Apidologie (2019) 50:215–229

DOI: 10.1007/s13592-019-00632-9

Original article

From the popular tRNA^{leu}-COX2 intergenic region to the mitogenome: insights from diverse honey bee populations of Europe and North Africa

Dora Henriques[1], Júlio Chávez-Galarza[1,2], Andreia Quaresma[1], Cátia José Neves[1], Ana Rita Lopes[1], Cecília Costa[3], Filipe O. Costa[4], José Rufino[5], Maria Alice Pinto[1]

[1]Centro de Investigação de Montanha (CIMO), Instituto Politécnico de Bragança, Campus de Sta. Apolónia, 5300-253, Bragança, Portugal
[2]Instituto Nacional de Innovación Agraria (INIA), Av. La Molina 1981, La Molina, Lima, Peru
[3]Consiglio per la Ricerca in Agricoltura e l'Analisi dell'Economia Agraria - Centro di Ricerca Agricoltura e Ambiente (CREA-AA), Via di Saliceto 80, 40128, Bologna, Italy
[4]Centre of Molecular and Environmental Biology (CBMA), Department of Biology, University of Minho, Campus de Gualtar, 4710-057, Braga, Portugal
[5]Research Centre in Digitalization and Intelligent Robotics (CeDRI), Polytechnic Institute of Bragança, Campus de Sta. Apolónia, 5300-253, Bragança, Portugal

Received 26 July 2018 – Revised 7 December 2018 – Accepted 10 January 2019

Abstract – The tRNA^{leu}-COX2 intergenic region of the mitochondrial DNA has been used for assessing diversity in honey bee (*Apis mellifera* L.) populations worldwide. However, differential mutation rates in different partitions of the mitogenome may produce incongruent results. In this study, we sequenced 123 mitogenomes of 7 subspecies from lineages A, M, and C. This allowed generating a comprehensive dataset to investigate the phylogenetic and phylogeographic congruence among the mitogenome, individual genes, and the tRNA^{leu}-COX2 region. We showed that the diversity patterns inferred from the tRNA^{leu}-COX2 marker are not fully paralleled by those obtained with the mitogenome and the individual genes; while the three lineages are supported by these, the African sub-lineages and the haplotypes are not. Thus, conclusions drawn from the tRNA^{leu}-COX2 region need to be taken with caution and this marker may not be appropriate to infer phylogenetic relationships between honey bee colonies.

Iberian honey bee / tRNA^{leu}-COX2 intergenic region / mitogenome

1. INTRODUCTION

Since publication in the 1980s of John Avise's seminal paper (Avise 1986) supporting the power of mitochondrial (mtDNA) markers in phylogeographic inference, thousands of studies on thousands of organisms have been undertaken using short segments of the cytoplasmatic DNA (reviewed by DeSalle et al. 2017). While these short segments can provide phylogeographical patterns, they may have limited resolution to solve very recent events (Jacobsen et al. 2012). Next-generation sequencing has made the sequencing of the whole mitochondrial genome (mitogenome) possible in a time- and cost-effective manner. The mitogenome has a higher resolution than single genes and has been examined in a variety of phylogenetic and phylogeographic studies to solve shallow evolutionary histories, such as in the dog (Pang et al. 2009), brown bear (Keis et al. 2013), bank vole (Filipi et al. 2015), and snub-

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13592-019-00632-9) contains supplementary material, which is available to authorized users.

Corresponding author: M. Pinto, apinto@ipb.ptA. Pinto, apinto@ipb.pt
Handling Editor : Klaus Hartfelder

nosed monkey (Hong et al. 2017), as well as many other organisms.

In 1985, the *Drosophila yakuba* mitogenome was first released (Clary and Wolstenholme 1985) and since then other insect mitogenomes have been sequenced (Simon and Hadrys 2013), including that of the Italian honey bee, *Apis mellifera ligustica* (Crozier and Crozier 1993). Like in most of the Metazoa, the honey bee mitogenome contains 37 genes, namely, two rDNAs that encode the mitochondrial ribosome RNA components, 22 tRNAs required for translation of the mitochondrial proteins, and 13 protein-coding genes encoding the essential components of the electron transport chain and oxidative phosphorylation.

Mitochondrial DNA diversity has been surveyed in honey bee populations around the world with a variety of molecular methods, including RFLPs (restriction fragment length polymorphisms), PCR-RFLPs, and direct sequencing of coding regions (Arias and Sheppard 1996; Chávez-Galarza et al. 2017; Techer et al. 2017). The early studies examined mtDNA variation by digesting the entire molecule with a combination of restriction enzymes, allowing for the construction of restriction site maps (Garnery et al. 1992; Smith and Brown 1988; Smith et al. 1991). Later on, restriction enzymes were employed in PCR-amplified coding regions of the mtDNA such as CYTB (cytochrome b; Crozier et al. 1991; Pinto et al. 2003), COX1 (cytochrome c oxidase subunit I; Bouga et al. 2005; Hall and Smith 1991; Nielsen et al. 2000; Stevanovic et al. 2010), ls-rRNA (large subunit ribosomal RNA; Hall and Smith 1991; Ozdil and Ilhan 2012a; Pinto et al. 2005), and ND5 (NADH dehydrogenase 5; Bouga et al. 2005; Ozdil and Ilhan 2012b). Among the different PCR-RFLP assays, the most popular is the *Dra* I test developed by Garnery et al. (1993). This assay consists of PCR amplification of the highly polymorphic non-coding region located between the $tRNA^{leu}$ and COX2 (cytochrome c oxidase subunit II) genes (originally named COI-COII intergenic region), followed by digestion with *Dra* I restriction enzyme. Garnery et al. (1993) showed that the combined length and restriction site polymorphisms of this region were able to assign honey bee populations to the morphology-based lineages (A, African; M, Western European; C, Eastern European) previously described by Ruttner (1988). Later on, screening of African and Middle Eastern populations led to proposal of a novel mtDNA lineage, dubbed Y, and to partitioning of lineage A variation into sub-lineages A_I, A_{II}, A_{III}, and Z (Alburaki et al. 2011; Franck et al. 2001).

In addition to being able to differentiate lineages and African sub-lineages, the *Dra* I test has several advantages: haplotype identification requires basic equipment available in any small-scale laboratory, it can be done in a time- and cost-effective manner, and there is a large catalog of haplotypes developed from honey bees collected across the world, which allows comparative studies (Chávez-Galarza et al. 2017; Rortais et al. 2011; Techer et al. 2017). On the other hand, the data produced by the *Dra* I test are not well suited to phylogenetic and phylogeographical inference due to the large number of indels, the duplication of large fragments in the $tRNA^{leu}$-COX2 region, and the uncertainty about whether this fragment captures the "real" maternal history.

The honey bees from the Iberian Peninsula have been intensively surveyed for mtDNA diversity with the *Dra* I test (Cánovas et al. 2008; Franck et al. 1998; Miguel et al. 2007; Pinto et al. 2013; Pinto et al. 2012). These studies revealed maternal patterns of unparalleled complexity, characterized by geographically structured co-existence of haplotypes belonging to lineages M and A, and to sub-lineages A_I, A_{II}, and A_{III}. In addition, Iberia harbors the highest levels of diversity reported so far, with hundreds of haplotypes (Cánovas et al. 2008; Chávez-Galarza et al. 2017; Franck et al. 1998; Miguel et al. 2007; Pinto et al. 2013). Therefore, this confined area of the *A. mellifera* distributional range offers a unique stage for studying honey bee genetic diversity patterns.

