A mirage of cryptic species: genomics uncover striking mito-nuclear discordance in the skipper butterfly *Thymelicus sylvestris*

Joan Carles Hinojosa[1], Darina Koubínová[2], Mark Szenteczki[3], Camille Pitteloud[4], Vlad Dincă[5], Nadir Alvarez[2]* & Roger Vila[1]*

[1] Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain

[2] Unit of Research and Collection, Museum of Natural History, Geneva, Switzerland

[3] Laboratory of Functional Ecology, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland

[4] Department of Environmental Systems Sciences, Institute of Terrestrial Ecosystems, ETHZ, Zürich, Switzerland

[5] Department of Ecology and Genetics, University of Oulu, Finland

* These authors participated equally and are considered joint senior authors

Correspondence: roger.vila@csic.es

Keywords: *mitochondrial DNA, genomics, RAD sequencing, despeciation, Lepidoptera, lineage fusion*

Running title: Mito-nuclear discordance in *Thymelicus sylvestris*

ABSTRACT

Mitochondrial DNA (mtDNA) sequencing has led to an unprecedented rise in the identification of cryptic species. However, it is widely acknowledged that nuclear DNA (nuDNA) sequence data are also necessary to properly define species boundaries. Next generation sequencing techniques provide a wealth of nuclear genomic data, which can be used to ascertain both the evolutionary history and taxonomic status of putative cryptic species. Here, we focus on the intriguing case of the butterfly *Thymelicus sylvestris* (Lepidoptera: Hesperiidae). We identified six deeply diverged mitochondrial lineages; three distributed all across Europe and found in sympatry, suggesting a potential case of cryptic species. We then sequenced these six lineages using double-digest restriction-site associated DNA sequencing (ddRADseq). Nuclear genomic loci contradicted mtDNA patterns and genotypes generally clustered according to geography, i.e., a pattern expected under the assumption of postglacial recolonization from different refugia. Further analyses indicated that this strong mtDNA/nuDNA discrepancy cannot be explained by incomplete lineage sorting, sex-biased asymmetries, NUMTs, natural selection, introgression or *Wolbachia*-mediated genetic sweeps. We suggest that this cyto-nuclear discordance was caused by long periods of geographic isolation followed by range expansions, homogenizing the nuclear but not the mitochondrial genome. These results highlight *T. sylvestris* as a potential case of multiple despeciation and/or lineage fusion events. We finally argue, since mtDNA and nuDNA do not necessarily follow the same mechanisms of evolution, their respective evolutionary history reflects complementary aspects of past demographic and biogeographic events.

INTRODUCTION

The study of cryptic biodiversity, diversity overlooked due to morphological similarities, has become a trending topic in recent years (Bickford et al., 2007; Struck et al., 2018). Cryptic biodiversity is apparently widespread, but its frequency among Metazoa and distribution across biogeographical regions is subject to debate (Pfenninger & Schwenk, 2007; Trontelj & Fišer, 2009). Despite their similarities, cryptic species may not always be sister taxa and can have different ecology and behaviour (McBride, Van Velzen, & Larsen, 2009; Vodă, Dapporto, Dincă, & Vila, 2015a). Ignoring the existence of this phenomenon can lead to incorrect assessments of biodiversity in the present as well as when modelling the future (Bálint et al., 2011), resulting in inadequate conservation management and eventually to biodiversity loss.

Assessing the status of potential cryptic species can be a challenging task. It is often recommended to combine morphological, ecological and genetic data, including multiple independent molecular markers of mitochondrial and nuclear DNA (e.g. Hernández-Roldán et al., 2016; Von Helversen et al., 2001). However, multi-copy nuclear markers such as *ITS2* may not always display a perfectly concerted evolution (Shapoval & Lukhtanov, 2015) and single-copy nuclear genes with sufficient variability have historically been scarce. Nowadays, massive amounts of nuclear genetic data can be obtained relatively easily using next-generation sequencing techniques based on restriction enzymes, i.e. double-digest restriction site-associated DNA sequencing (ddRADseq; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012), providing a solution to the limited resolution recovered with few DNA markers.

Although mitochondrial DNA (mtDNA) is widely used to assess genetic patterns, results obtained from mtDNA may differ from the ones based on nuclear DNA (Bernardo et al., 2019; Galtier, Nabholz, Glemin, & Hurst, 2009; Toews & Brelsford, 2012). Several mechanisms have been shown to generate mito-nuclear discordance, namely incomplete lineage sorting, sex-biased asymmetries, introgression, natural selection or genetic sweeps mediated by *Wolbachia* infection (Toews & Brelsford, 2012). Moreover, each genome, and even independent markers, may reflect distinct parts of the evolutionary history of an organism. Consequently, integrating markers of both genomes may help achieve a better understanding of biological processes, especially when processes such as the above mentioned are at work.

In relatively well-studied groups such as butterflies, it has been shown that cryptic taxa may represent a significant fraction of the total diversity. For example, about 25% of the west Mediterranean butterfly species could be considered to belong to a cryptic group given morphological similarity, with most of the cryptic taxa within groups occurring in allopatry (Vodă, Dapporto, Dincă, & Vila, 2015b). In Europe, it has been shown that nearly 28% of the currently accepted species include deeply diverged mitochondrial intraspecific lineages according to Generalized Mixed Yule-Coalescent (GMYC) model results (Dincă et al., 2015). Some of these lineages may represent cryptic species and directed studies are needed to ascertain their evolutionary history and taxonomic status.

The small skipper, *Thymelicus sylvestris* Poda, 1761 (Lepidoptera: Hesperiidae) was found by the GMYC model to display multiple deeply diverged mitochondrial lineages that were characteristic of the species level (Dincă et al., 2015). This fairly generalist species is widely distributed in the Western Palearctic, with its larvae feeding on various Poaceae plants (Tolman & Lewington, 2008). No strong morphological or ecological variability has been described within the species. However, four deeply diverged mitochondrial lineages recovered as separate entities by GMYC have been documented in Europe and North Africa (Dincă et al., 2015), some of which occur in sympatry. Minimum uncorrected pairwise distances of the *COI* barcode region among these lineages ranged between 1.8% and 3.6%, consistent with values reported for closely related species (Ashfaq, Akhtar, Khan, Adamowicz, & Hebert 2013; Huemer, Mutanen, Sefc, & Hebert, 2014). If these lineages actually represented real species, then *T. sylvestris* would be one of the most remarkable examples of cryptic diversity in Eurasian butterflies and their discovery would have notable implications for further research on butterflies, as well as for monitoring and nature conservation.

Here, we first sequenced the mitochondrial cytochrome *c* oxidase subunit I (*COI*) gene, a standard and widely used barcoding region, for a new set of *T. sylvestris* samples that spans the whole species' distribution. Then we used ddRADseq to validate the existence of cryptic taxa within *T. sylvestris* and to explore its phylogeographic history. We found no clear evidence of cryptic species within *T. sylvestris*. Nuclear loci mostly correlated with geography and intermediate specimens were present across its distribution range. We searched for an explanation for the strong mtDNA-nuDNA discrepancy and it was apparently not caused by incomplete lineage sorting, sex-biased asymmetries, NUMTs, selection processes, introgression or *Wolbachia*-mediated genetic sweeps. Thus, we suggest that the pattern observed is caused by geographic isolations followed by range expansions that produced generalized recombination of the nuclear genome. With all, *T. sylvestris* arises as a potential model to study despeciation and/or lineage fusion events. We propose that the two genomes respond differently to demographic and spatial events because of their particular evolutionary mechanisms – e.g. four-folds larger effective population size in mtDNA vs. nuDNA, meiotic segregation and recombination in nuDNA *vs*. maternally transmitted haploid mtDNA –, which may result in notably different patterns in particular conditions~~.~~

METHODS

Sample collection

We collected 63 samples of *T. sylvestris*, spanning their geographic distribution and all known *COI* main lineages (Table S1). We also added three *Thymelicus lineola* (Ochsenheimer, 1808) that correspond to three different GMYC entities reported in Dincă et al. (2015). Additionally, we added one *Thymelicus acteon* (Rottemburg, 1775) to the ddRADseq analyses as a root species, although it is not entirely clear whether *T. lineola* or *T. acteon* is a sister species of *T. sylvestris*. Butterflies collected from the field were dried, wings were stored separately as vouchers, and bodies were stored in ethanol 99% at -20ºC.

COI sequencing

We sequenced 44 additional individuals for the *COI* barcode region. DNA extraction and amplification were done following the protocol described in Dincă, Lukhtanov, Talavera, and Vila (2011) and using the same primers to obtain the 658 bp barcode fragment of *COI*. PCR amplification conditions were: initial denaturation at 92ºC for 60 s, followed by 5 cycles at 92ºC for 15 s, annealing at 48ºC for 45 s, extension at 62ºC for 150 s and other 30 cycles changing the annealing temperature to 52ºC with the final extension step at 62ºC for 7 min. PCR products were purified and Sanger sequenced by Macrogen Inc. Europe (Amsterdam, North Holland, the Netherlands). All sequences are available on GenBank (Table S1).

Mitochondrial analyses and phylogenetic reconstruction

DNA sequences were aligned with Geneious v11.0.5 (Kearse et al., 2012). The best fitting model was found using jModelTest v2.1.7 (Darriba, Taboada, Doallo, & Posada, 2012) and the phylogeny was reconstructed in BEAST v2.5.0 (Bouckaert et al., 2014). We estimated base frequencies, selected four gamma rate categories, and used a randomly generated initial tree. Rough estimates of node ages were obtained by applying a strict clock and a normal prior distribution centered on the mean between two generally accepted substitution rates for invertebrates, i.e. 1.5% and 2.3% uncorrected pairwise distance per million years (Quek, Davies, Itino and Pierce (2004) and Brower (1994), respectively). The standard deviation was tuned so that the 95% confidence interval of the posterior density coincided with the 1.5% and 2.3% rates. We did two independent runs of 20 million generations each, convergence was checked using TRACER 1.7.1 (Rambaut, 2018) with a 10% burn-in. To obtain an objective delimitation of the mitochondrial lineages, we ran bPTP (Zhang, Kapli, Pavlidis, & Stamatakis, 2013) on the Bayesian tree for 500,000 MCMC iterations, with thinning set to 100 and 10% burn-in.

We constructed a maximum parsimony haplotype network with the *COI* sequences, using the TCS Network method in PopART v1.7 (Clement, Snell, Walke, Posada, & Crandall, 2000). Minimum genetic distances between groups were calculated with MEGA v7.0.14 (Kumar, Stecher, & Tamura, 2016), using uncorrected p-distances (Collins, Boykin, Cruickshank, & Armstrong, 2012; Srivathsan & Meier, 2012) and the bootstrap method to estimate variance.

Mitochondrial sequences of the genes *COI* and *NADH5* recovered from the ddRADseq data were aligned and concatenated with Geneious v11.0.5 (Kearse et al., 2012) resulting in a 174 bp fragment. A maximum likelihood phylogeny was performed in Geneious v11.0.5 (Kearse et al., 2012) using PhyML v3.0 (Guindon et al., 2010) with GTR model and 1,000 bootstrap replicates.

ddRADseq library preparation

DNA was extracted from half thoraxes using a Qiagen DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA, USA) following the manufacturer’s recommended protocol. DNA was eluted in 50µl of EB buffer. DNA concentrations were quantified with a Qubit 2.0 fluorometer (Life Technologies, Norwalk, CT, USA) and ranged between 24ng/µL and 114ng/µl. Six µl of each sample was digested with MseI and SbfI, and P1 (24 different barcodes) + P2 (the same sequence for all samples) adaptors were ligated. Samples were cleaned with 0.8x AMPure XP magnetic beads (Beckman-Coulter, Brea, CA, USA). Following a double indexing approach, samples were individually amplified and tagged using Illumina sequencing primers. PCR conditions are listed in Table S2. Two replicates per sample were used to reduce PCR bias, and corresponding PCR products were then pooled. All the individual libraries were further pooled in equimolar ratio and cleaned with 1x AMPure XP beads. After quantification and quality check of the libraries pool with Fragment Analyzer (Agilent, Santa Clara, CA, USA), fragments at around 325 bp were selected using Pippin Prep (Sage Science, Beverly, MA, USA). A final cleaning step was conducted with 1x AMPure XP beads. Finally, the libraries were sequenced using HiSeq 2000 100pb paired-end sequencing at Lausanne Genomic Technologies Facility (Lausanne, Vaud, Switzerland).

ddRADseq data analyses

Alignment, SNP calling and initial filtering steps were carried out using pyRAD v3.0 (Eaton, 2014). We tested the following parameter values: clustering thresholds (c) 0.86 and 0.88; minimum number of taxa per locus (m) 10, 15, 25 and 40; minimum depth of coverage required to build a cluster (d) 6; maximum number of shared polymorphic sites in a locus (p) 2 and 4. Following initial exploratory analyses and quality checks, the final, optimal parameters we selected were c = 0.86, m = 10, d = 6 and p = 4.

