Apidologie (2017) 48:533–544

DOI: 10.1007/s13592-017-0498-2

Original article

Mitochondrial DNA variation of *Apis mellifera iberiensis*: further insights from a large-scale study using sequence data of the tRNA^{leu}-cox2 intergenic region

Julio Chávez-Galarza[1,2], Lionel Garnery[3,4], Dora Henriques[1,2]
Cátia J. Neves[1], Wahida Loucif-Ayad[5], J . Spencer Jonhston[6]
M . Alice Pinto[1]

[1]Mountain Research Centre (CIMO), Polytechnic Institute of Bragança, Campus de Sta. Apolónia, 5300-253, Bragança, Portugal
[2]Centre of Molecular and Environmental Biology (CBMA), University of Minho, Campus de Gualtar, 4710-057, Braga, Portugal
[3]Laboratoire Evolution, Génomes et Spéciation, CNRS, Bâtiment 13, Avenue de la Terrasse, 91198, Gif-sur-Yvette, France
[4]Saint Quentin en Yvelines, Université de Versailles, 45 Avenue des Etats-Unis, 78, Versailles, France
[5]Laboratory of Applied Animal Biology, University Badji-Mokhtar, Annaba, Algeria
[6]Department of Entomology, Texas A&M University, College Station, TX 77843-2475, USA

Received 8 June 2016 – Revised 23 October 2016 – Accepted 18 January 2017

Abstract – A large-scale survey of the Iberian honey bee (*Apis mellifera iberiensis*) diversity patterns, using sequence data of the tRNA^{leu}-cox2 mitochondrial DNA (mtDNA) region, demonstrates that earlier studies based on the *Dra* I test missed significant components of genetic variation. Based on results from this survey, existing haplotype names were revised and updated following a nomenclature system established earlier and extended herein for the intergenic region. A more complete picture of the complex diversity patterns of IHBs is revealed that includes 164 novel haplotypes, 113 belonging to lineage A and 51 to lineage M and within lineage A and 69 novel haplotypes that belong to sub-lineage A_I, 13 to A_{II}, and 31 to A_{III}. Within lineage M, two novel haplotypes show a striking architecture with features of lineages A and M, which based on sequence comparisons and relationships among haplotypes are seemingly ancestral. These data expand our knowledge of the complex architecture of the tRNA^{leu}-cox2 intergenic region in *Apis mellifera* and re-emphasizes the importance of Iberia as a source of honey bee mtDNA diversity.

Iberian honey bee / *Apis mellifera intermissa* / ancestral haplotype M / *Dra* I test

1. INTRODUCTION

The natural distribution of the honey bee, *Apis mellifera* L., encompasses Europe, Africa, and the Middle East (Ruttner 1988). Within this territory, adaptation to a wide range of environments has led to diversification of over 24 subspecies, all of which have been grouped into four lineages: the African (A), the western European (M), the eastern European (C), and the Middle Eastern (O) (Ruttner 1988). A sub-division of the African lineage into four sub-lineages (A_I, A_{II}, A_{III}, and Z) was later proposed (Alburaki et al. 2011).

Of those subspecies, the Iberian honey bee (IHB, *A. m. iberiensis*), which was allocated by Ruttner (1988) to lineage M, has been one of the most surveyed for genetic diversity due to its complex patterns of diversity and evolutionary history

Electronic supplementary material The online version of this article (doi:10.1007/s13592-017-0498-2) contains supplementary material, which is available to authorized users.

Corresponding author: M. Alice Pinto, apinto@ipb.pt
Manuscript editor: Marina Meixner

(Miguel et al. 2007, Cánovas et al. 2008, Pinto et al. 2014, Chávez-Galarza et al. 2015). The surveys have used a variety of markers, including morphology (Ruttner 1988; Cornuet and Fresnaye 1989; Arias et al. 2006), allozymes (Nielsen et al. 1994; Smith and Glenn 1995; Arias et al. 2006), mitochondrial DNA (Smith et al. 1991; Garnery et al. 1995, 1998; De la Rúa et al. 2005; Miguel et al. 2007; Cánovas et al. 2008; Pinto et al. 2012, 2013), microsatellites (Franck et al. 1998, 2001; Miguel et al. 2007; Cánovas et al. 2011), and, more recently, single nucleotide polymorphisms (SNPs; Chávez-Galarza et al. 2013, 2015).