Chávez-Galarza et al. (2017) carried out a comprehensive survey of maternal variation in Iberia by sequencing the $tRNA^{leu}$-COX2 fragment of 711 individuals collected across three north-south transects. Herein, we generated sequencing data of the complete mitogenome for 87 individuals selected from the 711 individuals to represent the Iberian variation range and for 36 individuals from six subspecies belonging to lineages A, M, and C. Using this large dataset, we aim to investigate the phylogenetic and phylogeographic

congruence among the mitogenome, the individual coding genes, and the tRNA^{leu}-COX2 region. Specifically, we addressed the following questions: (i) Do individual coding genes provide the same topology as the mitogenome? (ii) Do the mitogenome and the individual genes support the three lineages occurring in Europe? (iii) Do the mitogenome and the individual genes support the three African sub-lineages that have been identified from the tRNA^{leu}-COX2 region? (iv) Are the diversity patterns inferred from the mitogenome, each individual gene, and the tRNA^{leu}-COX2 intergenic region concordant?

2. METHODS

2.1. Sampling and DNA extraction

A total of 123 individuals, from seven subspecies, three lineages (A, M, and C), and three African sub-lineages (A_I, A_{II}, and A_{III}), were sampled across a wide area ranging from North Africa to North Europe (Figure1). The *Apis mellifera iberiensis* collection comprised 87 individuals from 16 sampling sites distributed throughout three north-south transects in Iberia: one along the Atlantic coast (31 individuals), one along the center (33 individuals), and another along the Mediterranean coast (23 individuals), as detailed in Chávez-Galarza et al. (2017). Each individual represents a single colony and apiary. The remaining 36 individuals represent six subspecies, including eight *A. m. mellifera* (Denmark, France, the Netherlands, Norway and Scotland), seven *A. m. sahariensis* (Algeria and Morocco), 12 *A. m. intermissa* (Algeria and Morocco), three *A. m. carnica* (Croatia and Serbia), four *A. m. ligustica* (Italy), and two *A. m. siciliana* (Sicily; Figure 1 and Table S1).

Samples were taken from inside the hives, placed in absolute ethanol, and then stored at − 20°C until DNA extraction. Genomic DNA was extracted from the thorax of the 123 individuals using a phenol/chloroform isoamyl alcohol (25:24:1) protocol (Sambrook et al. 1989).

2.2. tRNA^{leu}-COX2 intergenic region

The tRNA^{leu}-COX2-based haplotypes of 114 individuals (excluding the *A. m. sahariensis* and *A. m. siciliana* individuals) were obtained from Chávez-Galarza et al. (2017). The haplotypes of the remaining 9 individuals of *A. m. sahariensis* and *A. m. siciliana* were obtained by PCR amplification with the E2 and H2 primers using reaction and conditions established by Garnery et al. (1993), with slight modifications. PCR products were sent to STABVIDA Inc. (Oeiras, Portugal) for direct sequencing in both directions using the Sanger method. The alignment of the tRNA^{leu}-COX2 sequences was performed manually with MEGA 6.06 (Tamura et al. 2013) and the haplotypes were named following the nomenclature revised by Chávez-Galarza et al. (2017). GENEALEX 6.50 (Peakall and Smouse 2006) was used to estimate the mean number of haplotypes (Na), effective number of haplotypes (Ne), number of private haplotypes (Np), and unbiased diversity (uh) for the populations shown in Table I.

2.3. Mitogenome sequencing and filtering

Whole genome sequencing (WGS) was accomplished using the Illumina HiSeq 2500 platform with the aim of producing a mean coverage $\geq 10\times$. The 2×150 paired-end sequence reads were mapped against the reference honey bee genome Amel_4.5 (for further details, see Henriques et al. 2018). Here, we only used the reads mapped against the mitochondrial genome. Following SNP calling, several filters were applied to reduce poor mapping and spurious positions; SNPs, (i) with only one allele (monomorphic), (ii) with more than two alleles, (iii) with a quality score < 50, (iv) and called in ≤ 112 individuals, were removed from the dataset (Table S2). MtDNA data was intentionally misspecified to be diploid in the SNP calling process. The heterozygous positions were removed as they represented spurious SNPs (Table S3). Some problematic sites were identified by more than one filter. Of the 795 identified SNPs, only 645 were kept for further analysis. The assembled mitogenomes consisted of the nearly complete mtDNA; there was only a short fragment missing positioned within the tRNA^{leu}-COX2 region, corresponding to the P- and Q-element indels that characterize the

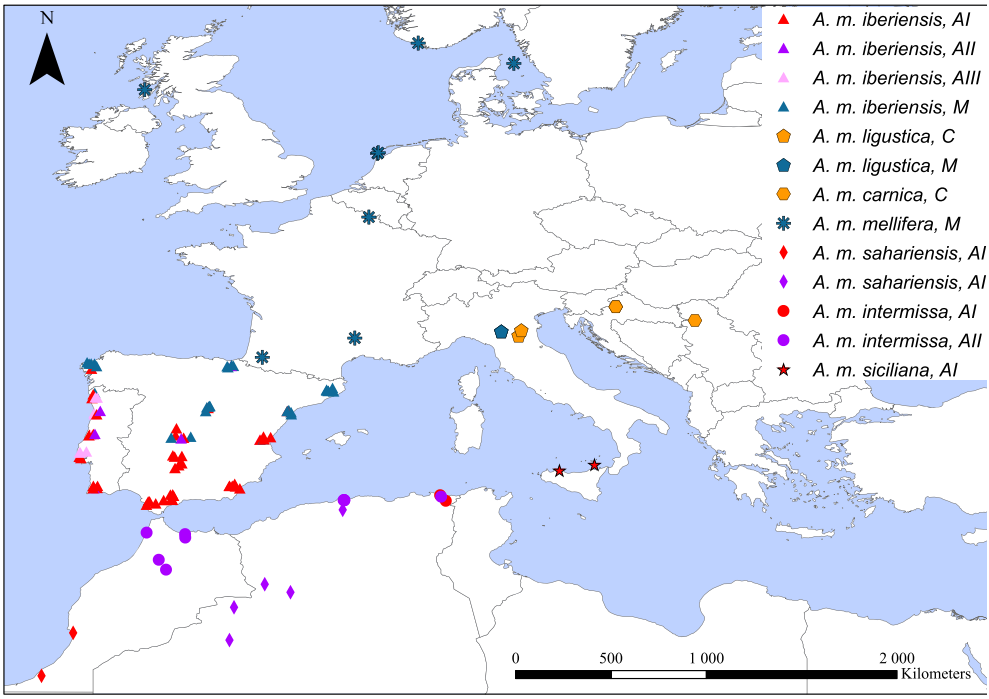

Figure 1 Geographical location of *A. m. iberiensis* ($N = 87$), *A. m. ligustica* ($N = 4$), *A. m. carnica* ($N = 3$), *A. m. mellifera* ($N = 8$), *A. m. sahariensis* ($N = 7$), *A. m. intermissa* ($N = 12$), and *A. m. siciliana* ($N = 2$). Each point represents a colony and the color/symbol codes for different subspecies, lineage (M or C), or African sub-lineage (A_I, A_{II}, or A_{III}), as determined by the $tRNA^{leu}$-COX2 variation.

reference *A. m. ligustica* mitogenome (Crozier and Crozier 1993). The mitogenome was partitioned into 15 datasets corresponding to the 13 protein-coding and 2 rRNA genes for comparisons.

2.4. Population and phylogenetic analyses

Summary statistics, including nucleotide diversity (π), haplotype diversity (H), number of haplotypes (Hd), and the number of polymorphic sites,

Table I. Diversity measures estimated from the $tRNA^{leu}$-COX2 sequence.