The pyRAD SNP matrix was improved by an additional rare allele filtering that removes alleles found in less than 5% of the samples. This dataset was used for SNPs phylogeny and STRUCTURE. *Thymelicus lineola* and *T. acteon* samples were excluded before running STRUCTURE in order to restrict the analyses to groups within *T. sylvestris*. The composition of the groups was calculated using STRUCTURE v2.3.4 (Pritchard, Stephens, & Donnelly, 2000) with the number of groups (K) allowed to vary between 1 to 8. *Thymelicus lineola* and *T. acteon* samples were excluded from STRUCTURE analyses in order to restrict the analyses to *T. sylvestris*. The selected burn-in was set at 75,000, followed by 250,000 MCMC replicates run to obtain the cluster data. 25 runs were done for each K, which were after combined in one per group with Clumpack v1.1 (Kopelman, Mayzel, Jakobsson, Rosenberg, & Mayrose, 2015). The optimal K was calculated using Structure Harvester v0.6.94 (Earl & vonHoldt, 2012). The plot was visualized in Distruct v1.1 (Rosenberg, 2004) and a map was prepared with QGIS v2.8.6 (www.qgis.org).

With the pyRAD output loci file we performed a search with Centrifuge v1.0.4 (Kim et al., 2016) in order to identify the loci of bacterial endosymbionts such as *Wolbachia.* An examination of mitochondrial genes present in the ddRADseq data was done with BLAST+ (Camacho et al., 2008). The same dataset but restricted to *T. sylvestris* was used for the BayeScan v2.1 analysis (Foll & Gaggiotti, 2008) to search for loci with a variability that significantly correlates with the mitochondrial lineages.

For the Extended Bayesian Skyline Plot (EBSP) we used 19 loci with more than 3 SNPs from a subset – lineages 1 and 5, encompassing all the available European distribution – of 47 samples. We used BEAST v2.5.0 (Bouckaert et al., 2014) setting a chain length of 2,500,000,000. Three runs were performed and a burn-in of 50-65% was applied. Convergence was checked with TRACER 1.7.1 (Rambaut, 2018).

ddRADseq data phylogenetic reconstruction

Using the alignment with all the loci generated by pyRAD, a phylogenetic analysis was performed through a maximum likelihood inference with RAxML v8.2.4 (Stamatakis, 2014) using GTR+CAT model and 1,000 bootstrap replicates. The same alignment was used for a quartet-based coalescent-approach with SVDquartets (Chifman & Kubatko, 2014). We ran 1,000 bootstrap replicates.

Two Bayesian phylogenies were explored with BEAST v2.5.0 (Bouckaert et al., 2014) using two independent sets of 200 randomly picked loci present in at least 30 samples. We calibrated the phylogeny with an intron mutation rate of 3.68% divergence per million years (Papadopoulou, Anastasiou, & Vogler, 2010). A GTR model with four gamma categories was selected, base frequencies were estimated and a randomly generated initial tree was used. Parameters were estimated using two runs of 100 million generations and convergence was checked with TRACER 1.7.1 (Rambaut, 2018). A 10% burn-in was applied.

An additional Bayesian tree was done with the SNPs dataset. We ran BEAST v2.5.0 (Bouckaert et al., 2014) twice with a GTR model with four gamma categories, estimated base frequencies and 20 million generations. Convergence was checked with TRACER 1.7.1 (Rambaut, 2018) and a 10% burn-in was applied.

Detecting *Wolbachia* infections

A total of 33 loci were retrieved with BLAST+ as likely belonging to *Wolbachia*. Individuals were considered infected if at least one these loci was detected. Then they were concatenated, edited and corrected manually with Geneious v11.0.5 (Kearse et al., 2012) to obtain a 2,873 bp alignment. We calculated the average distances (d_{XY}) and net mean distances (d_A) between groups with MEGA v7.0.14 (Kumar et al., 2016) using uncorrected p-distances (Collins et al., 2012; Srivathsan et al., 2012) and the bootstrap method to estimate variance.

RESULTS

Mitochondrial phylogenetics

bPTP identified eight entities (Figure S1) within *T. sylvestris*, but two were not considered since bPTP posterior probabilities and divergence time were too low (p.p. < 0.6, estimated mean age < 0.5 Mya) (Figure 1a). Additionally, bPTP separated the three *T. lineola* in three entities with a posterior probability of 1.

The distributions of the lineages (Figure 2a) were as follows: lineages 1 and 5 were widespread in Europe, lineage 2 appeared restricted to the Pindus mountain range (Greece) – not reported in Dincă et al. (2015) –, lineage 3 seemed to be present in the eastern half of Europe, lineage 4 was only found near the Danube delta, and a single individual from Crimea represented lineage 6 – not reported in Dincă et al. (2015). Some of the specimens studied were collected in sympatry but belong to different mitochondrial lineages, comprising groups 1-2, 1-3 and 1-5 (Table S1).

The six mitochondrial *T. sylvestris* lineages displayed minimum divergences of more than 1% between them and were supported by posterior probabilities of 1 in the Bayesian gene tree based on *COI* (Figure 1a, Table S3). Every lineage is also highlighted in the haplotype network (Figure 1b). According to our age estimates, all these lineages appeared in pairs in a window time of 0.48-1.56 million years ago (Mya). Lineages 1 and 2 emerged 0.93 (± 0.45) Million years ago (Mya), 3 and 4, 1.02 (± 0.48) Mya, and 5 and 6, 1.06 (± 0.5) Mya. The relationship between the ancestors of each pair is not fully resolved since posterior probabilities were low. Nonetheless the estimated age of their most recent common ancestor was 1.5 (± 0.65) Mya.

Amongst the ddRADseq loci we found two mitochondrial loci: 60 samples had a fragment of the *COI* gene (87 bp) and 14 samples had a fragment of the *NADH5* gene (87 bp). The maximum likelihood phylogeny constructed with these concatenated sequences is consistent with the pattern obtained from the barcode phylogeny (Figure S2).

ddRADseq data pre-processing

Total reads obtained from sequencing were 1.31x10^{8}. More than 86% of the raw sequencing data passed the quality filters. Reads retained per sample varied between 4.2x10^{5} and 3.05x10^{6}, with a median value of 1.58x10^{6} (Table S4). pyRAD grouped all the reads into 7.31x10^{6} clusters, with values ranging between 2.9x10^{3} and 1.9x10^{5} clusters per sample, with a median of 1.1x10^{5} and a mean depth of 12.02 reads per cluster (Table S4). pyRAD recovered 2,532 different loci and the number of loci recovered per sample ranged from 247 to 1,304 (Table S4), with a median of 867 loci. After the rare alleles filtering step done for the STRUCTURE analysis and SNPs phylogeny, we kept 1,360 SNPs.

Genetic structure based on SNPs

The Structure Harvester ΔK table for STRUCTURE runs from K=1 to K=8 based on ddRADseq SNPs indicated that K=7 was the most likely number of clusters (ΔK= 320.9). The STRUCTURE barplot (Figure S3) shows that the mitochondrial lineages do not correlate with the nuclear genetic diversity present in the ddRAD genetic clusters. The RAD cluster map (Figure 2b) exhibits a notable spatial effect on the genetic structure. Importantly, samples with intermediate cluster composition were common.

The Iberian Peninsula and Northern Africa showed mainly two clusters, both almost exclusively. Two clusters were found only in Eastern Europe. The other clusters are distributed across all Europe but with different weight depending on the longitude.

Loci and SNPs phylogenies

The maximum likelihood phylogeny based on ddRADseq loci (Figure 3) displayed a relatively poorly supported topology: individuals tended to group into small clusters, sometimes well supported, but with low divergence between them, compared to divergences with respect to *T. lineola* and *T. acteon*. Groups according to geographical distribution were obtained, but were commonly found in polyphyly.

Additional phylogenies (Figure S4; Figure S5) showed similar results but with the following slight differences. Bayesian phylogenies with 200 random loci (Figure S4a; Figure S4b) defined three main clades inside *T. sylvestris*. The groups did not have a geographic correlation, for example, samples from the Iberian peninsula were found in two groups while Italian individuals were present in all three clades. The estimated age of their most recent common ancestor was ca. 1.04 (± 0.1) Mya (Figure S4a; Figure S4b). The SVDquartets phylogeny (Figure S4c) only had good support for one clade; similarly to the signal retrieved from Bayesian phylogenies, no clear geographic correspondence was found in this dataset. SNPs phylogeny (Figure S5) had resemblances to the Bayesian phylogenies based on 200 loci: three clades with a similar composition were retrieved, but with some individuals shifted.

Virtually no mitochondrial lineage was recovered as monophyletic in the ddRADseq phylogenetic reconstructions (Figure 3; Figure S4; Figure S5). The only exception was mitochondrial lineage 4 (Figure 3; Figure S4c), which consisted of only three samples from south-eastern Romania. While the bootstrap support for this clade was 97 in the maximum likelihood phylogeny, it was recovered within a much wider clade that coincided within a genetic cluster in the STRUCTURE analysis.

Wolbachia data

From the ddRADseq data, we could extract and concatenate 33 loci belonging to *Wolbachia* from 47 samples: 44 *T. sylvestris*, 2 *T. lineola* and the *T. acteon* (Table S1). All mitochondrial lineages included infected individuals and infection rates per mtDNA lineage ranged between 33.3% and 81.5% – except for lineage 6, for which the single individual sampled was not infected – (Table S5). The divergence matrix of the concatenated *Wolbachia* loci did not show any divergence between *T. sylvestris* individuals belonging to different mitochondrial lineages (Table S6). In some cases, net mean distances between groups (d_A) yielded slightly negative values because differences among individuals within lineages were slightly higher than among lineages. *Thymelicus acteon* showed a strain distinct than that of *T. sylvestris*, with differences concentrated in four loci. Unfortunately, it was not possible to recover these four loci in *T. lineola and* the others were identical to those present in *T. sylvestris*. *Wolbachia* loci alignment was uploaded to Dryad (code: XXX).

Additional analyses

The BayeScan search for ddRADseq loci displaying the same pattern as mtDNA retrieved only one locus, which was a fragment of the *COI* gene itself, the only one we sequenced from the mitogenome. Finally, the plotted results of the EBSP analysis (Figure S6) showed a recent and abrupt increase of the *T. sylvestris* effective population size.

DISCUSSION

Six potential cryptic species uncovered in mitochondrial DNA data

Our species delimitation analysis on mtDNA data recovered six potential cryptic species with bPTP posterior probabilities > 0.6 and estimated mean age > 0.5 Mya (Figure S1), two more than previously reported using GMYC (Dincă et al., 2015). Genetic divergences among these mitochondrial lineages in some cases exceeded 3% minimum p-distance (lineage 5 compared to 3 and 4; Table S3). This degree of divergence may indeed be compatible with the hypothesis of multiple species, according to typical interspecific *COI* distances in butterflies. For example, when comparing DNA barcodes of 1004 Lepidoptera from two European sites separated by 1600 km, it has been found that the minimum distance to the nearest neighbour species averaged 7.17% but it was < 2% for 2.49% of recognized species (Huemer et al., 2014). In butterflies from Pakistan, 3.7% of the species analysed exhibited a minimum distance to the nearest neighbour species < 3% (Ashfaq et al., 2013). In Hesperiidae, Hebert, Penton, Burns, Janzen, and Hallwachs (2004) proposed ten new species derived from *Astraptes fulgerator* with an average divergence of 2.76% between them and similarly divergent mitochondrial lineages have been shown to represent actual cryptic species in the genus *Spialia* (Hernández-Roldán et al., 2016).

We estimate that the mtDNA of *T. sylvestris* experienced a first diversification period ca. around 1.25-1.5 Mya, that resulted into the three ancestors from which originated the final six lineages we detected. The confidence intervals of the age estimates for all the lineages widely overlap, with average times of 0.93-1.06 Mya. This suggests that they could have been diverged concomitantly within a short time range. In *T. lineola*, four lineages were reported in Dincă et al. (2015) and single representatives of three of them were analysed here as well. bPTP divided *T. lineola* in three entities with a posterior probability of 1, suggesting that this taxon can be another good candidate for a deeper study. *Thymelicus lineola* exhibit similar divergence times compared to those found in *T. sylvestris*. Considering that the habitat of *T. lineola* highly overlaps with that of *T. sylvestris* (Engler, Balkenhol, Filz, Habel, & Rödder, 2014) and that divergence times recovered within the two species were similar, it is possible that the same events that drove the emergence of mitochondrial lineages in *T. sylvestris* could also have affected *T. lineola*. In summary, the mitochondrial data indicate two major events that caused the isolation of *T. sylvestris* in different populations, a first ca. 1.25-1.5 Mya and a second ca. 1 Mya, with Quaternary glacial events likely involved.

Massive mito-nuclear discordance

Results obtained from the analysis of the ddRADseq data do not support the presence of cryptic species. Some structuring was found in *T. sylvestris* but the retrieved clades were not uniform across the analyses and not well supported in the maximum likelihood phylogeny with all the loci (Figure 3; Figure S4; Figure S5). Although three well supported clades were obtained with Bayesian phylogenies the sets of 200 loci (Figure S4a; Figure S4b) and with the SNPs tree (Figure S5), some samples shifted between both analyses. Geographical structure obtained is compatible with intraspecific genetic variability since specimens with intermediate genetic composition were frequent in the contact zones among most nuclear clusters (Figure 2b). Additionally, divergences among ddRADseq clades of *T. sylvestris* were much lower than divergences with respect to *T. lineola* or *T. acteon* (Figure 3).

