as commonly implemented in plant DNA barcoding (Elias *et al.* 2007; Dupuis *et al.* 2012). The use of nuclear genes could help to control for misleading COI patterns caused by pseudogenes or species hybridization. However, finding informative nuclear genes for recently diverged species (as we might suspect for *T. mariae*) is challenging because nuclear genes usually evolve much slower than mitochondrial ones. Dasmahapatra *et al.* (2010) demonstrated the utility of AFLPs to determine species limits in butterflies, but a multilocus approach with anonymous loci will probably defeat the purpose of DNA barcoding as a universal, easy to implement, fast and cost-effective tool for species diagnosis.

Conclusions

An efficient DNA barcoding tool for specific molecular diagnosis would be of great use in the taxonomic discrimination of cicadas. The analysis with COI sequences demonstrated that they are insufficient to determine species limits in some cicadas of genus *Tettigettalna* (*T. argentata* and *T. mariae*), and in such cases, male song patterns and geographical location data cannot be replaced by genetic data for specimens' assignment. On the other hand, COI analysis was useful to highlight hidden cryptic diversity (*T. helianthemi* and *T. defauti*). Further studies must be carried out with additional genes and more thorough morphological and acoustic analyses. Because male inherited songs play such an important role in the evolution of species-specific recognition in cicadas, future research should seek informative nuclear genes to overcome the limitations of maternally inherited mitochondrial genes. The use of next-generation sequencing data can significantly speed up the research effort and, at the same time, provide new genomic tools to investigate the complex taxonomic relationships of this group.

Acknowledgements

This work was financially supported by the project PTDC/BIA-BIC/115511/2009 from the Portuguese Foundation for Sci-

ence and Technology (FCT). We thank three anonymous reviewers and Jérôme Sueur for their useful comments, which helped to improve the manuscript.

References

Bergsten J, Bilton DT, Fujisawa T *et al.* (2012) The effect of geographical scale of sampling on DNA barcoding. *Systematics Biology*, **61**, 851–869.

Boulard M (1982) Les cigales du Portugal, Contribution à leur étude (Homoptera: Cicadidae). *Annales de la Société entomologique de France*, **18**, 181–198.

Boulard M (2006) Acoustic signals, diversity and behaviour of cicadas (Cicadidae, Hemiptera). In: *Insect Sounds and Communication* (eds Drosopoulos S & Claridge MF), pp. 331–349. CRC Press, Florida, USA.

Boulard M, Mondon B (1995) Vies et mémoires de cigales. Provence, Languedoc, Méditerranée. Editions de l'Equinoxe.

Buckley TR, Simon C (2007) Evolutionary radiation of the cicada genus Maoricicada Dugdale (Hemiptera: Cicadoidea) and the origins of the New Zealand alpine biota. *Biological Journal of the Linnean Society*, **91**, 419–435.

Claridge MF (1985) Acoustic signals in the Homoptera: behavior, taxonomy, and evolution. *Annual Review of Entomology*, **30**, 297–317.

Dasmahapatra KK, Elias M, Hill RI, Hoffman JI, Mallet J (2010) Mitochondrial DNA barcoding detects some species that are real, and some that are not. *Molecular Ecology Resources*, **10**, 264–273.

Deng J, Yu F, Zhang TX *et al.* (2012) DNA barcoding of six *Ceroplastes* species (Hemiptera: Coccoidea: Coccidae) from China. *Molecular Ecology Resources*, **12**, 791–796.

Dupuis JR, Roe AD, Sperling FAH (2012) Multi-locus species delimitation in closely related animals and fungi: one marker is not enough. *Molecular Ecology*, **21**, 4422–4436.

Elias M, Hill RI, Willmott KR *et al.* (2007) Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society B: Biological Sciences*, **274**, 2881–2889.

Fonseca PJ, Serrão EA, Pina-Martins F *et al.* (2008) The evolution of cicada songs contrasted with the relationships inferred from mitochondrial DNA (Insecta, Hemiptera). *Bioacoustics*, **18**, 17–34.

Gogala M, Gogala A (1999) A checklist and provisional atlas of the Cicadoidea fauna of Slovenia (Homoptera: Auchenorrhyncha). *Acta Entomologica Slovenica*, **7**, 119–128.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 968–971.

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, **41**, 95–98.

Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003a) Biological identification through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, **270**, 313–321.

Hebert PDN, Ratnasingham S, Waard JR (2003b) Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society B: Biological Sciences*, **270**, S96–S99.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM (2004a) Identification of birds through DNA barcodes. *PLoS Biology*, **2**, e312.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004b) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. *Proceedings of the National Academy of Sciences*, **101**, 14812–14817.

Hertach T (2008) A new cicada species for Switzerland: *Tettigetta argentata* (Olivier, 1790) (Hemiptera: Cicadoidea). *Bulletin de la Société Entomologique Suisse*, **81**, 209–214.

Hewitt GM (2011) Mediterranean peninsulas: the evolution of hotspots. In: *Biodiversity Hotspots* (eds Zachos FE & Habel JC), pp. 123–147. Springer, Amsterdam, the Netherlands.

Jinbo U, Kato T, Ito M (2011) Current progress in DNA barcoding and future implications for entomology. *Entomological Science*, **14**, 107–124.

Kaila L, Ståhls G (2006) DNA barcodes: evaluating the potential of COI to diffentiate closely related species of *Elachista* (Lepidoptera: Gelechioidea: Elachistidae) from Australia. *Zootaxa*, **1170**, 1–26.

Langhoff P, Authier A, Buckley TR, Dugdale JS, Rodrigo A, Newcomb RD (2009) DNA barcoding of the endemic New Zealand leafroller moth genera, *Ctenopseustis* and *Planotortix*. *Molecular Ecology Resources*, **9**, 691–698.