Lineage	Subspecies	N^a	Na^b	Private haplotypes	Ne ^{c}	uh^d
M	*A. m. iberiensis*	28	8	6	3.04	0.70
	A. m. mellifera	8	3	2	1.68	0.46
	A. m. ligustica	2	2	0	2.00	1.00
A	*A. m. iberiensis*	59	18	13	7.09	0.87
	A. m. intermissa	12	4	0	3.13	0.74
	A. m. sahariensis	7	3	0	2.33	0.67
	A. m. siciliana	2	2	1	2.00	1.00
C	*A. m. ligustica*	2	2	0	2.00	1.00
	A. m. carnica	3	2	0	1.80	0.67

[a] Number of individuals; [b] mean number of haplotypes; [c] number of effective haplotypes; [d] unbiased haplotype diversity

were calculated for each of the 16 datasets (mitogenome and individual genes) using DnaSP 5.10 (Rozas et al. 2003). Additional molecular indexes, such as number of transitions, transversions, and private substitution sites, were calculated using Arlequin 3.5.1.2 (Excoffier et al. 2005). The individuals were grouped by subspecies, and when subspecies carried haplotypes belonging to different lineages, the individuals were further grouped by lineage. The average number of pairwise differences between populations (π XY) and within populations (π X) was calculated by Arlequin. To avoid the bias introduced when sample sizes are small, diversity estimates for populations represented by < 5 individuals were not compared.

Phylogenetic relationships between haplotypes obtained from the mitogenome and the 15 genes were inferred using Median-Joining network analysis (Bandelt et al. 1999) in PopART (http://popart.otago.ac.nz), and the Bayesian phylogenetic analysis was performed using MrBayes 3.2.6 (Ronquist and Huelsenbeck 2003). The best-fitting evolutionary model for each gene and for the mitogenome was calculated according to the Bayesian information criterion (BIC) using the PartitionFinder v1.1.1 (Lanfear et al. 2012). The selected models were run on MrBayes 3.2.6 in the CIPRES platform (Miller et al. 2010). Four Bayesian independent runs with random starting trees were performed and posterior distributions of parameters, including the tree, were estimated using Markov chain Monte Carlo (MCMC) sampling. Samples were drawn every 100 MCMC steps over a total of 50,000,000 steps, with the first 125,000 trees discarded as burn-in. The homologous mitochondrial sequence of *Apis cerana* (GI:299829158), the sister species of *A. mellifera* , was used as an outgroup.

2.5. Dataset comparisons

Two different methods were used to delimitate intraspecific groups from the mitogenome and each of the 15 genes. In one method, the aligned sequences were analyzed by Automatic Barcode Gap Discovery (ABGD; Puillandre et al. 2012) with the default settings using either the Kimura-2-Parameter (K2P; Kimura 1980) or Jukes-Cantor (JC; Jukes et al. 1969) distance metrics. In the other method, the aligned sequences were analyzed by Bayesian Poisson Tree Processes (bPTP; Zhang et al. 2013), which delimits groups based on the phylogenetic species concept. We used the Bayesian non-ultrametric phylograms obtained by MrBayes as input, which was submitted to the Exelixis Lab web-server (http://species.h-its.org/ptp/). The bPTP analysis was run for 500 ,000 MCMC generations, with a thinning value of 100 and a burn-in of 25%.

The phylogenetic topologies generated by MrBayes 3.2.6 from each dataset were compared using the APE package in R (R Development Core Team 2014; Paradis et al. 2004) with the PH85 distance (Penny and Hendy 1985). The pairwise tree distances were then used by Past 3.08 (Hammer et al. 2001) to create a neighbor-joining (NJ) dendrogram, which represents gene groupings by topology similarities.

3. RESULTS

3.1. Distribution of SNPs in the mitogenome

The sequencing coverage of each individual ranged from 2523 to 7758× (Table S1). A total of 795 SNPs were identified in the 16,343-bp reference mitochondrial genome, of which 150 did not pass one or more filtering criteria (Table S2). The remaining 645 SNPs were distributed across all genes with ls-rRNA containing the lowest proportion (2.41%) and ATP8 (ATP synthase protein 8) having the highest (5.66%), relative to the size of the gene. The 13 protein-coding genes contained 506 SNPs (421 transitions and 85 transversions), of which 151 were in non-synonymous positions. ND4 (NADH dehydrogenase, subunit 4; 1344 bp) carried the highest number of non-synonymous SNPs (24) whereas the shortest genes ATP8 (159 bp) and ND4L (NADH dehydrogenase, subunit 4L; 264 bp) carried the lowest with only 3 non-synonymous SNPs each (Table S4).

3.2. tRNA^{leu}-COX2 intergenic region

Using the nomenclature system developed by Garnery et al. (1993) and revised by Rortais et al. (2011), for the M lineage, and Chávez-Galarza et al. (2017), for the A lineage, a total of 34 haplotypes

and 83 variants were identified for the 123 individuals. The most common haplotypes were A2 (19 individuals, 15.4%), M4 (15 individuals, 12.2%), and A1 (14 individuals, 11.4%). The haplotypes were grouped into lineages C (5 individuals of *A. m. carnica* and *A. m. ligustica*), M (38 individuals of *A. m. iberiensis*, *A. m. mellifera*, *and A. m. ligustica*), and A (80 individuals of *A. m. iberiensis*, *A. m. intermissa*, *A. m. sahariensis*, and *A. m. siciliana*). The A lineage was further partitioned into sub-lineages A_I (51 individuals of *A. m. iberiensis*, *A. m. intermissa*, *A. m. sahariensis*, and *A. m. siciliana*), A_{II} (20 individuals of *A. m. iberiensis*, *A. m. intermissa*, *A. m. sahariensis*), and A_{III} (9 individuals of *A. m. iberiensis*; Table S1). Remarkably, the three sub-lineages coexist in Iberia with the most common and widespread being A_I (45 individuals), whereas A_{II} was the least common (5 individuals). Sub-lineage A_{III} was confined to the Atlantic part of Iberia (Figure 1).

3.3. Diversity across mitochondrial genes

A total of 115 haplotypes was identified in the 123 mitogenomes (Table S4). The genes with the highest number of haplotypes were CYTB (44), ND4 (43), and COX1 (41), and these were also the lengthiest (≥ 1152 bp; Table S4). The gene with the highest haplotypic diversity was CYTB (1152 bp long; Hd = 0.910) and the gene with the lowest was ATP8 (159 bp long; Hd = 0.522; Table S4). This diversity measure together with the number of haplotypes, parsimony-informative sites, and the average number of nucleotide differences were highly correlated with the length of the gene ($r^2 = 0.85$, 0.94, 0.99, and 0.99, respectively).

3.4. Comparison of diversity levels

Diversity estimates of the mitogenome were not fully matched by most individual genes and the $tRNA^{leu}$-COX2 region. Nucleotide diversity inferred from the mitogenome generated the highest values for *A. m. sahariensis* ($\pi = 23.05$; Table II and S5), closely followed by the M lineage component of *A. m. iberiensis* ($\pi = 22.77$), and the lowest for the A lineage component of *A. m. iberiensis* ($\pi = 10.85$). While this pattern was concurrently captured by ND2 (NADH dehydrogenase, subunit 2), ND3 (NADH dehydrogenase, subunit 3), and COX1, this was not the case for all the other genes. In COX2, for example, the M lineage component of *A. m. iberiensis* exhibited the lowest nucleotide diversity across all different populations ($\pi = 0.21$; Table S5). Contrasting with the mitogenome and individual genes, the haplotypic diversity estimated from the intergenic region generated the highest values for the A lineage component of *A. m. iberiensis* (uh = 0.87; Table I).

3.5. Phylogeographical structure inferred from mitogenomes

The network inferred from the mitogenomes clustered the 115 haplotypes into three main haplogroups, which coincided with the lineages A, M, and C previously identified from $tRNA^{leu}$-COX2 variation. However, the mitogenomes did not support a partitioning of African variation into sub-lineages (Figure 2). Yet, three geographical clusters could be identified in the African haplogroup: one containing individuals mostly from Algeria (N Africa), one containing individuals from southeastern Iberia and north Africa nearby the Strait of Gibraltar (N Africa and S Iberia), and the other one containing individuals mostly from the Atlantic coast of Iberia (Iberian A lineage; Figure 3). Haplotypes previously defined as belonging to sub-lineages A_I and A_{II} were widespread within the African haplogroup and while A_{III} haplotypes were more confined in the network, they were connected with haplotypes A_I and A_{II}.