Instead, phylogeographic patterns obtained using ddRADseq loci were not concordant with lineages identified by mitochondrial *COI* sequencing. The mitochondrial lineages were polyphyletic in all the phylogenies made with ddRADseq data (Figure 3; Figure S4; Figure S5), with the exception of the poorly represented lineage 4, which may be a product of isolation-by-distance. Furthermore, BayeScan was not able to recover any nuclear locus displaying the same six-lineage pattern displayed by *COI*. We obtained the six mitochondrial linages using mtDNA sequences recovered from the ddRADseq loci (Figure S2). Hence, we discard the possibility of significant mistakes and/or contaminations.

Conflicts between nuclear and mitochondrial genomes have been observed in many taxa and these can be caused by a number of processes (Toews & Brelsford, 2012). Considering our case, in which we found a strong discrepancy between a mitochondrial gene and a wide set of nuclear sequences, we rejected the following hypotheses:

1) **Incomplete lineage sorting**. The mitochondrial genome reflects a deep differentiation among lineages, with estimated mean ages ranging between 0.93 and 1.06 Mya. Such ancient divergence times are not easily compatible with a lack of coalescence in the nuclear markers. Even if we assume that some nuclear markers may not have coalesced, an important fraction should have. On the contrary, we did not detect any nuclear locus that correlates with the deep mitochondrial lineages.

2) **Sex-biased asymmetries**. As the mitochondrial genome is only inherited maternally, sex-biased asymmetries may result in different patterns in the mitochondrial and nuclear genomes (Toews & Brelsford, 2012). It is not known if on sex disperses more than the other

in *T. sylvestris* and, generally, we cannot assume that males disperse more than females in butterflies (e.g. Baguette, Vansteenwegen, Convi, & Nève, 1996; Petit, Moilanen, Hanski, & Baguette, 2001). Disjunct distributions or severely fragmented populations are scenarios in which sex-biased dispersal capabilities could play a strong role in mtDNA-nuDNA discrepancy (e.g. Nietlisbach et al., 2012), but this is not the case of *T. sylvestris*. Moreover, females represent half of the effective population and it is most unexpected not finding any nuclear locus among thousands that reflect the matrilineal history. Although we cannot entirely discard this hypothesis, we consider unlikely that sex-biased asymmetries cause such a strong mito-nuclear discordance in *T. sylvestris*.

3) **NUMTs**. We discarded the effects of nuclear mitochondrial DNA segments (NUMTs) (Ribeiro, 2012) because no double peaks and no stop codons were detected in any of the *COI* sequences obtained with Sanger sequencing. Also, a single sequence was obtained for each mtDNA locus obtained with ddRADseq (Figure S2). Moreover, the existence of six differentiated lineages would be hard to explain because it would require multiple nuclear capture events. The same reasoning – except for the stop codons – applies to discarding heteroplasmy as a potential cause for the pattern observed.

4) **Selection**. Natural selection can be a source of divergence (Cheviron & Brumfield, 2009; Irwin, 2012). In the case of *T. sylvestris* adaptive selection on *COI* producing divergent lineages can be discarded because all the mutations detected were synonymous. Moreover, no apparent ecological or life history differentiation has been documented in this species, and the six mitochondrial lineages are not distributed following any obvious environmental cline.

5) **Introgression**. Introgression is another mechanism that may produce deep mitochondrial divergence within a species (Muñoz, Baxter, Linares, & Jiggins, 2012). It implies interspecific hybridization that produces the substitution of the mitochondrial genome of a species for the

one of another species. In the case of *T. sylvestris*, introgression from several different species should have occurred to fully explain the obtained mitochondrial pattern only with introgression. However, *T. sylvestris* was recovered as monophyletic with respect to the closely related *T. lineola* and *T. acteon* in the *COI* gene tree (Figure 1b & figure 3), as was the case in a previous study that covered ca. 60% of the European butterfly species and included numerous related Pyrginae taxa (Dincă et al., 2015). Although relatively unlikely, past mtDNA introgression events followed by the extinction of the donor species could be an explanation for the discrepancy.

6) ***Wolbachia***. In insects, one of the most frequent phenomena resulting in mito-nuclear discrepancy is the presence of *Wolbachia* (Toews & Brelsford, 2012). This bacteria is maternally inherited and sometimes cause male-killing or cytoplasmic incompatibility (Hurst & Jiggins, 2005; Jiggins, 2003; Ritter *et al*., 2013; Werren, Baldo, & Clark, 2008). As a result, the infection may cause selective sweeps where a mitochondrial genome is associated with a *Wolbachia* strain. Our endosymbiont search found *Wolbachia* infection to be widespread, but no substantial differences in *Wolbachia* were shown among the mitochondrial lineages (Table S5), neither in terms of infection rates nor in terms of strain identity. In fact, all mitochondrial lineages were apparently widely infected by the same strain. Thus, infection by *Wolbachia* should not affect gene flow among *T. sylvestris* mitochondrial lineages. Nevertheless, we cannot discard past infections by different strains of *Wolbachia* could have promoted lineage creation; but no trace of these strains were detected in present day butterfly populations.

***Thymelicus sylvestris*, a despeciation model?**

Thymelicus sylvestris current genetic pattern might be the signature of speciation reversal (despeciation) or lineage fusion processes that occurred in the past. These imply long isolation periods and posterior secondary contacts when genetic variability is pooled. The mitochondrial genome is inherited only from females and does not segregate or recombine. As a result, past isolation events are theoretically detectable in the mtDNA in the form of diverged haplotypes, even after populations come back in contact with one another and fuse. This is especially true in species with large population sizes and under the scenario of population growth, when haplotypes can be maintained for a long time. On the contrary, due to meiotic segregation and recombination, the fingerprints of previous periods of isolation can be rapidly lost in the nuDNA due to gene flow. Thus, the ideal scenario for this hypothesis requires: 1) The existence of long isolation periods; 2) large population growth that causes secondary contact and avoids fixation of one mtDNA lineage; 3) gene flow is currently maintained. Worth noting, instead of a single long period of isolation, several consecutive shorter periods followed by generalised gene flow would result in a similar genetic pattern, as long as the mtDNA lineages are maintained across these events. This could have taken place during the Quaternary glacial cycles for relatively mobile and widespread organisms and could explain that mitochondrial lineages notably older than the last glacial event are routinely detected within species. Complementary, the EBSP analysis indicated a recent abrupt increase of the population size for *T. sylvestris*, which supports the hypothesis of geographic isolation followed by a population growth and posterior secondary contact (Figure S6). The increase in population size could be linked to a range expansion, probably postglacial, that allowed gene flow among previously isolated populations. The mostly gradual differentiation of the nuDNA that correlates with geography (Figure 2) reflects the postglacial scenario of gene flow across the range of the species.

Cases of speciation reversal involving deeply diverged lineages are scarce and they have not been yet documented in butterflies. Whether *T. sylvestris* could have experienced speciation reversal instead of lineage fusion cannot be ensured, but divergences found in the barcode region are compatible with those found between butterfly species, which makes possible that speciation reversal processes occurred. Hence *T. sylvestris* could be a potentially useful model to study speciation reversal and/or lineage fusion phenomena.

Despeciation and lineage fusion driven by the isolation and secondary contact mechanism leading to mito-nuclear discordance may also have happened in other organisms, especially in those that at present display high population numbers and important levels of gene flow. For example, a recent case has been described in birds (Kearns et al., 2018). In wide-scale DNA barcoding surveys, highly diverged mtDNA haplotypes are sometimes detected at low frequency and scattered across the range of common species, as in the butterfly *Melitaea didyma* (Esper, 1778) (Pazhenkova & Lukhtanov, 2016). It is likely that some of such lineages are remnants of ancient evolutionary events that were not entirely lost because of high effective population size.

We argue that mitochondrial genomes might be key for detecting ongoing or past despeciation events. Thus, the different behaviour of mitochondrial and nuclear genomes should not be seen as an obstacle, but an opportunity. As they reflect different processes, they offer a chance to better understand the evolutionary history of the organisms in a wide sense.

Conclusions

Single-marker approaches display a reasonable compromise between cost, time and data quality for preliminary assessments of mega-diverse faunas, poorly known taxonomic groups and potential cryptic diversity. However, in this study we add evidence to the fact that these approaches are actually not sufficient for unambiguously delimiting species and no taxonomic decisions should be taken solely based on such data. Based on our wider sampling across the distribution range of *T. sylvestris* we found not four, but six deeply diverged mitochondrial lineages. In contrast, ddRADseq loci did not show evidence of cryptic species within *T. sylvestris*: data mostly correlated with geography and intermediate specimens were common at the contact zones. The strong mtDNA/nuDNA discrepancy detected is apparently not caused by incomplete lineage sorting, sex-biased asymmetries, NUMTs, selection processes, introgression or *Wolbachia*-mediated genetic sweeps. Instead, the hypothesis of geographic isolation followed by range expansion that produced generalized recombination of the nuclear genome seems to be the most plausible. Thus, we argue *T. sylvestris* could be a good model for studying despeciation and/or lineage fusion. Finally, we propose that the nuclear and mitochondrial genomes respond differently to demographic and spatial events because of their particular evolutionary mechanisms – e.g. four-folds larger effective population size in mtDNA vs. nuDNA, and meiotic segregation and recombination in nuDNA vs. maternally transmitted haploid mtDNA – which may result in notably different patterns in particular conditions.

Acknowledgments

We thank Raluca Vodă and Leonardo Dapporto for helping us to obtain the samples. Financial support for this research was provided by project CGL2016-76322-P (AEI/ERDF, UE) to Roger Vila and by predoctoral fellowship BES-2017-080641 to Joan Carles Hinojosa. Nadir Alvarez was funded by project PP00P3_172899 from the Swiss National Science Foundation. Computational resources for certain analyses were supplied by the Ministry of Education, Youth and Sports of the Czech Republic under the Projects CESNET (Project No. LM2015042) and CERIT-Scientific Cloud (Project No. LM2015085) provided within the program Projects of Large Research, Development and Innovations Infrastructures.

Tables and Figures

Table S1. Samples used in this study.