Mitochondrial DNA (mtDNA) has indisputably been the most popular and longest used marker in Iberia. Early mtDNA analyses of IHBs digested the entire molecule with a battery of restriction endonucleases (reviewed by Meixner et al. 2013). This approach was soon replaced by the *Dra* I test, which consists of a single digestion, with the restriction enzyme *Dra* I, of the PCR-amplified $tRNA^{leu}$-cox2 intergenic fragment (Garnery et al. 1993). The relatively low cost and scoring simplicity, combined with the high information content of the $tRNA^{leu}$-cox2 intergenic region, have made this molecular tool the most widely used for mtDNA identification (Meixner et al. 2013). Over 3300 honey bee colonies collected across Iberia have been scored using the *Dra* I test (Franck et al. 1998; Miguel et al. 2007; Cánovas et al. 2008; Pinto et al. 2013), which has revealed the co-occurrence of the two highly divergent M and A lineages forming a northeastern-southwestern cline (Smith et al. 1991; Miguel et al. 2007; Cánovas et al. 2008). The clinal pattern has been confirmed by a recent study using genome-wide nuclear SNPs (Chávez-Galarza et al. 2015), strengthening the hypothesis of an ancestral secondary contact (Smith et al. 1991).

The great amount of haplotypes generated by the complexity of the intergenic region (see Supplementary Material for a detailed description of its architecture) has made necessary to develop a nomenclature system in conjunction with the *Dra* I test (Garnery et al. 1993). However, assigning names to haplotypes scored from complex gel banding patterns has not always been a straightforward endeavor, and this has been especially true in the presence of high levels of variation such as in Iberia. Complicating matters further, the increasing availability of sequence data has uncovered increasing amounts of variation that had been missed by the *Dra* I test (Rortais et al. 2011). Accordingly, haplotype names have not always followed the nomenclature criteria earlier established, leading to haplotype misnaming. In an attempt to correct many nomenclature problems, recently, Rortais et al. (2011) carried out a revision of the highly diverse lineage M. Using a large collection of sequences that were newly developed or available in GenBank, these authors amended many of the names of M haplotypes. Continued revision of African haplotype nomenclature is certainly warranted, given the great amount of mtDNA diversity that has been reported for Iberia.

The purpose of this study was threefold: (i) to provide a more complete survey of mtDNA variation of the $tRNA^{leu}$-cox2 region by sequencing a large collection of individuals sampled across the entire distributional range of the IHB, (ii) to revise the haplotype names of African ancestry by integrating information from the newly developed sequences with all the sequences deposited in GenBank, (iii) to refine the current status of haplotype diversity patterns in the IHB from sequence data using the revised collection of haplotypes. Mitochondrial DNA variation of IHBs has been extensively documented using restriction digestion data from the whole mtDNA molecule (Smith et al. 1991) or from the $tRNA^{leu}$-cox2 region (Franck et al. 1998; Miguel et al. 2007; Cánovas et al. 2008; Pinto et al. 2013). This, however, is the first study on mtDNA diversity patterns inferred strictly from sequence data. Using sequence data, a more detailed picture of mtDNA patterns of IHBs emerges showing that the *Dra* I test failed to reveal important amounts of existing variation.

2. MATERIAL AND METHODS

2.1. Sampling

Sampling was carried out in 2010 across three north-south transects in Iberia (Figure 1). One transect extended along the Atlantic coast (AT-Atlantic transect), one through the center (CT-

central transect), and another along the Mediterranean coast (MT-Mediterranean transect). A total of 711 individuals, each representing a single random colony, was collected from 237 apiaries (three colonies per apiary) grouped into 23 sites (see details in Chávez-Galarza et al. 2013). In addition to the Iberian sample, 31 *A. m. intermissa* (lineage A), 34 *A. m. mellifera* (lineage M), 17 *A. m. ligustica* (lineage C), and 19 *A. m. carnica* (lineage C) individuals, which have previously been identified by genome-wide SNPs (Chávez-Galarza et al. 2013, Pinto et al. 2014), were used herein as reference populations. The reference individuals were sampled in 2010 and 2011 across Europe and North Africa (Supplementary Material Figure S1). All 812 individuals were stored in ethanol at −20°C.