Interestingly, although previous studies using the $tRNA^{leu}$-COX2 marker did not report the presence of sub-lineages within lineage M, four well-defined clusters were captured by the mitogenomes: two of them were formed by individuals mainly from Iberia (Iberian M lineage and NW Iberia and M lineage *A. m. ligustica*); one by individuals from Iberia, France, and Italy (Iberian and S French M lineage); and another by individuals from North Europe (N Europe M lineage; Figure 3).

3.6. Phylogenetic patterns: gene comparisons

The mitogenome patterns were not fully concordant with those reconstructed from individual

Table II. Diversity measures estimated from the mitogenomes.

Lineage	Subspecies	N [a]	Ts [b]	Tv [c]	Private haplotypes	π [d]
M	*A. m. iberiensis*	28	121	23	99	22.77
	A. m. mellifera	8	56	9	34	20.32
	*A. m. ligustica**	2	0	0	0	0.00
A	*A. m. iberiensis*	59	115	30	104	10.85
	A. m. intermissa	12	42	7	27	13.64
	A. m. sahariensis	7	67	5	42	23.05
	*A. m. siciliana**	2	31	0	16	31.00
C	*A. m. ligustica**	2	11	2	9	13.00
	*A. m. carnica**	3	22	7	23	19.33

[a] Number of individuals; [b] number of transitions; [c] number of transversions; [d] nucleotide diversity. *Diversity estimates should be interpreted with caution because $N < 5$

genes and the tRNA^{leu}-COX2 region (Fig. S1, Table S6). While all genes (except s-rRNA) supported the presence of lineages A, M, and C, they disagreed regarding their relationships. Apart from COX2, CYTB, and ND2, which suggested a closer relationship between lineages C and M, the remaining genes were concordant with the mitogenome by placing lineage C closer to A than

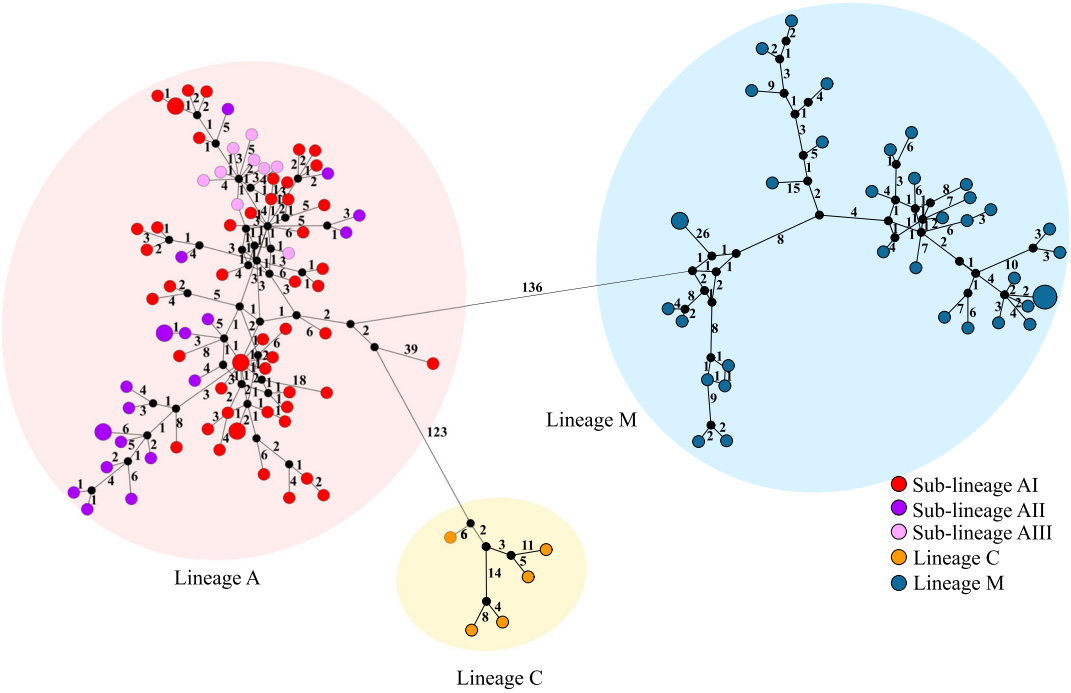

Figure 2 Median-joining network inferred from the mitogenomes. The network supports the A, C, and M evolutionary lineages but does not cluster African haplotypes by sub-lineages. The size of the circles is proportional to haplotype frequencies and the colors correspond to the African sub-lineages and lineages, as determined by the tRNA^{leu}-COX2 marker. Branch length is proportional to genetic distances between haplotypes. Hypothetical (unsampled or extinct) haplotypes are denoted as filled black circles. The values along the branches indicate the number of mutational steps between nodes.

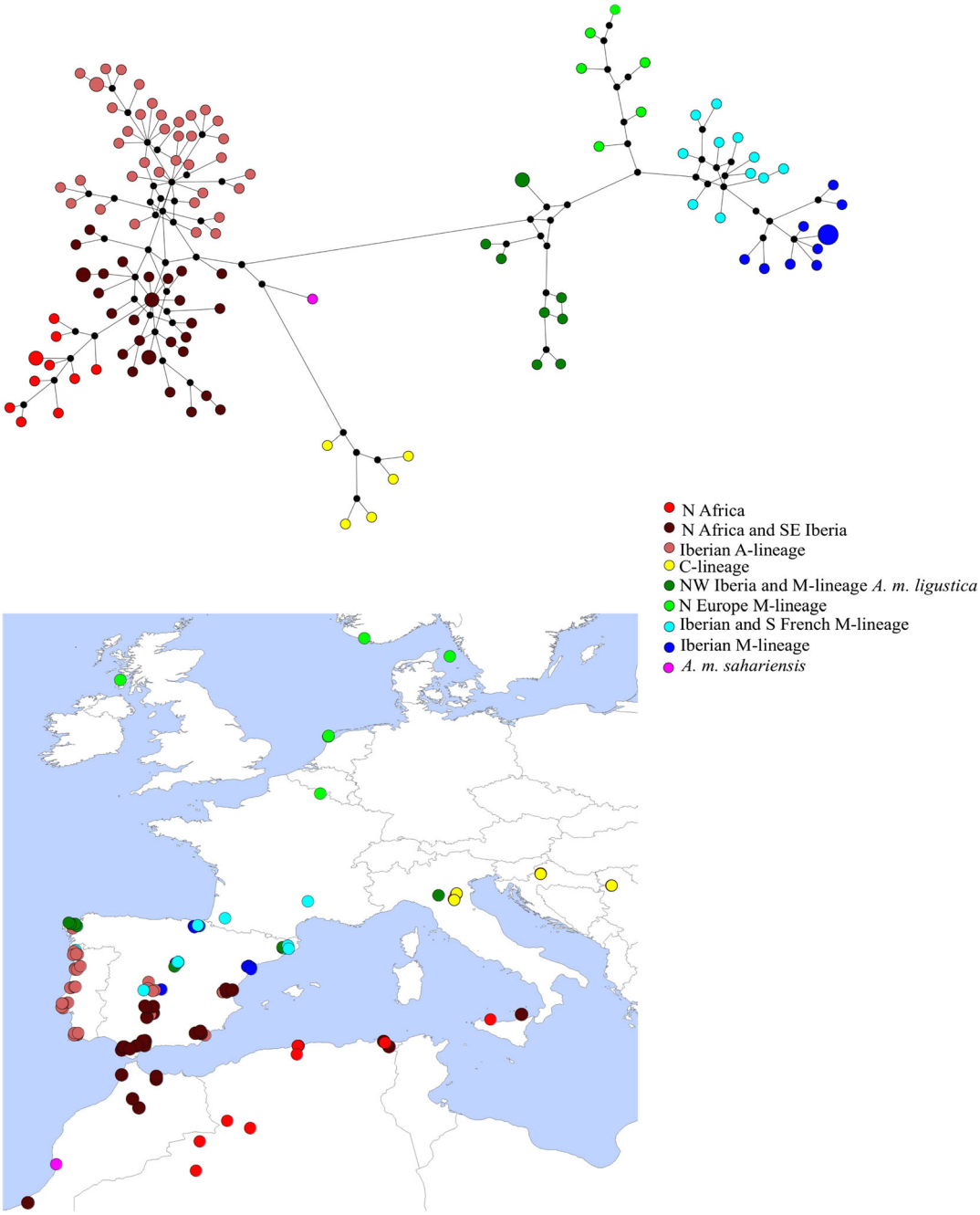

Figure 3 Median-Joining network inferred from the mitogenomes. Phylogeographical relationships show the existence of three and four well-defined African and Western European geographical clusters, respectively. The haplogroups are represented by the same color in the network and map. Hypothetical (unsampled or extinct) haplotypes are denoted as filled black circles in the network.