Sample ID	Genbank code	Species	*COI* lineage	Country	Latitude (deg)	Longitude (deg)	*Wolbachia* infected
Rvcoll06M948	HQ005242	*T. sylvestris*	1	Romania	45,09	26,53	Yes
Rvcoll06V657	HQ005243	*T. sylvestris*	3	Romania	46,70	23,55	Yes
Rvcoll06V658	HQ005234	*T. sylvestris*	1	Romania	46,70	23,55	Yes
Rvcoll07C554	MK812966	*T. sylvestris*	5	Germany	50,44	8,92	No
Rvcoll07D561	HQ005235	*T. sylvestris*	1	Romania	45,83	23,11	Yes
Rvcoll07D567	HQ005238	*T. sylvestris*	1	Romania	45,83	23,11	Yes
Rvcoll07D570	HQ005240	*T. sylvestris*	3	Romania	45,83	23,11	Yes
Rvcoll07D904	HQ005236	*T. sylvestris*	3	Romania	45,14	23,75	No
Rvcoll07D942	HQ005244	*T. sylvestris*	1	Romania	45,14	23,75	Yes
Rvcoll08A015	HQ005239	*T. sylvestris*	1	Romania	46,72	23,65	Yes
Rvcoll08H340	HM901574	*T. sylvestris*	5	Spain	40,37	-3,37	Yes
Rvcoll08J187	GU676540	*T. sylvestris*	5	Portugal	41,88	-7,72	Yes
Rvcoll08L005	HM901261	*T. sylvestris*	5	Spain	37,15	-3,49	Yes
Rvcoll08M051	GU676167	*T. lineola*	-	Spain	40,67	-2,67	No
Rvcoll08M397	HQ005233	*T. sylvestris*	4	Romania	45,27	28,18	No
Rvcoll08M409	HQ005232	*T. sylvestris*	4	Romania	45,27	28,18	No
Rvcoll08M474	HQ005229	*T. sylvestris*	4	Romania	44,82	28,69	Yes
Rvcoll08M502	HQ005228	*T. sylvestris*	3	Romania	46,48	23,73	No
Rvcoll08M596	HQ005231	*T. sylvestris*	3	Romania	46,79	25,68	Yes
Rvcoll08P489	MK812953	*T. sylvestris*	5	Spain	42,88	-4,88	Yes
Rvcoll10B804	MK812965	*T. sylvestris*	1	France	44,25	6,67	Yes
Rvcoll11D879	KP870957	*T. sylvestris*	1	Spain	36,80	-3,93	Yes
Rvcoll11E182	KP870245	*T. sylvestris*	5	Spain	42,19	1,85	Yes
Rvcoll11F133	MK812947	*T. sylvestris*	5	Morocco	35,12	-5,19	Yes
Rvcoll11F699	MK812930	*T. sylvestris*	5	Morocco	34,09	-4,18	No
Rvcoll11F863	MK812922	*T. sylvestris*	5	Morocco	32,99	-5,08	Yes
Rvcoll11H773	MK812926	*T. sylvestris*	1	Italy	37,85	14,71	Yes
Rvcoll11H977	MK812951	*T. sylvestris*	1	Italy	37,80	13,99	Yes
Rvcoll11I099	MK812929	*T. sylvestris*	1	Italy	37,99	14,88	Yes
Rvcoll11I260	MK812948	*T. sylvestris*	1	Italy	39,86	16,06	Yes
Rvcoll11I262	MK812959	*T. sylvestris*	5	Italy	39,86	16,07	No
Rvcoll11I501	KP870751	*T. sylvestris*	1	Spain	37,08	-3,51	Yes
Rvcoll11J141	MK812942	*T. sylvestris*	1	Switzerland	46,33	7,97	Yes
Rvcoll11J902	MK812940	*T. sylvestris*	3	Bulgaria	41,85	24,95	Yes
Rvcoll12N789	MK812939	*T. sylvestris*	1	Greece	39,07	26,38	Yes
Rvcoll12O632	MK812945	*T. sylvestris*	5	France	44,20	7,07	Yes
Rvcoll12R407	MK812949	*T. sylvestris*	5	Italy	43,07	11,29	Yes
Rvcoll14A332	MK812958	*T. sylvestris*	5	Italy	43,07	13,23	Yes
Rvcoll14B776	MK812962	*T. sylvestris*	2	Albania	40,62	20,49	No
Rvcoll14B843	MK812924	*T. sylvestris*	5	Bosnia	44,04	16,61	Yes
Rvcoll14D089	MK812931	*T. sylvestris*	1	Bulgaria	41,62	24,70	No
Rvcoll14E536	MK812963	*T. sylvestris*	5	Romania	44,65	21,92	No
Rvcoll14E875	MK812927	*T. sylvestris*	5	Serbia	44,36	21,89	Yes
Rvcoll14F196	MK812941	*T. sylvestris*	5	Serbia	43,40	22,37	No
Rvcoll14F208	MK812961	*T. lineola*	-	Serbia	43,40	22,37	Yes
Rvcoll14F661	MK812956	*T. sylvestris*	2	Greece	40,21	21,06	No
Rvcoll14F685	MK812955	*T. sylvestris*	2	Greece	40,23	20,96	Yes
Rvcoll14G273	MK812925	*T. sylvestris*	1	Greece	37,78	22,22	No
Rvcoll14G416	-	*T. sylvestris*	2	Greece	40,56	21,23	Yes
Rvcoll14G456	MK812923	*T. sylvestris*	1	Greece	40,86	21,20	Yes
Rvcoll14G730	MK812933	*T. sylvestris*	1	Greece	39,90	20,78	No
Rvcoll14H212	MK812954	*T. sylvestris*	1	Greece	38,48	22,51	Yes
Rvcoll14H908	MK812935	*T. sylvestris*	1	France	44,84	2,22	Yes
Rvcoll14I368	MK812938	*T. sylvestris*	1	Italy	42,58	11,14	Yes
Rvcoll14I413	MK812950	*T. sylvestris*	1	Italy	40,69	15,00	No
Rvcoll14I428	-	*T. acteon*	-	Italy	40,83	15,07	Yes
Rvcoll14J760	MK812943	*T. sylvestris*	5	France	43,45	2,68	Yes
Rvcoll14J892	MK812934	*T. lineola*	-	France	45,14	2,71	Yes
Rvcoll14V062	MK812944	*T. sylvestris*	3	Ukraine	50,18	36,40	Yes
Rvcoll14W437	MK812967	*T. sylvestris*	1	UK	53,33	-3,83	Yes
Rvcoll14W456	MK812936	*T. sylvestris*	1	UK	53,30	-2,20	Yes
Rvcoll15F148	MK812960	*T. sylvestris*	1	Portugal	39,44	-9,20	Yes
Rvcoll15G033	MK812946	*T. sylvestris*	5	Belgium	49,80	5,74	No
Rvcoll15G054	MK812937	*T. sylvestris*	5	France	49,07	7,51	No
Rvcoll15I358	MK812964	*T. sylvestris*	3	Austria	46,84	13,44	Yes
Rvcoll15I830	MK812928	*T. sylvestris*	1	Austria	47,15	13,38	No
Rvcoll15Q170	MK812932	*T. sylvestris*	6	Russia	-	-	No

Table S2. Reaction conditions used for the ddRADseq protocol.

Digestion		
Step	**Temperature (ºC)**	**Time (min)**
Digestion	37	180
Deactivation	65	20
Ligation		
Step	**Temperature (ºC)**	**Time (min)**
Ligation	16	180
Amplification I		
Step	**Temperature (ºC)**	**Time (s)**
Denaturation I	98	30
Denaturation II (21x)	98	20
Annealing (21x)	60	30
Elongation I (21x)	72	40
Elongation II	72	600
Amplification II		
Step	**Temperature (ºC)**	**Time (min)**
Denaturation	98	3
Annealing	30	2
Elongation	72	12

Table S3. Minimum genetic distances in substitutions per site between mitochondrial lineages based on the *COI* gene.

	***T. sylvestris* 1**	***T. sylvestris* 2**	***T. sylvestris* 3**	***T. sylvestris* 4**	***T. sylvestris* 5**	***T. sylvestris* 6**	***T. lineola***	***T. acteon***
***T. sylvestris* 1**	-							
***T. sylvestris* 2**	0,015	-						
***T. sylvestris* 3**	0,024	0,028	-					
***T. sylvestris* 4**	0,018	0,024	0,025	-				
***T. sylvestris* 5**	0,026	0,03	0,032	0,036	-			
***T. sylvestris* 6**	0,021	0,021	0,028	0,027	0,024	-		
T. lineola	0,048	0,047	0,062	0,059	0,059	0,05	-	
T. acteon	0,044	0,052	0,055	0,052	0,056	0,052	0,055	-

Table S4. Information about the ddRADseq data obtained per sample. *Number of reads* represents those kept by pyRAD in step 2 and *number of clusters* refers to those grouped by the same program in step 3. *Number of final loci* refers to the final dataset obtained after the pyRAD pipeline.

Sample ID	Nº reads	Nº clusters	Nº final loci
Rvcoll06M948	1561276	70650	600
Rvcoll06V657	2627484	180871	454
Rvcoll06V658	2337117	123025	463
Rvcoll07C554	1569094	123408	359
Rvcoll07D561	2550644	25978	450
Rvcoll07D567	3052100	17633	377
Rvcoll07D570	2459958	150353	551
Rvcoll07D904	2146762	39419	504
Rvcoll07D942	2028139	85031	597
Rvcoll08A015	603635	32785	361
Rvcoll08H340	1726692	101099	348
Rvcoll08J187	1387694	119051	307
Rvcoll08L005	2198392	107697	356
Rvcoll08M051	1210939	72953	97
Rvcoll08M397	2143444	107928	350
Rvcoll08M409	2536849	128497	453
Rvcoll08M474	2485892	169062	392
Rvcoll08M502	1334270	109548	313
Rvcoll08M596	1391610	120012	375
Rvcoll08P489	1418951	113330	297
Rvcoll10B804	2031928	144978	335
Rvcoll11D879	1400999	79377	316
Rvcoll11E182	1449520	83184	332
Rvcoll11F133	2034555	159631	297
Rvcoll11F699	2057332	186027	359
Rvcoll11F863	2185676	161054	391
Rvcoll11H773	1611848	123076	304
Rvcoll11H977	1811334	141541	299
Rvcoll11I099	2024356	89001	336
Rvcoll11I260	1574977	115528	311
Rvcoll11I262	2552344	162981	404
Rvcoll11I501	2300921	160377	312
Rvcoll11J141	1530917	114924	280
Rvcoll11J902	804383	97673	597
Rvcoll12N789	734549	88903	375
Rvcoll12O632	2625527	176926	350
Rvcoll12R407	1565894	129625	328
Rvcoll14A332	1681505	140860	372
Rvcoll14B776	2486652	170703	342
Rvcoll14B843	1772199	152548	409
Rvcoll14D089	1380763	115543	377
Rvcoll14E536	1939240	147322	352
Rvcoll14E875	1635480	130564	345
Rvcoll14F196	1299228	103868	392
Rvcoll14F208	2775485	59022	103
Rvcoll14F661	1918426	177639	424
Rvcoll14F685	2319183	154006	413
Rvcoll14G273	1433744	117006	328
Rvcoll14G416	2227680	138865	347
Rvcoll14G456	1358224	42883	338
Rvcoll14G730	1568452	125045	402
Rvcoll14H212	1305983	98862	434
Rvcoll14H908	421910	2882	201
Rvcoll14I368	1482695	101055	526
Rvcoll14I413	1943818	133091	349
Rvcoll14I428	1579429	109693	150
Rvcoll14J760	746545	41434	234
Rvcoll14J892	1688506	93433	104
Rvcoll14V062	1533771	127030	303
Rvcoll14W437	1359917	97762	347
Rvcoll14W456	1519416	115574	431
Rvcoll15F148	647169	68979	272
Rvcoll15G033	554014	76389	492
Rvcoll15G054	1184402	92471	421
Rvcoll15I358	811188	63866	276
Rvcoll15I830	667234	55454	261
Rvcoll15Q170	488065	38736	301

Table S5. Number and percentage of individuals infected by *Wolbachia* per mitochondrial lineage. Individuals were considered infected if at least one *Wolbachia* locus was present.

COI lineage	**Nº individuals**	**Nº *Wolbachia* infected**	***Wolbachia* infected (%)**
1	27	22	81,5
2	4	2	50
3	8	6	75
4	3	1	33,3
5	20	13	65
6	1	0	0

Table S6. Genetic distances in substitutions per site between concatenated *Wolbachia* loci for infected individuals partitioned according to mitochondrial lineage. Above-right: net mean distances between groups (d_A); bottom-left: average distances between groups (d_{XY}).

	***T. sylvestris* 1**	***T. sylvestris* 2**	***T. sylvestris* 3**	***T. sylvestris* 4**	***T. sylvestris* 5**	***T. lineola***	***T. acteon***
***T. sylvestris* 1**	**-**	-0,001	-0,001	-0,001	0	-0,001	0,030
***T. sylvestris* 2**	0	-	0	0	0	0	0,030
***T. sylvestris* 3**	0	0	-	0	0	0	0,010
***T. sylvestris* 4**	0	0	0	-	0	0	0,130
***T. sylvestris* 5**	0	0	0	0	-	0	0,020
T. lineola	0	0	0	0	0	-	0
T. acteon	0,030	0,030	0,010	0,130	0,020	0	-