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

Linares MC, Soto-Calderón ID, Lees DC, Anthony NM (2009) High mitochondrial diversity in geographically widespread butterflies of Madagascar: a test of the DNA barcoding approach. *Molecular Phylogenetics and Evolution*, **50**, 485–495.

Marshall DC, Slon K, Cooley JR, Hill KBR, Simon C (2008) Steady Plio-Pleistocene diversification and a 2-million-year sympatry threshold in a New Zealand cicada radiation. *Molecular Phylogenetics and Evolution*, **48**, 1054–1066.

Marshall DC, Hill KBR, Marske KA, Chambers C, Buckley TR, Simon C (2012) Limited, episodic diversification and contrasting phylogeography in a New Zealand cicada radiation. *BMC Evolutionary Biology*, **12**, 177.

Médail F, Quézel P (1999) Biodiversity hotspots in the Mediterranean basin: setting global conservation priorities. *Conservation Biology*, **13**, 1510–1513.

Meier R, Shiyang K, Vaidya G, Nq PK (2006) DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic Biology*, **55**, 715–728.

Meier R, Zhang G, Ali F (2008) The use of mean instead of smallest interspecific distances exaggerates the size of the "barcoding gap" and leads to misidentification. *Systematic Biology*, **57**, 809–813.

Meyer CP, Paulay G (2005) DNA barcoding: error rates based on comprehensive sampling. *PLoS Biology*, **3**, e422.

Moritz C, Cicero C (2004) DNA barcoding: promise and pitfalls. *PLoS Biology*, **2**, e354.

Moulton MJ, Song H, Whiting MF (2010) Assessing the effects of primer specificity on eliminating numt coamplification in DNA barcoding: a case study from Orthoptera (Arthropoda: Insecta). *Molecular Ecology Resources*, **10**, 615–627.

Nast J (1972) *Palaearctic Auchenorrhyncha (Homoptera), an Annotated Check List*. Polish Scientific Publications, Warsaw, Poland.

Nylander JAA (2004) MrModeltest v2. Program distributed by the author. Evolutionary Biology Center, Uppsala University.

Pina-Martins F, Paulo OS (2008) Concatenator: sequence data matrices handling made easy. *Molecular Ecology Resources*, **8**, 1254–1255.

Puissant S, Sueur J (2010) A hotspot for Mediterranean cicadas (Insecta: Hemiptera: Cicadidae): new genera, species and songs from southern Spain. *Systematics and Biodiversity*, **8**, 555–574.

Quartau JA, Boulard M (1995) *Tettigetta mariae*, nouvelle Cigale lusitanienne (Homoptera Cicadoidea Tibicinidae). *EPHE, Biologie et Evolution des Insectes*, **7**, 105–110.

Ronquist F, Teslenko M, van der Mark P *et al.* (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.

Seabra SG, Pina-Martins F, Marabuto E *et al.* (2010) Molecular phylogeny and DNA barcoding in the meadow-spittlebug *Philaenus spumarius* (Hemiptera, Cercopidae) and its related species. *Molecular Phylogenetics and Evolution*, **56**, 462–467.

Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences*, **105**, 13486–13491.

Sueur J (2006) Insect species and their songs. In: *Insect Sounds and Communication* (eds Drosopoulos S & Claridge MF), pp. 207–217. CRC Press, Florida, USA.

Sueur J, Puissant S, Simões PC, Seabra S, Boulard M, Quartau JA (2004) Cicadas from Portugal: revised list of species with eco-ethological data (Hemiptera: Cicadidae). *Insect Systematics and Evolution*, **35**, 177–187.

Sueur J, Vanderpool D, Simon C, Ouvrard D, Bourgoin T (2007) Molecular phylogeny of the genus Tibicina (Hemiptera, Cicadidae): rapid radiation and acoustic behaviour. *Biological Journal of the Linnean Society*, **91**, 611–626.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, **28**, 2731–2739.

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, **22**, 4673–4680.

Valentini A, Pompanon F, Taberlet P (2008) DNA barcoding for ecologists. *Trends in Ecology and Evolution*, **24**, 110–117.

Virgilio M, Backeljau T, Nevado B, De Meyer M (2010) Comparative performances of DNA barcoding across insect orders. *BMC Bioinformatics*, **11**, 206.

Virgilio M, Jordaens K, Breman FC, Backeljau T, De Meyer M (2012) Identifying insects with incomplete DNA barcode libraries, African fruit flies (Diptera: Tephritidae) as a test case. *PLoS ONE*, **7**, e31581.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **360**, 1847–1857.

Wiemers M, Fiedler K (2007) Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). *Frontiers in Zoology*, **4**, 8.

Williams KS, Simon C (1995) The ecology, behavior, and evolution of periodical cicadas. *Annual Review of Entomology*, **40**, 269–295.

Žurovcová M, Havelka J, Starý P *et al.* (2010) "DNA barcoding" is of limited value for identifying adelgids (Hemiptera: Adelgidae) but supports traditional morphological taxonomy. *European Journal of Entomology*, **107**, 147–156.

P.C.S., V.L.N., J.A.Q., S.G.S. and O.S.P. designed the study. B.N., E.M., V.L.N., T.H., P.C.S. and R.M. collected the specimens. R.M. and E.M. obtained and edited the cicadas' photos. V.L.N. obtained and analysed sequence data and wrote the article. All authors contributed to the final version of the text.

Data Accessibility

DNA sequences: GenBank accession numbers are listed in Table S1 (Supporting information), KC807229-KC807318.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Sampled locations, GPS coordinates and GenBank accession numbers associated with each collected cicada specimen

Table S2 DNA barcoding identification of 88 cicada specimens using Best Match (BM), Best Close Match (BCM) and All Species Barcodes (ASB) criteria