to M. The Bayesian phylogenetic tree inferred from the mitogenomes was concordant with the network analysis (Figure 4). In addition to the main phylogenetic clades coincidental with the

three evolutionary lineages, the topology largely supported the subgroups within A and M. The only exception was the north African and south-eastern Iberian groups (N Africa and SE Iberia) that did not form a single clade like in the network (Figure 3). Interestingly, individuals sharing the tRNA^{leu}-COX2 haplotype were not always clustered together. Instead, clusters tended to reunite individuals from the same geographical area. For instance, five of the seven A1 individuals from southern Portugal clustered together and were more related with one A9 from the same location than with the other two A1 individuals from the center of Portugal.

The NJ dendrogram, representing gene groupings by topology similarities, suggested that the topology reconstructed from the mitogenome is remarkably different from the individual coding genes (Figure 5). Among the 15 genes, COX1 was the most similar to the mitogenome, although it was still distant (PH85 = 53; Table S7). Individual genes produced different results, with only three pairs of genes (ATP8:ND4L, ND3:ATP6, ND6:COX3) being able to reconstruct similar topologies between each other.

The bPTP and ABGD methods were not concordant in the number of groups for single genes (Table S8). The bPTP was the most conservative method, identifying for the great majority of genes a single group, and for ND6 (NADH dehydrogenase, subunit 6) and ND4 three groups matching the three lineages. The ABGD method showed a variable number of clusters, ranging from one in ND5 (using the Jukes Cantor distance metric) and ND4L (using both Jukes Cantor and Kimura distance metrics) to 36 in ND1 (NADH dehydrogenase, subunit 1). Only ND2 and COX1 were congruent with the mitogenome capturing the three groups (Figure 4).

4. DISCUSSION

Mitochondrial DNA has been widely employed for assessing genetic diversity and structure of honey bee populations in the native and introduced ranges. Among the diverse array of mtDNA genes that have been interrogated using sequence or PCR-RFLP data (Collet et al. 2007; Ferreira et al. 2009; Pinto et al. 2005), the tRNA^{leu}-COX2 intergenic region has been largely preferred with hundreds of haplotypes described so far (Meixner et al. 2013). By assembling the mitogenomes of 123 individuals representing seven subspecies, we had an unprecedented opportunity to compare patterns obtained from the nearly complete chromosome with those obtained from individual genes and the popular tRNA^{leu}-COX2 marker. The 123 mitogenomes were sequenced at extremely high coverage (2523×–7758×), strengthening the reliability of the 645 detected SNPs. A total of 115 haplotypes was identified in the 123 individuals, revealing a remarkable diversity also reported in other mitogenome studies (Wragg et al. 2016).

While all mitochondrial genes are linked, it has been shown for many organisms, and now for the honey bee, that different regions evolve at different rates, leading to incongruent phylogenies and phylogeographical patterns (Duchene et al. 2012; Keis et al. 2013; Meiklejohn et al. 2014; Sasaki et al. 2005; Zardoya and Meyer 1996). The networks reconstructed from most individual genes showed low resolution and were incongruent among each other. This can be illustrated by ATP6, ATP8, and COX3 which distinguished three haplogroups and placed lineage C closer to A whereas ND2 and CYTB distinguished five haplogroups and placed lineage C closer to M. As expected, the mitogenome provided greater resolution than individual genes discriminating nine haplogroups and placing lineage C closer to A, a pattern concordant with nuclear SNPs (Whitfield et al. 2006). Notably, even the complete mitogenome was unable of distinguishing the populations at the subspecies level. The northern African individuals from nearby the Strait of Gibraltar (*A. m. intermissa*), from southeastern Spain (*A. m. iberiensis*), and one from Sicily (*A. m. siciliana*) were grouped together regardless of the subspecies, suggesting a recent shared evolutionary history.

The lengthiest genes (COX1, ND4, CYTB, and ND5) generated phylogenies closer to that of the mitogenome than the shorter ones. This is in agreement with the theoretical and empirical studies showing that longer sequences better resolve phylogenies than those based on shorter sequences (DeFilippis and Moore 2000; Saitou and