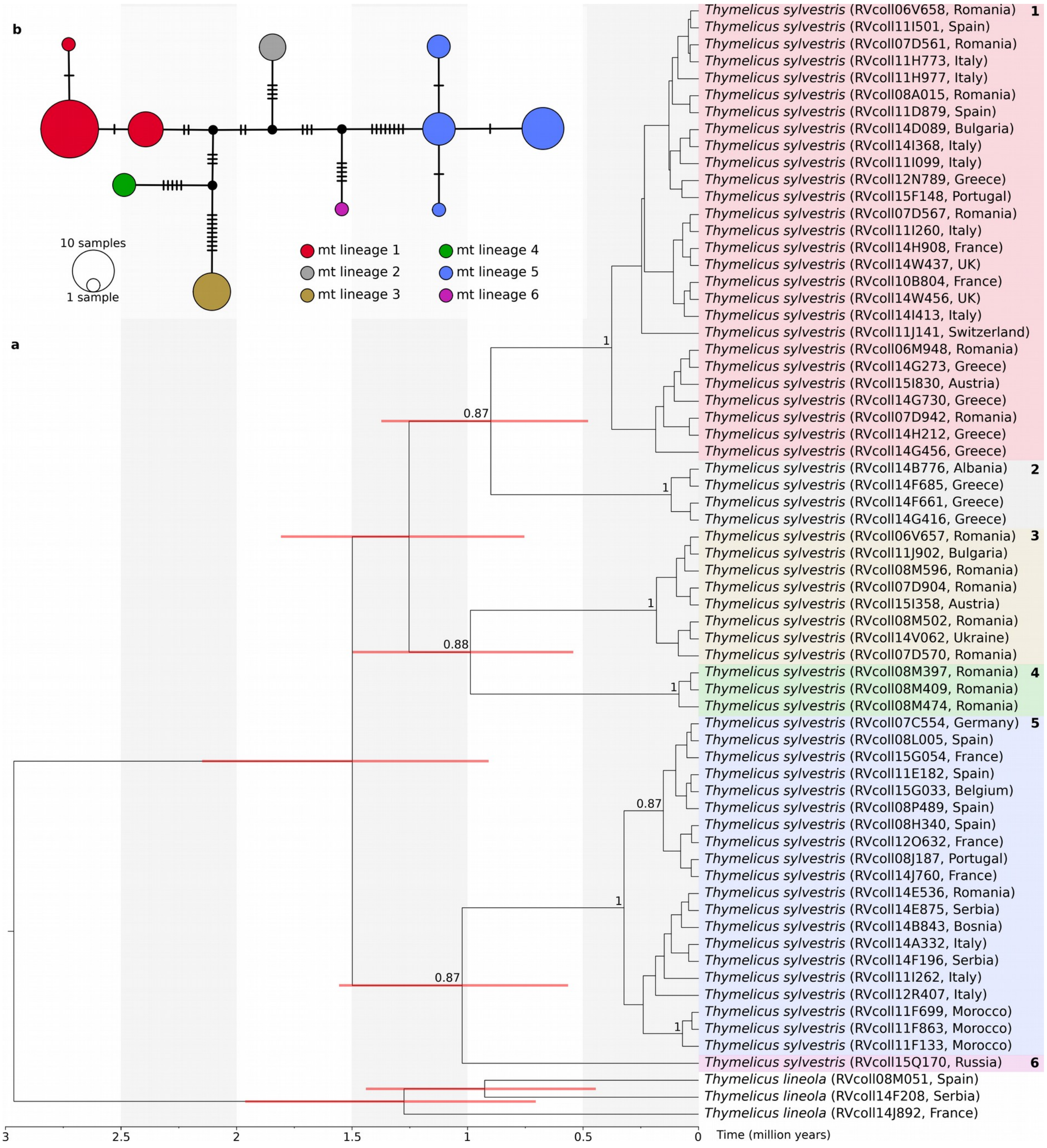
b
10 samples
1 sample
● mt lineage 1
● mt lineage 2
● mt lineage 3
● mt lineage 4
● mt lineage 5
● mt lineage 6
a
1
0.87
0.88
0.87
1
1
1
0.87
1
1
Thymelicus sylvestris (RVcoll06V658, Romania)
Thymelicus sylvestris (RVcoll11I501, Spain)
Thymelicus sylvestris (RVcoll07D561, Romania)
Thymelicus sylvestris (RVcoll11H773, Italy)
Thymelicus sylvestris (RVcoll11H977, Italy)
Thymelicus sylvestris (RVcoll08A015, Romania)
Thymelicus sylvestris (RVcoll11D879, Spain)
Thymelicus sylvestris (RVcoll14D089, Bulgaria)
Thymelicus sylvestris (RVcoll14I368, Italy)
Thymelicus sylvestris (RVcoll11I099, Italy)
Thymelicus sylvestris (RVcoll12N789, Greece)
Thymelicus sylvestris (RVcoll15F148, Portugal)
Thymelicus sylvestris (RVcoll07D567, Romania)
Thymelicus sylvestris (RVcoll11I260, Italy)
Thymelicus sylvestris (RVcoll14H908, France)
Thymelicus sylvestris (RVcoll14W437, UK)
Thymelicus sylvestris (RVcoll10B804, France)
Thymelicus sylvestris (RVcoll14W456, UK)
Thymelicus sylvestris (RVcoll14I413, Italy)
Thymelicus sylvestris (RVcoll11J141, Switzerland)
Thymelicus sylvestris (RVcoll06M948, Romania)
Thymelicus sylvestris (RVcoll14G273, Greece)
Thymelicus sylvestris (RVcoll15I830, Austria)
Thymelicus sylvestris (RVcoll14G730, Greece)
Thymelicus sylvestris (RVcoll07D942, Romania)
Thymelicus sylvestris (RVcoll14H212, Greece)
Thymelicus sylvestris (RVcoll14G456, Greece)
1
Thymelicus sylvestris (RVcoll14B776, Albania)
Thymelicus sylvestris (RVcoll14F685, Greece)
Thymelicus sylvestris (RVcoll14F661, Greece)
Thymelicus sylvestris (RVcoll14G416, Greece)
2
Thymelicus sylvestris (RVcoll06V657, Romania)
Thymelicus sylvestris (RVcoll11J902, Bulgaria)
Thymelicus sylvestris (RVcoll08M596, Romania)
Thymelicus sylvestris (RVcoll07D904, Romania)
Thymelicus sylvestris (RVcoll15I358, Austria)
Thymelicus sylvestris (RVcoll08M502, Romania)
Thymelicus sylvestris (RVcoll14V062, Ukraine)
Thymelicus sylvestris (RVcoll07D570, Romania)
3
Thymelicus sylvestris (RVcoll08M397, Romania)
Thymelicus sylvestris (RVcoll08M409, Romania)
Thymelicus sylvestris (RVcoll08M474, Romania)
4
Thymelicus sylvestris (RVcoll07C554, Germany)
Thymelicus sylvestris (RVcoll08L005, Spain)
Thymelicus sylvestris (RVcoll15G054, France)
Thymelicus sylvestris (RVcoll11E182, Spain)
Thymelicus sylvestris (RVcoll15G033, Belgium)
Thymelicus sylvestris (RVcoll08P489, Spain)
Thymelicus sylvestris (RVcoll08H340, Spain)
Thymelicus sylvestris (RVcoll12O632, France)
Thymelicus sylvestris (RVcoll08J187, Portugal)
Thymelicus sylvestris (RVcoll14J760, France)
Thymelicus sylvestris (RVcoll14E536, Romania)
Thymelicus sylvestris (RVcoll14E875, Serbia)
Thymelicus sylvestris (RVcoll14B843, Bosnia)
Thymelicus sylvestris (RVcoll14A332, Italy)
Thymelicus sylvestris (RVcoll14F196, Serbia)
Thymelicus sylvestris (RVcoll11I262, Italy)
Thymelicus sylvestris (RVcoll12R407, Italy)
Thymelicus sylvestris (RVcoll11F699, Morocco)
Thymelicus sylvestris (RVcoll11F863, Morocco)
Thymelicus sylvestris (RVcoll11F133, Morocco)
5
Thymelicus sylvestris (RVcoll15Q170, Russia)
6
Thymelicus lineola (RVcoll08M051, Spain)
Thymelicus lineola (RVcoll14F208, Serbia)
Thymelicus lineola (RVcoll14J892, France)
3
2.5
2
1.5
1
0.5
0
Time (million years)

Figure 1. a) Bayesian inference chronogram based on the *COI* barcode region with posterior probabilities > 0.70 indicated. The X-axis indicates time in million years and the red bars show the 95% HPD range for the posterior distribution of node ages. **b)** Maximum parsimony haplotype network based on *COI* barcode region. Every mutation is indicated with a bar and circle sizes are proportional to the number of samples represented. Colours correspond to mitochondrial lineages (bPTP entities with posterior probabilities > 0.6 and estimated mean age > 0.5 Mya).

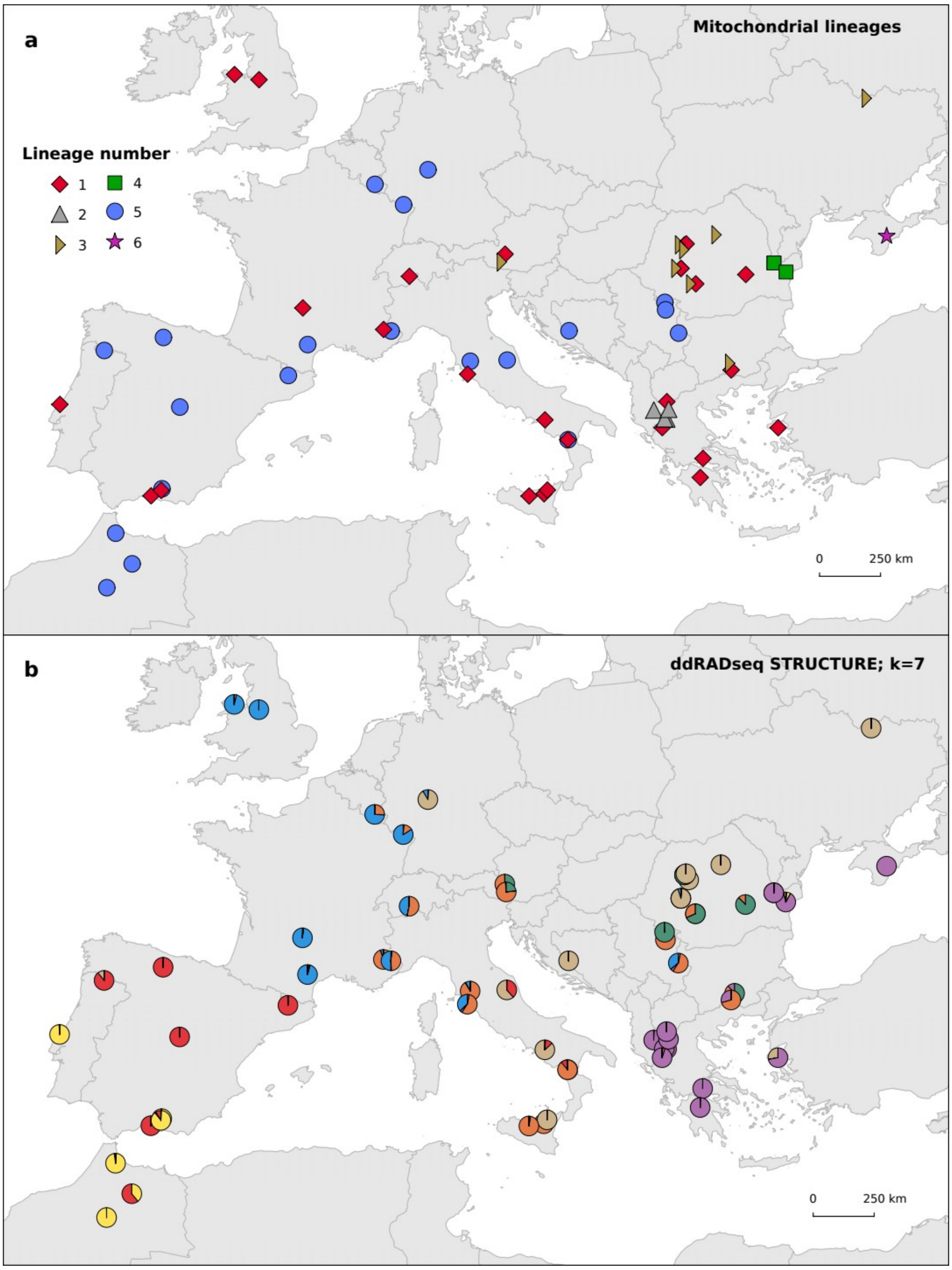

Figure 2. a) Sample distribution and results for the mitochondrial *COI* sequences. Colours and shapes indicate mitochondrial lineages (the same colour code as in Figure 1 and 3 is used). **b)** STRUCTURE results (K = 7) based on 1,360 SNPs plotted on a map. Pie charts represent percentages of SNPs of single individuals attributed to different clusters, as delimited by STRUCTURE. Colours represent clusters defined by STRUCTURE and match those of Figure 3 and Figure S3.

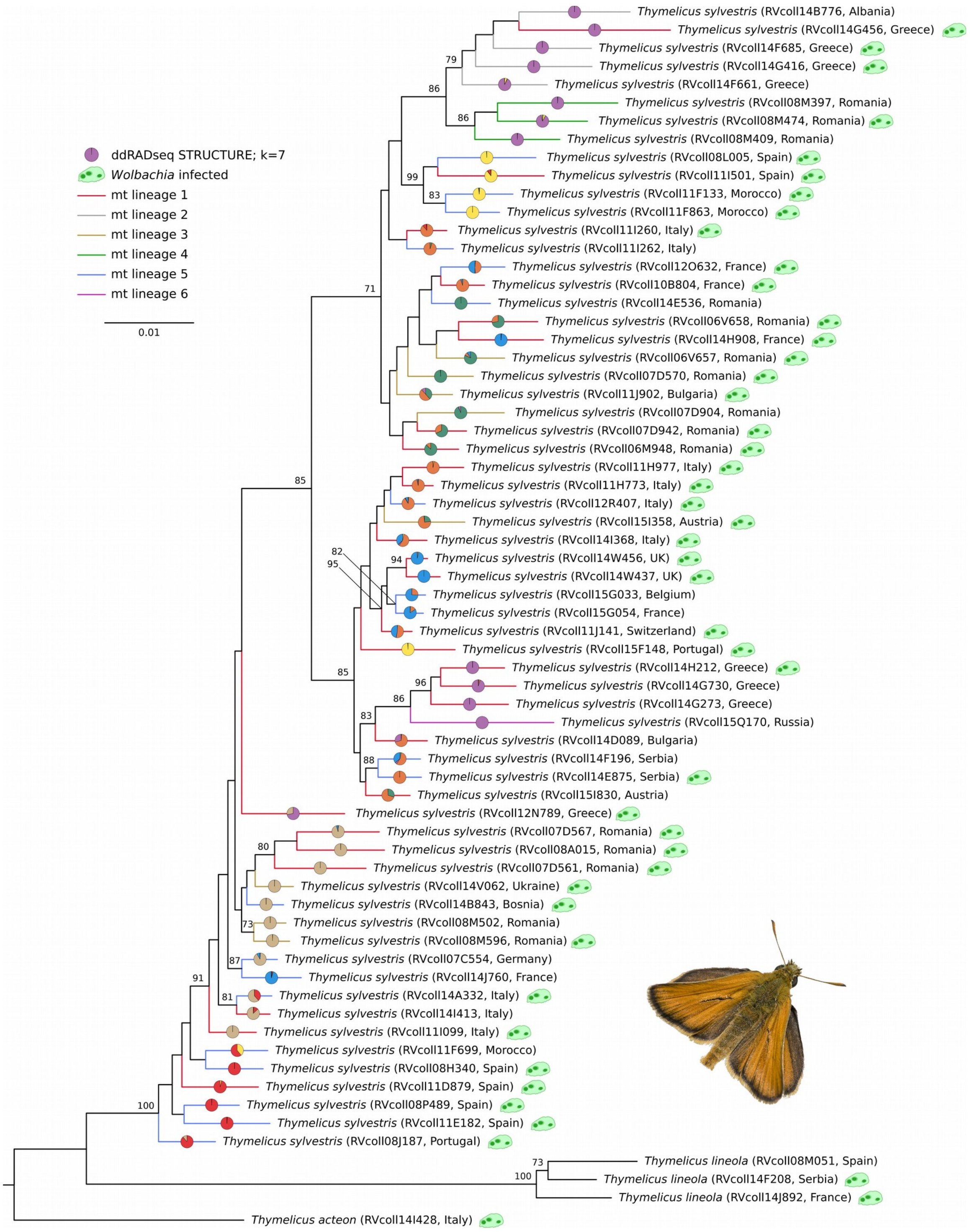

Figure 3. Maximum Likelihood inference tree made with 2,532 ddRADseq loci. Terminal branches are coloured according to the mitochondrial lineage of each specimen (colours match those of Figure 1 and Figure 2). STRUCTURE results, represented as pie charts, are also added to the branches (colours match those of Figure 2 and Figure S3). Individuals for which at least one *Wolbachia* locus was obtained are marked as infected. Scale indicates substitutions per site. The lower right corner illustrates a male *T. sylvestris*. Photo: Vlad Dincă.

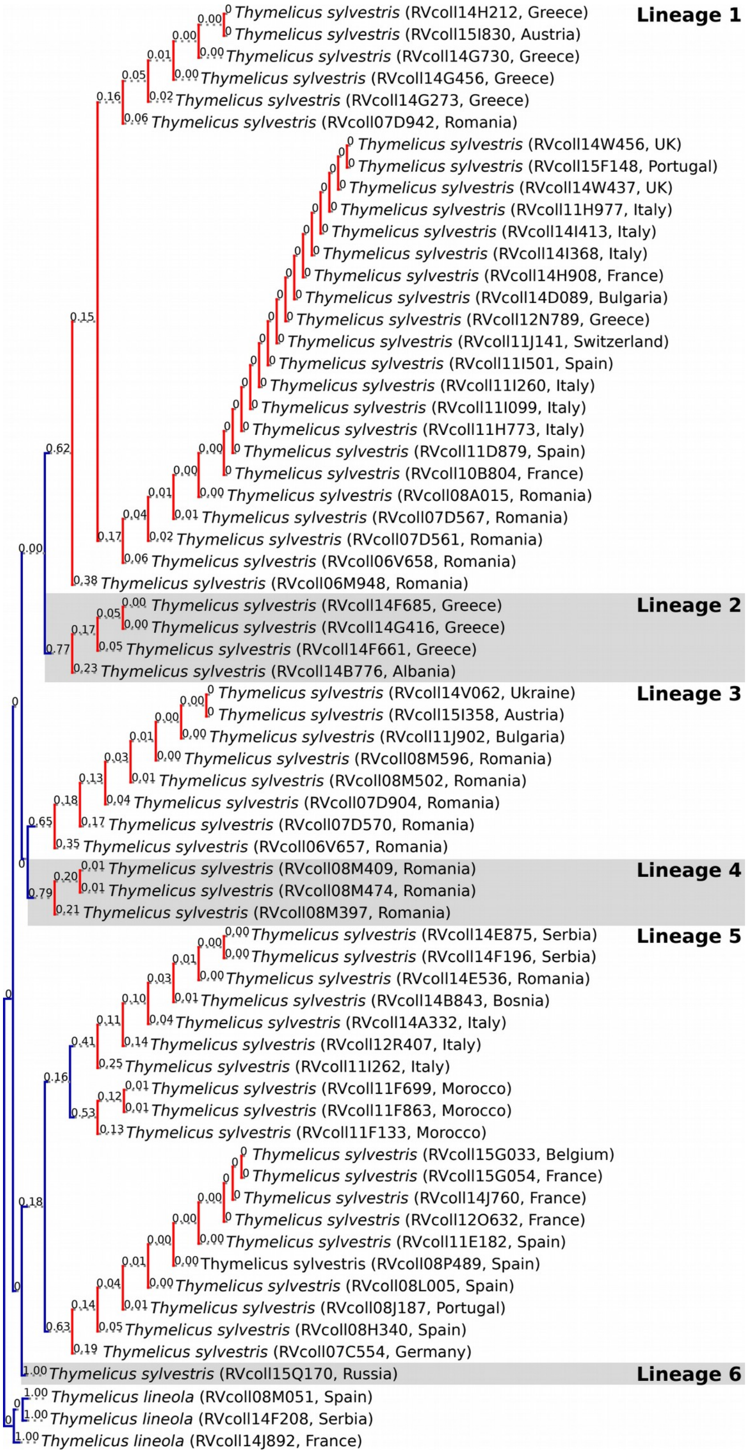
Lineage 1
Thymelicus sylvestris (RVcoll14H212, Greece)
Thymelicus sylvestris (RVcoll15I830, Austria)
Thymelicus sylvestris (RVcoll14G730, Greece)
Thymelicus sylvestris (RVcoll14G456, Greece)
Thymelicus sylvestris (RVcoll14G273, Greece)
Thymelicus sylvestris (RVcoll07D942, Romania)
Thymelicus sylvestris (RVcoll14W456, UK)
Thymelicus sylvestris (RVcoll15F148, Portugal)
Thymelicus sylvestris (RVcoll14W437, UK)
Thymelicus sylvestris (RVcoll11H977, Italy)
Thymelicus sylvestris (RVcoll14I413, Italy)
Thymelicus sylvestris (RVcoll14I368, Italy)
Thymelicus sylvestris (RVcoll14H908, France)
Thymelicus sylvestris (RVcoll14D089, Bulgaria)
Thymelicus sylvestris (RVcoll12N789, Greece)
Thymelicus sylvestris (RVcoll11J141, Switzerland)
Thymelicus sylvestris (RVcoll11I501, Spain)
Thymelicus sylvestris (RVcoll11I260, Italy)
Thymelicus sylvestris (RVcoll11I099, Italy)
Thymelicus sylvestris (RVcoll11H773, Italy)
Thymelicus sylvestris (RVcoll11D879, Spain)
Thymelicus sylvestris (RVcoll10B804, France)
Thymelicus sylvestris (RVcoll08A015, Romania)
Thymelicus sylvestris (RVcoll07D567, Romania)
Thymelicus sylvestris (RVcoll07D561, Romania)
Thymelicus sylvestris (RVcoll06V658, Romania)
Thymelicus sylvestris (RVcoll06M948, Romania)
Lineage 2
Thymelicus sylvestris (RVcoll14F685, Greece)
Thymelicus sylvestris (RVcoll14G416, Greece)
Thymelicus sylvestris (RVcoll14F661, Greece)
Thymelicus sylvestris (RVcoll14B776, Albania)
Lineage 3
Thymelicus sylvestris (RVcoll14V062, Ukraine)
Thymelicus sylvestris (RVcoll15I358, Austria)
Thymelicus sylvestris (RVcoll11J902, Bulgaria)
Thymelicus sylvestris (RVcoll08M596, Romania)
Thymelicus sylvestris (RVcoll08M502, Romania)
Thymelicus sylvestris (RVcoll07D904, Romania)
Thymelicus sylvestris (RVcoll07D570, Romania)
Thymelicus sylvestris (RVcoll06V657, Romania)
Lineage 4
Thymelicus sylvestris (RVcoll08M409, Romania)
Thymelicus sylvestris (RVcoll08M474, Romania)
Thymelicus sylvestris (RVcoll08M397, Romania)
Lineage 5
Thymelicus sylvestris (RVcoll14E875, Serbia)
Thymelicus sylvestris (RVcoll14F196, Serbia)
Thymelicus sylvestris (RVcoll14E536, Romania)
Thymelicus sylvestris (RVcoll14B843, Bosnia)
Thymelicus sylvestris (RVcoll14A332, Italy)
Thymelicus sylvestris (RVcoll12R407, Italy)
Thymelicus sylvestris (RVcoll11I262, Italy)
Thymelicus sylvestris (RVcoll11F699, Morocco)
Thymelicus sylvestris (RVcoll11F863, Morocco)
Thymelicus sylvestris (RVcoll11F133, Morocco)
Thymelicus sylvestris (RVcoll15G033, Belgium)
Thymelicus sylvestris (RVcoll15G054, France)
Thymelicus sylvestris (RVcoll14J760, France)
Thymelicus sylvestris (RVcoll12O632, France)
Thymelicus sylvestris (RVcoll11E182, Spain)
Thymelicus sylvestris (RVcoll08P489, Spain)
Thymelicus sylvestris (RVcoll08L005, Spain)
Thymelicus sylvestris (RVcoll08J187, Portugal)
Thymelicus sylvestris (RVcoll08H340, Spain)
Thymelicus sylvestris (RVcoll07C554, Germany)
Thymelicus sylvestris (RVcoll15Q170, Russia)
Lineage 6
Thymelicus lineola (RVcoll08M051, Spain)
Thymelicus lineola (RVcoll14F208, Serbia)
Thymelicus lineola (RVcoll14J892, France)

Figure S1. Entities estimated with bPTP – clades in red – on a Bayesian *COI* gene-tree with their posterior probability indicated. The mitochondrial lineages considered in this study – those with bPTP posterior probabilities > 0.6 and estimated mean age > 0.5 Mya – are shown.

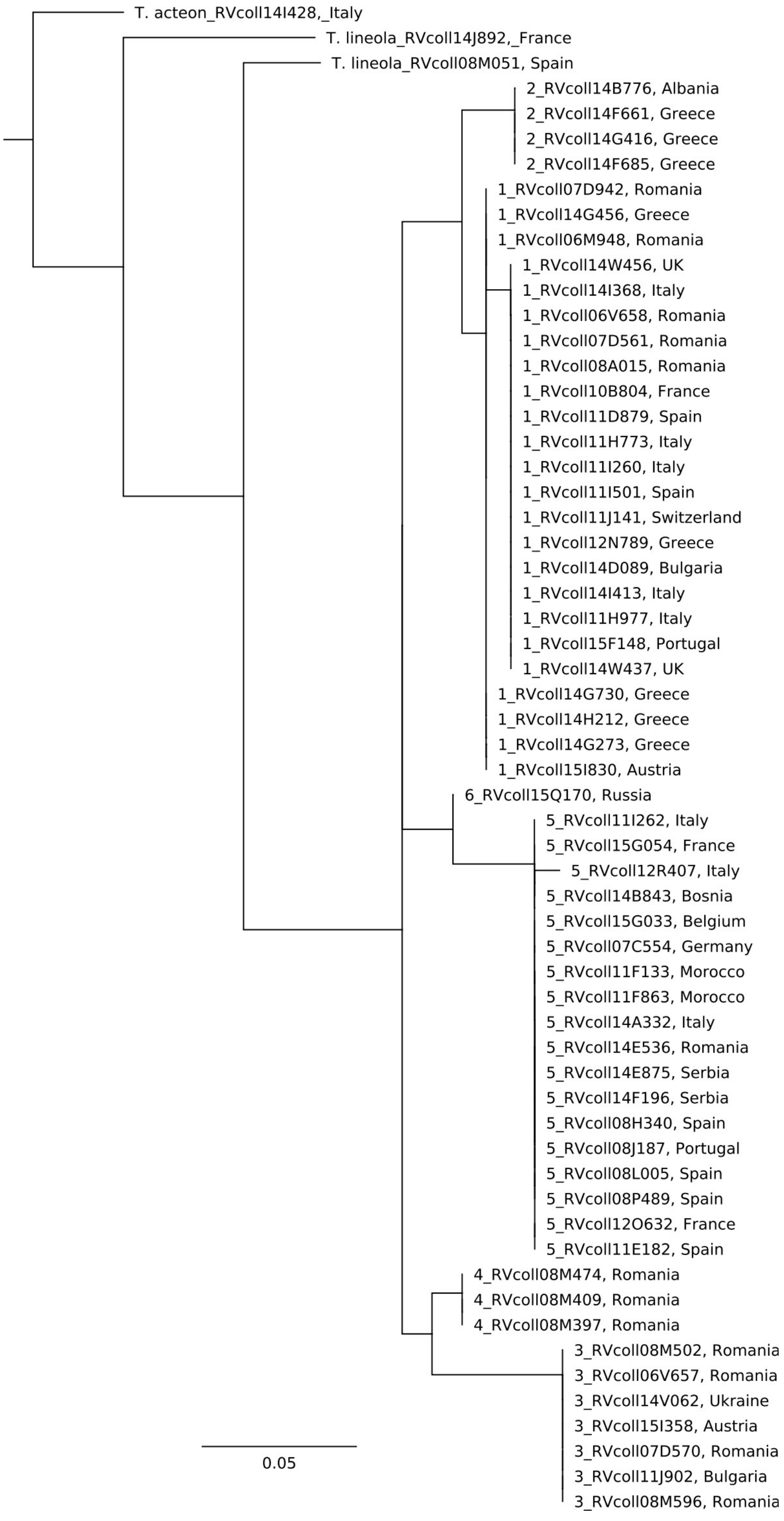

Figure S2. Maximum Likelihood inference tree based on the concatenated fragments of *COI* (87 bp) and *NADH5* (87 bp) genes. Scale indicates substitutions per site. Mitochondrial lineage numbers are written at the beginning of the specimen labels.