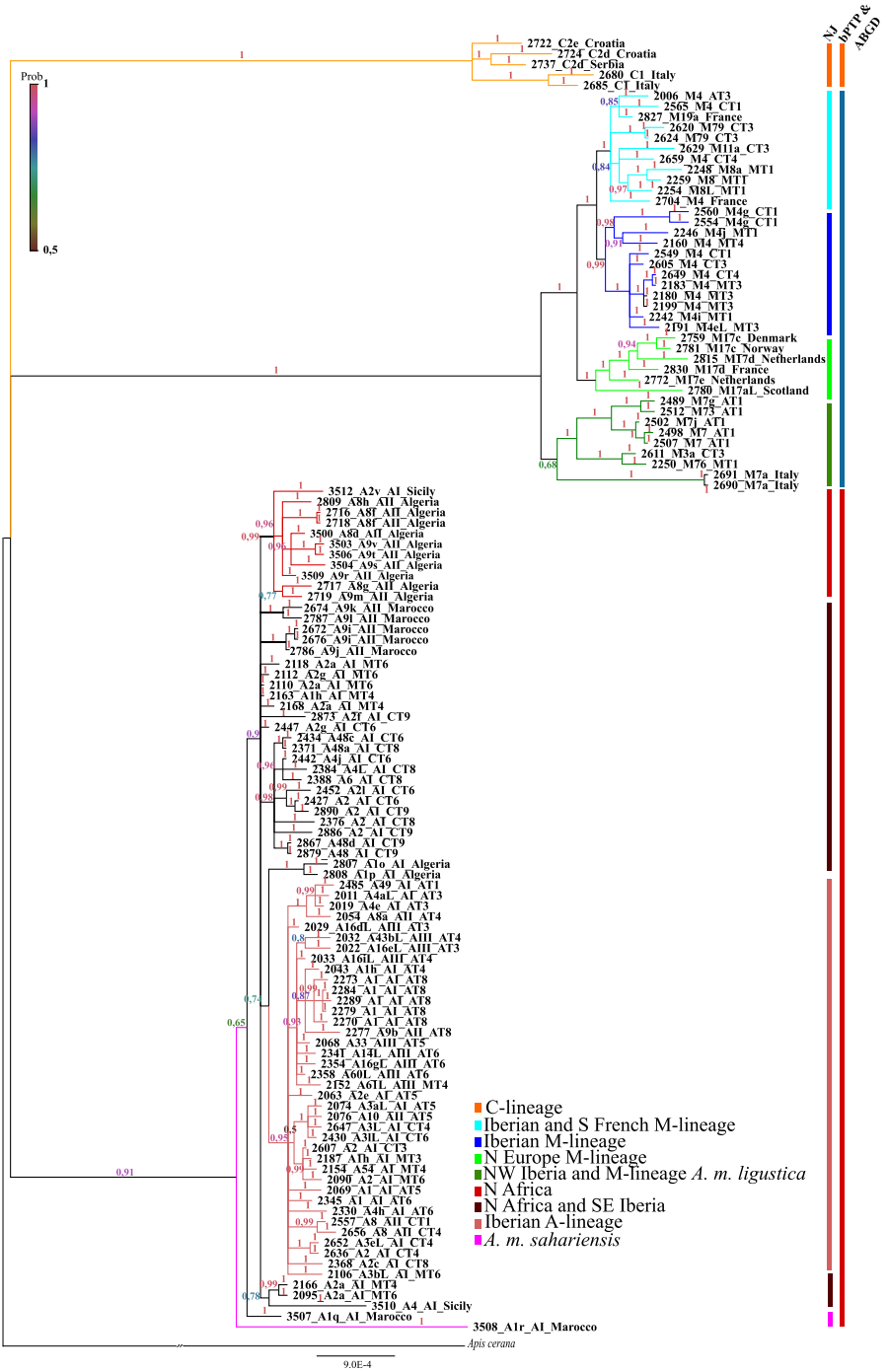

Figure 4 Bayesian phylogenetic tree inferred from the mitogenomes. Phylogenetic relationships are concordant with those inferred from the network analysis, as depicted in the first vertical bar. The group delimitation analyses (bPTP and ABGD) support the evolutionary lineages C (orange), M (blue), and A (red), as depicted in the second vertical bar. Values along the branches of the phylogenetic tree indicate the bootstrap support and are colored from the less to the most probable.

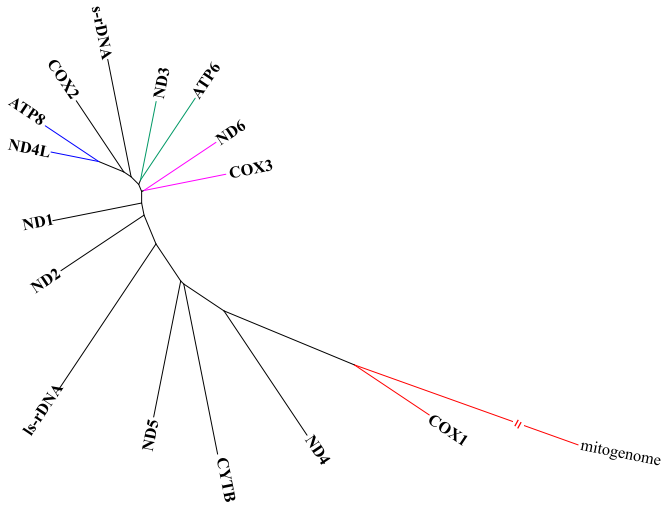

Figure 5 Neighbor-joining dendrogram for the PH85 topology distance between the single genes and the mitogenome. The dendrogram suggests that the topology reconstructed from the mitogenome is remarkably different from the individual coding genes.

Nei 1986). Among the lengthiest genes, COX1 generated the topology closest to the mitogenome, which is coincidently the standard DNA barcoding region for animals (Schmidt et al. 2015). Some of the lengthiest genes revealed to be more powerful for resolving relationships within M lineage (e.g., ND5 and ND2 distinguished three haplogroups), while others were more powerful for resolving relationships within A lineage (e.g., COX1, ND4, and CYTB distinguished two haplogroups). This finding suggests that either we use the full mitogenome to have the highest phylogenetic resolution across lineages, or the potential to resolve relationships depends not only on the gene but also on the samples under scrutiny.

The patterns captured by the mitogenome and individual genes were not congruent with those of the tRNA^{leu}-COX2 region. The networks generated from the mitogenome and individual genes clustered the haplotypes into three main well-defined haplogroups that matched lineages A, M, and C originally identified by the *Dra* I test (Garnery et al. 1993). However, the African sub-lineages proposed by Franck et al. (2001) were neither supported by the mitogenome and individual genes nor identified by the delimitation methods. Remarkably, even at the haplotypic level, individuals sharing the *Dra* I-based haplotype were not always clustered together. These results provide evidence for the homoplasious evolution of the intergenic region. It was also interesting to note that the remarkably high diversity levels observed in Iberian honey bees of A lineage ancestry (this study and Cánovas et al. 2008; Franck et al. 2001; Franck et al. 1998; Miguel et al. 2007; Pinto et al. 2013), as compared with the putative ancestral African populations (Whitfield et al. 2006), were not supported by any individual gene or the mitogenome. The higher variation in Iberian populations has been explained by a higher mutation rate related to the larger length of the tRNA^{leu}-cox2 sequence, which offers more targets for site mutations and more possibilities for Q duplications/deletions (Cornuet and Garnery 1991; Franck et al. 1998). The complex architecture of the intergenic region formed by a P element (which vary in sequence and length) combined with a variable number of Q elements (which also vary in sequence and length) suggests a complex mode of evolution where homoplasy has an important role.

This study shows that diversity patterns inferred from the popular tRNA^{leu}-COX2 intergenic region did not accurately reflect the intraspecific relationships and variation of the complete mitochondrial genome. While the mitogenome supports the three evolutionary lineages defined by the *Dra* I test, the African sub-lineages, and even

the haplotypes, were not fully supported. Due to the extremely high variability, this marker has revealed powerful for uncovering recent events, such as the historical colonization of the new world (Collet et al. 2006). On the other hand, homoplasy makes the intergenic region unsuited for inferring phylogenetic relationships between individuals.

ACKNOWLEDGEMENTS

We are deeply indebted to John C. Patton, Phillip San Miguel, Paul Parker, and Rick Westerman, University of Purdue, for resequencing the whole genomes, and to Pilar de la Rúa and to Carlo Amodeo for providing *A. m. sahariensis* and *A. m. siciliana* samples, respectively. Analyses were performed using computational resources at the Polytechnic Institute of Bragança. DH was supported by a PhD scholarship (SFRH/BD/84195/2012) from the Portuguese Science Foundation (FCT). This work was supported by FCT through the programs COMPETE/QREN/EU (PTDC/BIA-BEC/099640/2008) and the 2013-2014 BiodivERsA/FACCE-JPI joint call for research proposals, with the national funders FCT (Portugal), CNRS (France), and MEC (Spain).

AUTHOR CONTRIBUTIONS

MAP, DH, and JC-G conceived the ideas and designed the methodology. DH performed most of the analyses with assistance of JC-G, AQ, CJN, ARL, and FCO. JR provided the computational resources and assisted with the computationally demanding analyses. CC provided honey bee samples. MAP and DH wrote the manuscript with input from FCO and CC. All the authors critically reviewed the manuscript for important intellectual content.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interest The authors declare that they have no conflict of interest.