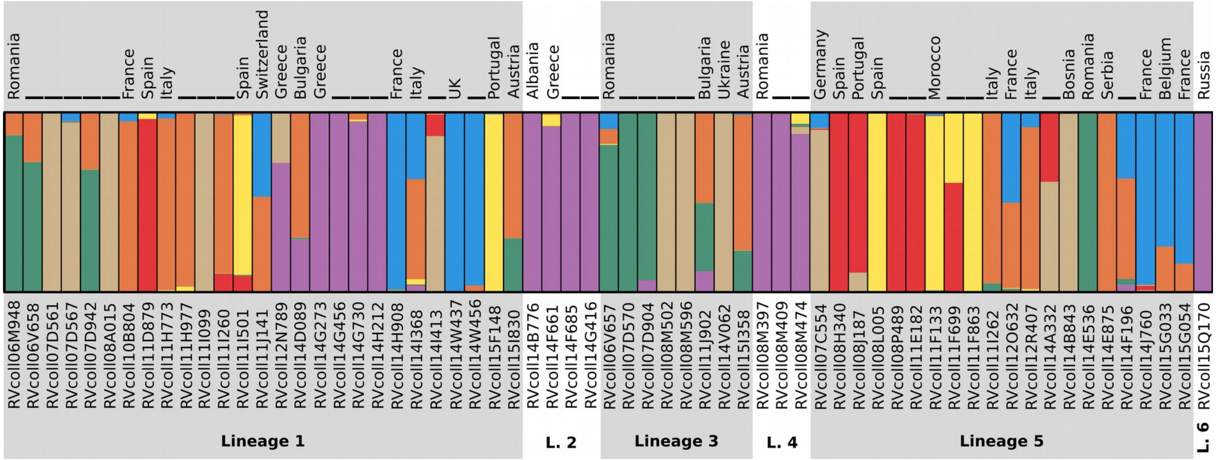

Figure S3. STRUCTURE plot based on 1,360 SNPs, obtained with K = 7. Every colour represents a cluster delimited by STRUCTURE. Samples are ordered based on mitochondrial lineages, which are indicated below.

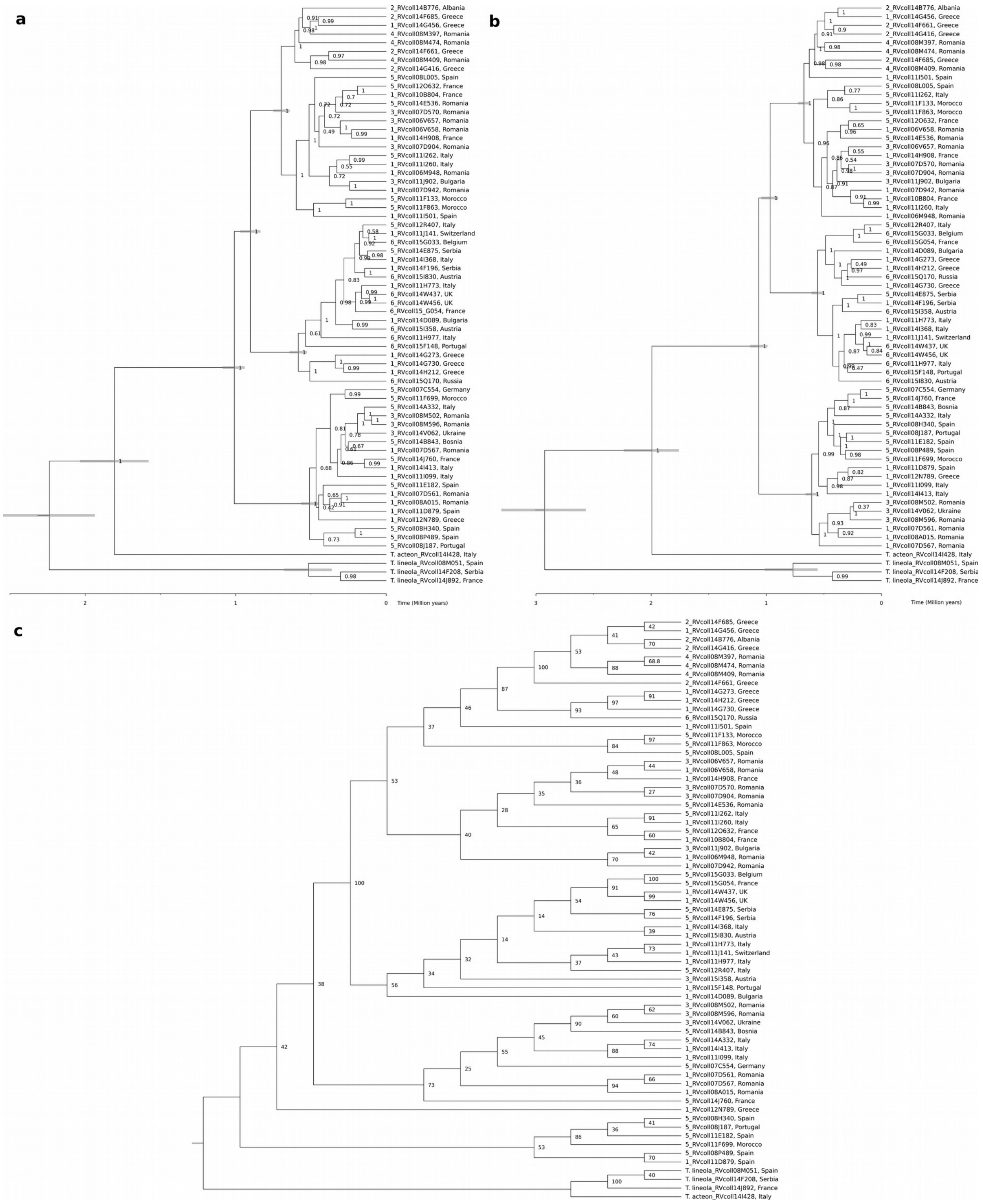

Figure S4. a & b) Bayesian inference phylogeny made with two different random sets of 200 loci with a minimum sample coverage of 30. The X-axis indicates time in million years and bars show the 95%HPD range for the posterior distribution of node ages. **c)** SVDquartets phylogeny made with 2,532 loci. Mitochondrial lineage numbers are written at the tip of the specimen labels in all the figures.

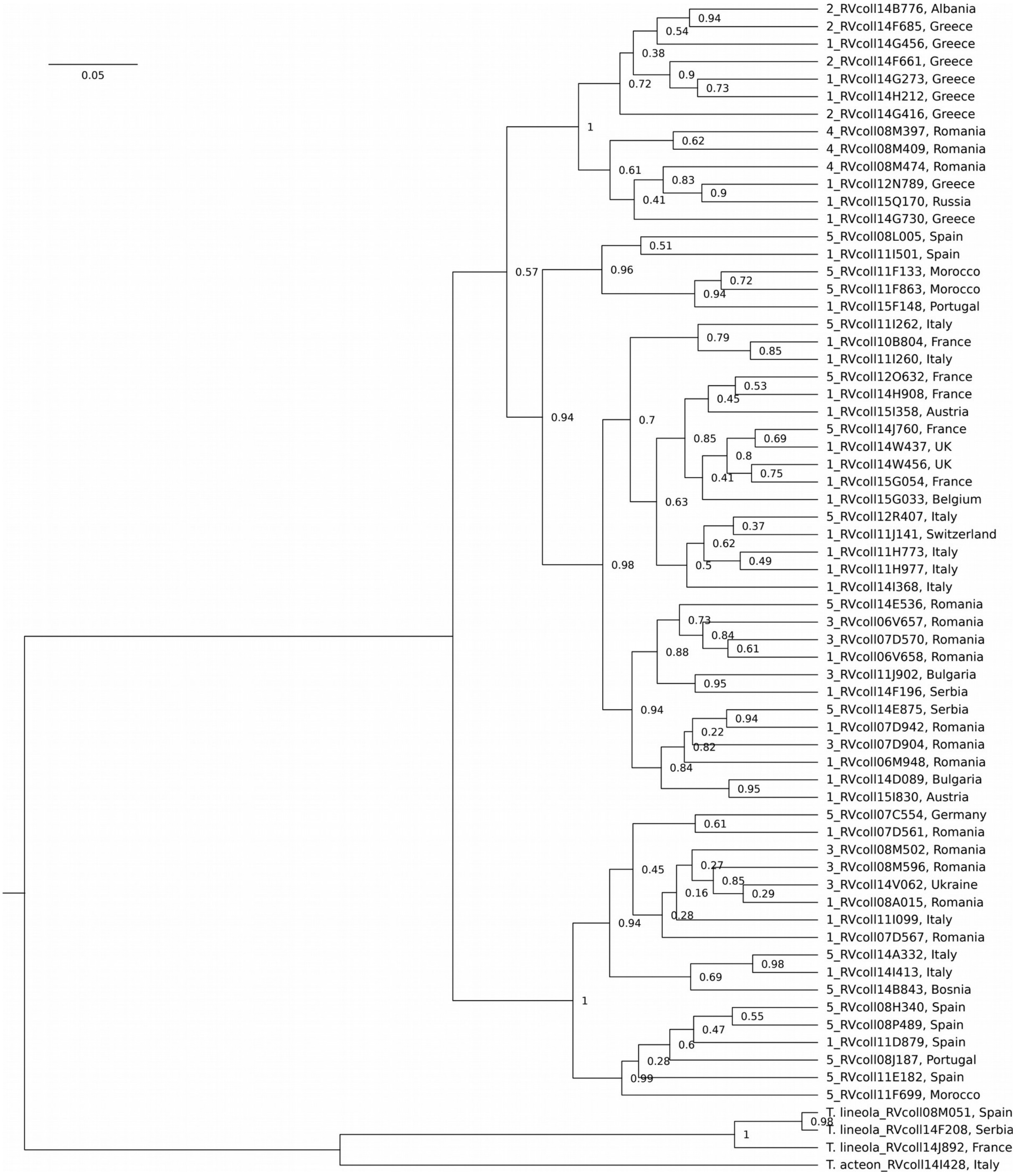

Figure S5. Bayesian inference phylogeny based on 1,360 SNPs. Scale indicates substitutions per site. Mitochondrial lineage numbers are written at the tip of the specimen labels.

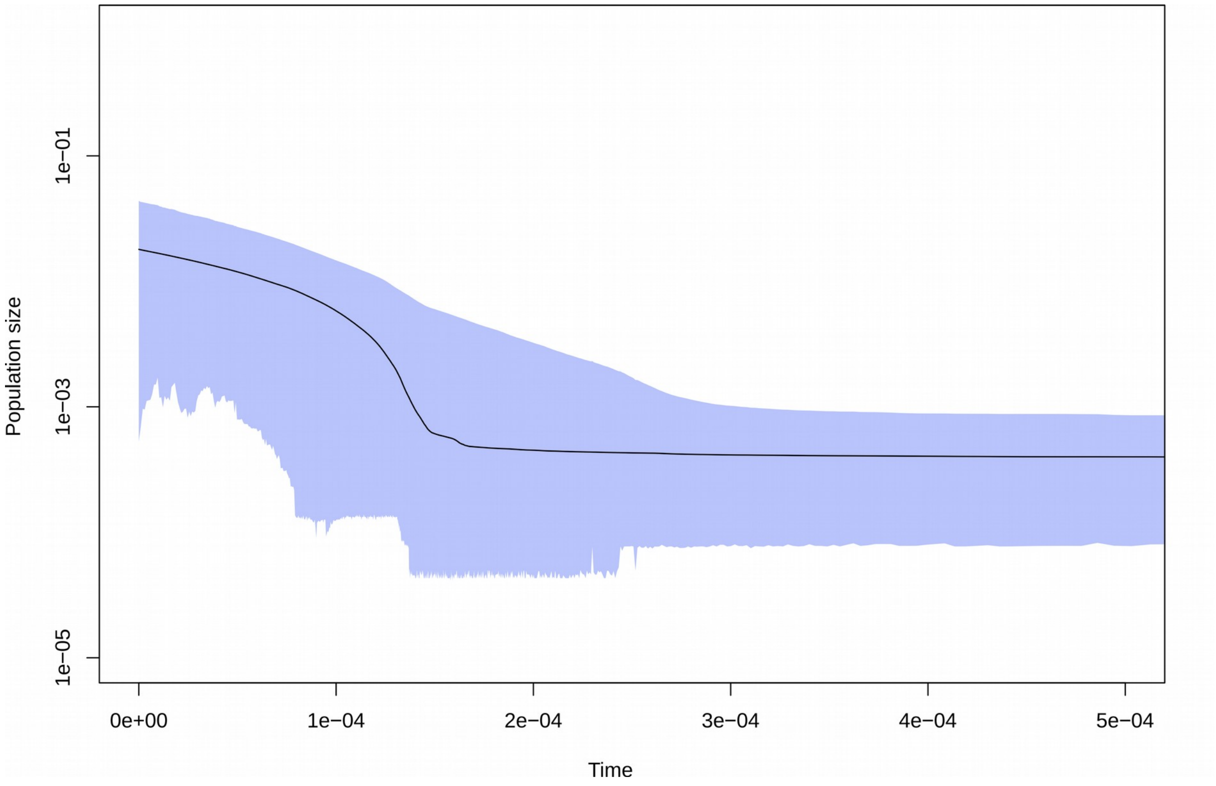

Figure S6. Plotted results of the EBSP analysis based on ddRADseq data. The 19 highest represented loci from a subset of 47 samples were used.