De la familière région intergénique tRNA^{leu}-COX2 au mitogenome: informations tirées de diverses populations d'abeilles mellifères d'Europe et d'Afrique du Nord

Abeille à miel ibérique / région intergénique tRNA^{leu}-COX2 / mitogenome

Von der bekannten tRNA^{leu}-COX2 intergenischen Region zum Mitogenom: Einblicke von verschiedenen Honigbienen-Populationen in Europa und Nordafrika

Iberische Honigbiene / tRNA ^{leu}-COX2 intergenische Region / Mitogenom

REFERENCES

Alburaki, M., Moulin, S., Legout, H., Alburaki, A., Garnery, L. (2011) Mitochondrial structure of Eastern honeybee populations from Syria, Lebanon and Iraq Apidologie **42**, 628–641. https://doi.org/10.1007/s13592-011-0062-4

Arias, M.C., Sheppard, W.S. (1996) Molecular phylogenetics of honey bee subspecies (*Apis mellifera* L) inferred from mitochondrial DNA sequence Mol. Phylogenet. Evol. **5**, 557–566. https://doi.org/10.1006/mpev.1996.0050

Avise, J. (1986) Mitochondrial DNA and the evolutionary genetics of higher animals Philos. Trans. R. Soc. B: Biol. Sci. **312**, 325–342

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

Bouga, M., Harizanis, P. C., Kilias, G., Alahiotis, S. (2005) Genetic divergence and phylogenetic relationships of honey bee *Apis mellifera* (Hymenoptera: Apidae) populations from Greece and Cyprus using PCR – RFLP analysis of three mtDNA segments Apidologie. **36**, 335–344. https://doi.org/10.1051/apido:2005021

Cánovas, F., De la Rúa, P., Serrano, J., Galián, J. (2008) Geographical patterns of mitochondrial DNA variation in *Apis mellifera iberiensis* (Hymenoptera: Apidae) J. Zool. Syst. Evol. Res. **46**, 24–30. https://doi.org/10.1111/j.1439-0469.2007.00435.x

Chávez-Galarza, J., Garnery, L., Henriques, D., Neves, C.J., Loucif-Ayad, W., Jonhston, J.S., Pinto, M.A. (2017) Mitochondrial DNA variation of *Apis mellifera iberiensis*: further insights from a large-scale study using sequence data of the tRNAleu-cox2 intergenic region Apidologie. **48**, 533–544. https://doi.org/10.1007/s13592-017-0498-2

Clary, D.O., Wolstenholme, D.R. (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code J. Mol. Evol. **22**, 252–271

Collet, T., Ferreira, K. M., Arias, M. C., Soares, A. E. E., Del Lama, M. A. (2006) Genetic structure of

Africanized honeybee populations (*Apis mellifera L.*) from Brazil and Uruguay viewed through mitochondrial DNA COI–COII patterns Heredity (Edinb.) **97**, 329–335. https://doi.org/10.1038/sj.hdy.6800875

Collet, T., Arias, M. C., Del Lama, M. A. (2007) 16S mtDNA variation in *Apis mellifera* detected by PCR-RFLP Apidologie. **38**:47–54. https://doi.org/10.1051/apido:2006056

Cornuet, J., Garnery, L. (1991) Mitochondrial DNA variability in honeybees and its phylogeographic implications Apidologie. **22**, 627–642

Crozier, R.H., Crozier, Y.C. (1993) The mitochondrial genome of the honeybee *Apis-mellifera*-Complete sequence and genome organization Genetics. **133**, 97–117

Crozier, Y.C., Koulianos, S., Crozier, R.H. (1991) An improved test for africanized honeybee mitochondrial-DNA Experientia. **47**, 968–969. https://doi.org/10.1007/bf01929894

DeFilippis, V.R., Moore, W.S. (2000) Resolution of phylogenetic relationships among recently evolved species as a function of amount of DNA sequence: An empirical study based on woodpeckers (Aves: Picidae) Mol. Phylogenet. Evol. **16**, 143–160

DeSalle, R., Schierwater, B., Hadrys, H. (2017) MtDNA: The small workhorse of evolutionary studies Front. Biosci. (Landmark Ed.) **22**, 873–887

Duchene, S., Frey, A., Alfaro-Núñez, A., Dutton, P.H., Thomas P. Gilbert, M., Morin, P.A. (2012) Marine turtle mitogenome phylogenetics and evolution Mol. Phylogenet. Evol. **65**, 241–250. https://doi.org/10.1016/j.ympev.2012.06.010

Excoffier, L., Estoup, A., Cornuet, J. M. (2005) Bayesian analysis of an admixture model with mutations and arbitrarily linked markers Genetics. **169**, 1727–1738. https://doi.org/10.1534/genetics.104.036236

Ferreira, K.M., Silva, O.L.E., Arias, M.C., Del Lama, M.A. (2009) Cytochrome-b variation in *Apis mellifera* samples and its association with COI-COII patterns Genetica. **135**, 149–155. https://doi.org/10.1007/s10709-008-9264-8

Filipi, K., Marková, S., Searle, J. B., Kotlík, P. (2015) Mitogenomic phylogenetics of the bank vole *Clethrionomys glareolus*, a model system for studying end-glacial colonization of Europe Mol. Phylogenet. Evol. **82**, 245–257. https://doi.org/10.1016/j.ympev.2014.10.016

Franck, P., Garnery, L., Solignac, M., Cornuet, J.-M. (1998) The origin of west european subspecies of honeybees (*Apis mellifera*): new insights from microsatellites and mitochondrial data Evolution. **52**, 1119–1134. https://doi.org/10.1111/j.1558-5646.1998.tb01839.x

Franck, P., Garnery, L., Loiseau, A., Oldroyd, B.P., Hepburn, H.R., Solignac, M., Cornuet, J.M. (2001) Genetic diversity of the honeybee in Africa: microsatellite and mitochondrial data Heredity. **86**, 420–430. https://doi.org/10.1046/j.1365-2540.2001.00842.x

Garnery, L., Cornuet, J.M., Solignac, M. (1992) Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis Mol. Ecol. **1**, 145–154. https://doi.org/10.1111/j.1365-294X.1992.tb00170.x

Garnery, L., Solignac, M., Celebrano, G., Cornuet, J.M. (1993) A simple test using restricted PCR-amplified mitochondrial DNA to study the genetic structure of *Apis mellifera L* Experientia 49:1016–1021. https://doi.org/10.1007/bf02125651

Hall, H.G., Smith, D.R. (1991) Distinguishing African and European honeybe matrilines using amplified mitochondrial DNA Proc. Natl. Acad. Sci. U. S. A. **88**:4548–4552. https://doi.org/10.1073/pnas.88.10.4548

Hammer, Ø., Harper, D.A.T., Ryan, P.D. (2001) Paleontological statistics software: package for education and data analysis. Palaeontol. Electron.

Henriques, D., Parejo, M., Vignal, A., Wragg, D., Wallberg, A., Webster, M.T., Pinto, M.A. (2018) Developing reduced SNP assays from whole-genome sequence data to estimate introgression in an organism with complex genetic patterns, the Iberian honeybee (*Apis mellifera iberiensis*). Evol. Appl.

Hong, Y., Duo, H., Hong, J., Yang, J., Liu, S., Yu, L., Yi, T. (2017) Resequencing and comparison of whole mitochondrial genome to gain insight into the evolutionary status of the Shennongjia golden snub-nosed monkey (SNJ *R. roxellana*) Ecol. Evol. **7**, 4456–4464. https://doi.org/10.1002/ece3.3011

Jacobsen, M.W., Hansen, M.M., Orlando, L., Bekkevold, D., Bernatchez, L., Willerslev, E., Gilbert, M.T.P. (2012) Mitogenome sequencing reveals shallow evolutionary histories and recent divergence time between morphologically and ecologically distinct European whitefish (*Coregonus spp.*) Mol. Ecol. **21**, 2727–2742. https://doi.org/10.1111/j.1365-294X.2012.05561.x

Jukes, T.H., Cantor, C.R., Munro, H.N. (1969) Evolution of protein molecules Mammalian protein Metabolism. **3**, 132

Keis, M., Remm, J., Ho, S.Y.W., Davison, J., Tammeleht, E. et al. (2013) Complete mitochondrial genomes and a novel spatial genetic method reveal cryptic phylogeographical structure and migration patterns among brown bears in north-western Eurasia J. Biogeogr. **40**, 915–927. https://doi.org/10.1111/jbi.12043

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences J. Mol. Evol. **16**, 111–120 0022–2844

Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S. (2012) PartitionFinder: Combined selection of partitioning schemes and substitution models for phylogenetic analyses Mol. Biol. Evol. **29**, 1695–1701. https://doi.org/10.1093/molbev/mss020

Meiklejohn, K.A., Danielson, M.J., Faircloth, B.C., Glenn, T.C., Braun, E.L., Kimball, R.T. (2014) Incongruence among different mitochondrial regions: a case study using complete mitogenomes Mol. Phylogenet. Evol. **78**, 314–323 1055–7903

Meixner, M.D., Pinto, M.A., Bouga, M., Kryger, P., Ivanova, E., Fuchs, S. (2013) Standard methods for characterising subspecies and ecotypes of *Apis mellifera* J. Apic. Res. **52**: 4, 1–28.https://doi.org/10.3896/ibra.1.52.4.05

Miguel, I., Iriondo, M., Garnery, L., Sheppard, W.S., Estonba, A. (2007) Gene flow within the M evolutionary lineage of *Apis mellifera*: role of the Pyrenees, isolation by distance and post-glacial re-colonization routes in the western Europe Apidologie. **38**, 141–155. https://doi.org/10.1051/apido:2007007

Miller, M.A., Pfeiffer, W., Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE), 14–14 Nov. 2010. pp 1–8. https://doi.org/10.1109/GCE.2010.5676129

Nielsen, D.I., Ebert, P.R., Page, R.E., Hunt, G.J., Guzman-Novoa, E. (2000) Improved polymerase chain reaction-based mitochondrial genotype assay for identification of the africanized honey bee (Hymenoptera : Apidae) Ann. Entomol. Soc. Am. **93**, 1–6. https://doi.org/10.1603/0013-8746(2000)093[0001:ipcrbm]2.0.co;2

Ozdil, F., Ilhan, F. (2012a) Diversity of *Apis mellifera* Subspecies from Turkey revealed by sequence analysis of mitochondrial 16s rDNA region Biochem. Genet. **50**, 748–760. https://doi.org/10.1007/s10528-012-9517-1

Ozdil, F., Ilhan, F. (2012b) Genetic Divergence of Turkish *Apis mellifera* Subspecies Based on Sequencing of ND5 Mitochondrial Segment Sociobiology. **59**, 225–234

Pang, J.-F., Kluetsch, C., Zou, X.-J., Zhang, A.-B., Luo, L.-Y. et al. (2009) mtDNA data indicate a single origin for dogs south of Yangtze river, less than 16,300 Years Ago, from numerous wolves Mol. Biol. Evol. **26**, 2849–2864. https://doi.org/10.1093/molbev/msp195

Paradis, E., Claude, J., Strimmer, K. (2004) APE: Analyses of Phylogenetics and Evolution in R language Bioinformatics. **20**, 289–290. https://doi.org/10.1093/bioinformatics/btg412

Peakall, R.O.D., Smouse, P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol Resour. **6**, 288–295 1471–8286

Penny, D., Hendy, M.D. (1985) The use of tree comparison metrics Syst. Zool. **34**, 75–82. https://doi.org/10.2307/2413347

Pinto, M.A., Johnston, J.S., Rubink, W.L., Coulson, R.N., Patton, J.C., Sheppard, W.S. (2003) Identification of Africanized honey bee (Hymenoptera: Apidae) mitochondrial DNA: validation of a rapid polymerase chain reaction-based assay Ann. Entomol. Soc. Am. **96**, 679–684

Pinto, M.A., Rubink, W.L., Patton, J.C., Coulson, R.N., Johnston, J.S. (2005) Africanization in the United States: replacement of feral European honey bees (*Apis mellifera* L.) by an African hybrid swarm. Genetics

Pinto, M.A., Muñoz, I., Chávez-Galarza, J., De la Rúa, P. (2012) The Atlantic side of the Iberian Peninsula: a hot-spot of novel African honey bee maternal diversity Apidologie **43**, 663–673

Pinto, M. A., Henriques, D., Guedes, H., Munoz, I., Azevedo, J., De la Rúa, P. (2013) Maternal diversity patterns of Ibero-Atlantic populations reveal further complexity of Iberian honeybees Apidologie. **44**, 430–439

Puillandre, N., Lambert, A., Brouillet, S., Achaz, G. (2012) ABGD, Automatic Barcode Gap Discovery for primary species delimitation Mol. Ecol. **21**, 1864–1877

R Core Team (2014) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria

Ronquist, F., Huelsenbeck, J.P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models Bioinformatics **19**, 1572–1574

Rortais, A., Arnold, G., Alburaki, M., Legout, H., Garnery, L. (2011) Review of the *Dra*I COI-COII test for the conservation of the black honeybee (*Apis mellifera mellifera*) Conserv. Genet. Resour. **3**, 383–391

Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods Bioinformatics **19**, 2496–2497

Ruttner, F. (1988) Biogeography and taxonomy of honeybees. Springer Science & Business Media

Saitou, N., Nei, M. (1986) The number of nucleotides required to determine the branching order of three species, with special reference to the human-chimpanzee-gorilla divergence J. Mol. Evol. **24**, 189–204

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989) Molecular Cloning, 2nd edn. Cold Spring Harbor Laboratory Press, New York.

Sasaki, T., Nikaido, M., Hamilton, H., Goto, M., Kato, H. et al. (2005) Mitochondrial phylogenetics and evolution of mysticete whales Syst. Biol. **54**, 77–90. https://doi.org/10.1080/10635150590905939

Schmidt, S., Schmid-Egger, C., Morinière, J., Haszprunar, G., Hebert, P. D. (2015) DNA barcoding largely supports 250 years of classical taxonomy: identifications for Central European bees (Hymenoptera, Apoidea partim) Mol. Ecol. Resour. **15**, 985–1000

Simon, S., Hadrys, H. (2013) A comparative analysis of complete mitochondrial genomes among Hexapoda Mol. Phylogenet. Evol. **69**, 393–403

Smith, D., Brown, W.M. (1988) Polymorphisms in mitochondrial DNA of European and Africanized honeybees (*Apis mellifera*) Experientia. **44**, 257–260

Smith, D., Palopoli, M., Taylor, B., Garnery, L., Cornuet, J.-M., Solignac, M., Brown, W. (1991) Geographical overlap of two mitochondrial genomes in Spanish honeybees (*Apis mellifera iberica*) J. Hered. **82**, 96–100

Stevanovic, J., Stanimirovic, Z., Radakovic, M., Kovacevic, S.R. (2010) Biogeographic study of the honey bee *(Apis mellifera* L.) from Serbia, Bosnia and Herzegovina and Republic of Macedonia based on mitochondrial DNA

analyses Russ. J. Genet. **46**, 603–609 https://doi.org/10.1134/s1022795410050145

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0 Mol. Biol. Evol. **30**, 2725–2729 https://doi.org/10.1093/molbev/mst197

Techer, M. A., Clemencet, J., Simiand, C., Preeaduth, S., Azali, H. A., Reynaud, B., Delatte, H. (2017) Large-scale mitochondrial DNA analysis of native honey bee *Apis mellifera* populations reveals a new African sub-group private to the South West Indian Ocean islands BMC Genet. **18**: 1, 53. https://doi.org/10.1186/s12863-017-0520-8

Whitfield, C.W., Behura, S.K., Berlocher, S.H., Clark, A.G., Johnston, J.S. et al. (2006) Thrice out of Africa: ancient and recent expansions of the honey bee, *Apis mellifera* Science. **314**, 642–645

Wragg, D., Marti-Marimon, M., Basso, B., Bidanel, J.-P., Labarthe, E., Bouchez, O., Le Conte, Y., Vignal, A. (2016) Whole-genome resequencing of honeybee drones to detect genomic selection in a population managed for royal jelly Sci. Rep. **6**, 27168. https://doi.org/10.1038/srep27168

Zardoya, R., Meyer, A. (1996) Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates Mol. Biol. Evol. **13**, 933–942

Zhang, J., Kapli, P., Pavlidis, P., Stamatakis, A. (2013) A general species delimitation method with applications to phylogenetic placements Bioinformatics. **29**, 2869–2876 1460–2059