References

Ashfaq, M., Akhtar, S., Khan, A. M., Adamowicz, S. J., & Hebert, P. D. N. (2013). DNA barcode analysis of butterfly species from Pakistan points towards regional endemism. *Molecular Ecology Resources*, 13(5), 832–843. doi: 10.1111/1755-0998.12131

Baguette, M., Vansteenwegen, C., Convi, I., & Nève, G. (1998). Sex-biased density-dependent migration in a metapopulation of the butterfly *Proclossiana eunomia*. *Acta Oecologica*, 19(1), 17–24. doi: 10.1016/S1146-609X(98)80004-0

Bálint, M., Domisch, S., Engelhardt, C. H. M., Haase, P., Lehrian, S., Sauer, J., ... Nowak, C. (2011). Cryptic biodiversity loss linked to global climate change. *Nature Climate Change*, 1(6), 313–318. doi: 10.1038/nclimate1191

Bernardo, P. H., Sánchez-Ramírez, S., Sánchez-Pacheco, S. J., Álvarez-Castañeda, S. T., Aguilera-Miller, E. F., Mendez-de la Cruz, F. R., & Murphy, R. W. (2019). Extreme mito-nuclear discordance in a peninsular lizard: the role of drift, selection, and climate. *Heredity*.doi: 10.1038/s41437-019-0204-4

Bickford, D., Lohman, D.J., Sodhi, N.S., Ng, P.K.L., Meier, R., Winker, K., ... Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology and Evolution*, 22(3), 148–155. doi: 10.1016/j.tree.2006.11.004

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C-H., Xie, D., ... Drummond, A. J. (2014). BEAST 2: a software platform for bayesian evolutionary analysis. *PLoS Computational Biology*, 10(4), e1003537. doi: 10.1371/journal.pcbi.1003537

Brower, A. V. Z. (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 91(14), 6491–6495. doi: 10.1073/pnas.91.14.6491

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2008). BLAST+: architecture and applications. *BMC Bioinformatics*, 10, 421. doi: 10.1186/1471-2105-10-421

Cheviron, Z. A., & Brumfield, R. T. (2009). Migration-selection balance and local adaptation of mitochondrial haplotypes in Rufous-Collared Sparrows (*Zonotrichia capensis*) along an elevational gradient. *Evolution*, 63(6), 1593–1605. doi: 10.1111/j.1558-5646.2009.00644.x

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, 30(23), 3317–3324. doi: 10.1093/bioinformatics/btu530

Clement, M., Snell, Q., Walke, P., Posada, D., & Crandall, K. (2000). TCS: estimating gene genealogies. *Molecular Ecology*, 9(10), 1657–1659. doi: 10.1046/j.1365-294x.2000.01020.x

Collins, R.A., Boykin, L.M., Cruickshank, R.H., & Armstrong, C.F. (2012). Barcoding’s next top model: an evaluation of nucleotide substitution models for specimen identification. *Methods in Ecology and Evolution*, 3(3), 457–465. doi: 10.1111/j.2041-210X.2011.00176.x

Dapporto, L., Vodă, R., Dincă, V., & Vila, R. (2014). Comparing population patterns for genetic and morphological markers with uneven sample sizes: An example for the butterfly *Maniola jurtina*. *Methods in Ecology and Evolution*, 5(8), 834–843. doi: 10.1111/2041-210X.12220

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, 9(8), 772–772. doi: 10.1038/nmeth.2109

Dincă, V., Lukhtanov, V. A., Talavera, G., & Vila, R. (2011). Unexpected layers of cryptic diversity in wood white *Leptidea* butterflies. *Nature Communications*, 2(May), 324. doi: 10.1038/ncomms1329.

Dincă, V., Montagud, S., Talavera, G., Hernández-Roldán, J., Munguira, M. L., García-Barros, E., ... Vila, R. (2015). DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. *Scientific Reports*, 5(January), 12395. doi: 10.1038/srep12395.

Duplouy, A., Hurst, G. D. D., O'Neill, S. L., & Charlat, S. (2010). Rapid spread of male-killing *Wolbachia* in the butterfly *Hypolimnas bolina*. *Evolutionary Biology*, 23(1), 231–235. doi: 10.1111/j.1420-9101.2009.01891.x

Hernández-Roldán, J. L., Dapporto, L., Dincă, V., Vicente, J. C., Hornett, E., Šíchová, J., ... Vila, R. (2016). Integrative analyses unveil speciation linked to host plant shift in *Spialia* butterflies. *Molecular Ecology*, 25(17), 4267–4284. doi: 10.1111/mec.13756

Huemer, P., Mutanen, M., Sefc, K. M. & Hebert, P. D. N. (2014). Testing DNA barcode performance in 1000 species of European Lepidoptera: Large geographic distances have small genetic impacts. *PLoS ONE*, 9(12), 1–21. doi: 10.1371/journal.pone.0115774

Earl, D. A., & vonHoldt, B.M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. doi: 10.1007/s12686-011-9548-7

Eaton, D. A. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, 30(13), 1844–1849. doi: 10.1093/bioinformatics/btu121

Engler, J. O., Balkenhol, N., Filz, K. J., Habel, J. C., & Rödder, D. (2014). Comparative landscape genetics of three closely related sympatric Hesperid butterflies with diverging ecological traits. *PLoS ONE*, 9(9), e106526. doi: 10.1371/journal.pone.0106526

Foll, M., & Gaggiotti, O. E. (2008). A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180(2), 977–993. doi: 10.1534/genetics.108.092221

Galtier, N., Nabholz, B., Glemin, S., & Hurst, G. D. D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, 18(22), 4541–4550. doi: 10.1111/j.1365-294X.2009.04380.x

Guindon, S., Dufayard J. F., Lefort V., Anisimova M., Hordijk W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, 59(3), 307–321. doi: 10.1093/sysbio/syq010

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (41), 14812–14817. doi: 10.1073/pnas.0406166101

Hurst, G. D. D., & Jiggins, F. M. (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society B.*, 272 (1572), 1525–1534. doi: 10.1098/rspb.2005.3056

Irwin, D. E. (2012). Local adaptation along smooth ecological gradients causes phylogeographic breaks and phenotypic clustering. *The American Naturalist*, 180 (1), 35–49. doi: 10.1086/666002

Jiggins, F. M. (2003). Male-killing *Wolbachia* and mitochondrial DNA: selective sweeps, hybrid introgression and parasite population dynamics. *Genetics*, 164(1), 5–12.

Kearns, A. M., Restani, M., Szabo, I., Schrøder-Nielsen, A., Kim, J. A., Richardson, H. M., ... Omland, K. E. (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications*, 9(1), 906. doi: 10.1038/s41467-018-03294-w

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647–1649. doi: 10.1093/bioinformatics/bts199

Kim, D., Song, L., Breitwieser, F. P., & Salzberg, S. L. (2016). Centrifuge: Rapid and sensitive classification of metagenomic sequences. *Genome Research*, 26(12), 1721–1729. doi: 10.1101/gr.210641.116

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, 15(5), 1179–1191. doi: 10.1111/1755-0998.12387

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution*, 33(7), 1870–1874. doi: 10.1093/molbev/msw054

McBride, C. S., Van Velzen, R., & Larsen, T. B. (2009). Allopatric origin of cryptic butterfly species that were discovered feeding on distinct host plants in sympatry. *Molecular Ecology*, 18(17), 3639–3651. doi: 10.1111/j.1365-294X.2009.04309.x

Muñoz, A. G., Baxter, S. W., Linares, M., & Jiggins, C. D. (2011). Deep mitochondrial divergence within a *Heliconius* butterfly species is not explained by cryptic speciation or endosymbiotic bacteria. *BMC Evolutionary Biology*, 11(1), 358. doi: 10.1186/1471-2148-11-358

Nietlisbach, P., Arora, N., Nater, A., Goossens, B., Van Schaik, C. P., & Krützen, M. (2012). Heavily male-biased long-distance dispersal of orang-utans (genus: *Pongo*), as revealed by Y-chromosomal and mitochondrial genetic markers. *Molecular Ecology*, 21(13), 3173–3186. doi: 10.1111/j.1365-294X.2012.05539.x

Papadopoulou, A., Anastasiou, I. & Vogler, A. P. (2010). Revisiting the insect mitochondrial molecular clock: the mid-Aegean trench calibration. *Molecular Biology and Evolution*, 27(7), 1659–1672. doi: 10.1093/molbev/msq051

Pazhenkova E. A., & Lukhtanov, V. A. (2016). Chromosomal and mitochondrial diversity in *Melitaea didyma* complex (Lepidoptera, Nymphalidae), eleven deeply diverged DNA barcode groups in one non-monophyletic species?. *Comparative Cytogenetics*, 10(4), 697–717. doi: 10.3897/CompCytogen.v10i4.11069

Pearman, P. B., D'Amen, M., Graham, C. H., Thuiller, W., & Zimmermann, N. E. (2010). Within-taxon niche structure: niche conservatism, divergence and predicted effects of climate change. *Ecography*, 33(6), 990–1003. doi: 10.1111/j.1600-0587.2010.06443.x

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: An inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, 7(5), e37135. doi: 10.1371/journal.pone.0037135

Petit, S., Moilanen, A., Hanski, I., & Baguette, M. (2001). Metapopulation dynamics of the fritillary butterfly: movements between patches. *Oikos*, 92(3), 491–500. doi: 10.1034/j.1600-0706.2001.920310.x

Quek, S. P., Davies, S. J., Itino, T., & Pierce, N. E. (2004). Codiversification in an ant-plant mutualism: stem texture and the evolution of host use in *Crematogaster* (Formicidae: Myrmicinae) inhabitants of *Macaranga* (Euphorbiaceae). *Evolution*, 58(3), 554–570. doi: 10.1111/j.0014-3820.2004.tb01678.x

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. (2018). Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*. 67(5), 901–904. doi: 10.1093/sysbio/syy032

Ribeiro, L. (2012). Mitochondrial pseudogenes in insect DNA barcoding: differing points of view on the same issue. *Biota Neotropica*, 12(3), 301–308. doi: 10.1590/S1676-06032012000300029

Ritter, S., Michalski, S. G., Settele, J., Wiemers, M., Fric, Z.F., Sielezniew, M., ... Durka, W. (2013). *Wolbachia* infections mimic cryptic speciation in two parasitic butterfly species, *Phengaris teleius* and *P. nausithous* (Lepidoptera: Lycaenidae). *PLoS ONE*, 8(11), e78107. doi: 10.1371/journal.pone.0078107

Rosenberg, N. A. (2004). DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, 4(1), 137–138. doi: 10.1046/j.1471-8286.2003.00566.x

Shapoval N. A., & Lukhtanov V. A. (2015). Intragenomic variations of multicopy *ITS2* marker in *Agrodiaetus* blue butterflies (Lepidoptera, Lycaenidae). *Comparative Cytogenetics*, 9(4), 483–497. doi: 10.3897/CompCytogen.v9i4.5429

Song, H., Buhay, J. E., Whiting, M. F., & Crandall, K. A. (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 of the United States of America*, 105(36), 13486–13491. doi: 10.1073/pnas.0803076105

Srivathsan, A., & Meier, R. (2012). On the inappropriate use of Kimura-2-parameter (K2P) divergences in the DNA-barcoding literature. *Cladistics*, 28(2), 190–194. doi: 10.1111/j.1096-0031.2011.00370.x

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, 30(9), 1312–1313. doi: 10.1093/bioinformatics/btu033

Struck, T. H., Feder, J. L., Bendiksby, M., Birkeland, S., Cerca, J., Gusarov, V.I., ... Dimitrov, D. (2018). Finding evolutionary processes hidden in cryptic species. *Trends in Ecology and Evolution*, 33(3), 153–163. doi: 10.1016/j.tree.2017.11.007

Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 21(16), 3907–3930. doi: 10.1111/j.1365-294X.2012.05664.x

Tolman, T., & Lewington, R. (2008). Collins butterfly guide: The most complete field guide to the butterflies of Britain and Europe. London, UK: HarperCollins Publishers.

Trontelj, P., & Fišer, C. (2009). Cryptic species diversity should not be trivialised. *Systematics and Biodiversity*, 7(1), 1-3. doi: 10.1017/S1477200008002909

Vodă, R., Dapporto, L., Dincă, V., & Vila, R. (2015a). Why do cryptic species tend not to co-occur? A case study on two cryptic pairs of butterflies. *PLoS ONE*, 10(2), 1–18. doi: 10.1371/journal.pone.0117802

Vodă, R., Dapporto, L., Dincă, V., & Vila, R. (2015b). Cryptic matters: overlooked species generate most butterfly beta-diversity. *Ecography*, 38(4), 405–409. doi: 10.1111/ecog.00762

Von Helversen, O., Heller, K. G., Mayer, F., Nemeth, A., Volleth, M., & Gombkötö, P. (2001). Cryptic mammalian species: A new species of whiskered bat (*Myotis alcathoe* n. sp.) in Europe. *Naturwissenschaften*, 88(5), 217–223. doi: 10.1007/s001140100225

Werren, J. H., Baldo, L., & Clark, M. E. (2008). *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, 6(10), 741–751. doi: 10.1038/nrmicro1969

Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics*, 29(22), 2869–2876. doi: 10.1093/bioinformatics/btt499

Data accsessibility

COI sequences were deposited in Genbank under de following codes: XXX. Raw ddRADseq reads can be found in the Bioproject XXX. *Wolbachia* loci alignment was uploaded to Dryad (code: XXX).

Author contributions

N. Alvarez, V. Dincă, J.C. Hinojosa and R. Vila designed the study. Funding was secured by R. Vila and N. Álvarez. Laboratory protocols were done by J.C. Hinojosa and C. Pitteloud. J.C. Hinojosa, D. Koubínová and M. Szenteczki perfomed the data analyses. All authors participated in writing the manuscript.

ORCID

J. C. Hinojosa: https://orcid.org/0000-0002-6318-4252