2.2. Mitochondrial DNA analysis

The tRNA^{leu}-cox2 mtDNA intergenic region encompasses the 3' end of the tRNA^{leu} gene, the P and Q elements, and the 5' end of the cox2 gene (see Supplementary Material for a detailed description of the architecture of this region and how it is used to discriminate lineages and African sub-lineages). This region was PCR-amplified using the primers and reaction conditions detailed by Garnery et al. (1993). The PCR products were sent to Macrogen (Seoul, Korea) for direct Sanger sequencing in both directions. The 812 sequences (711 *A. m. iberiensis*, 31 *A. m. intermissa*, 34 *A. m. mellifera*, 17 *A. m. ligustica*, and 19 *A. m. carnica*) were checked for base calling and aligned with 77 sequences available in GenBank using MEGA 6.06 (Tamura et al. 2013). Sequence data for *A. m. mellifera*, *A. m. ligustica*, and *A. m. carnica* were previously reported in Pinto et al. (2014). Haplotypes of A, M, and C ancestry were identified and named following the nomenclature system revised in Supplementary Material.

2.3. Genetic diversity and phylogenetic analysis

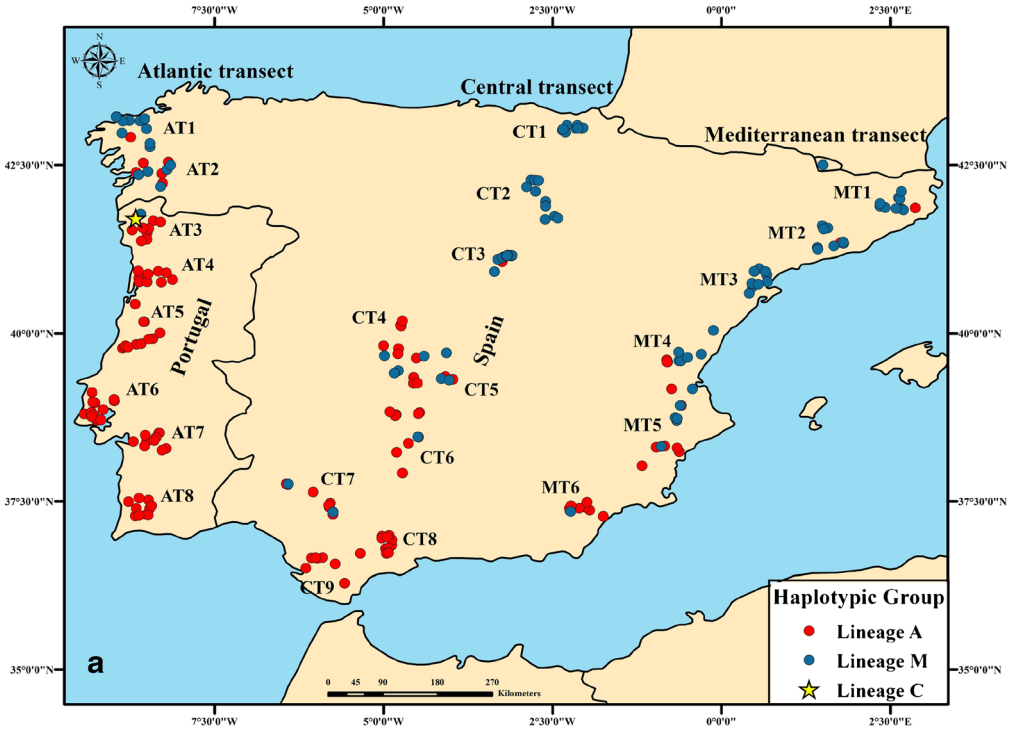
GENEALEX 6.5 (Peakall and Smouse 2012) was used to estimate mean number of haplotypes per locus (Na), effective number of haplotypes (Ne), number of private haplotypes (Np), and unbiased diversity (uh) per sampling site. Genetic differentiation among the Iberian sampling sites and the reference subspecies was estimated using Φ_{PT} values, which were then employed to perform a principal coordinate analysis (PCo) using GENEALEX 6.5 and to build a neighbor-joining tree using POPULATIONS 1.2.32 (Langella 2011).

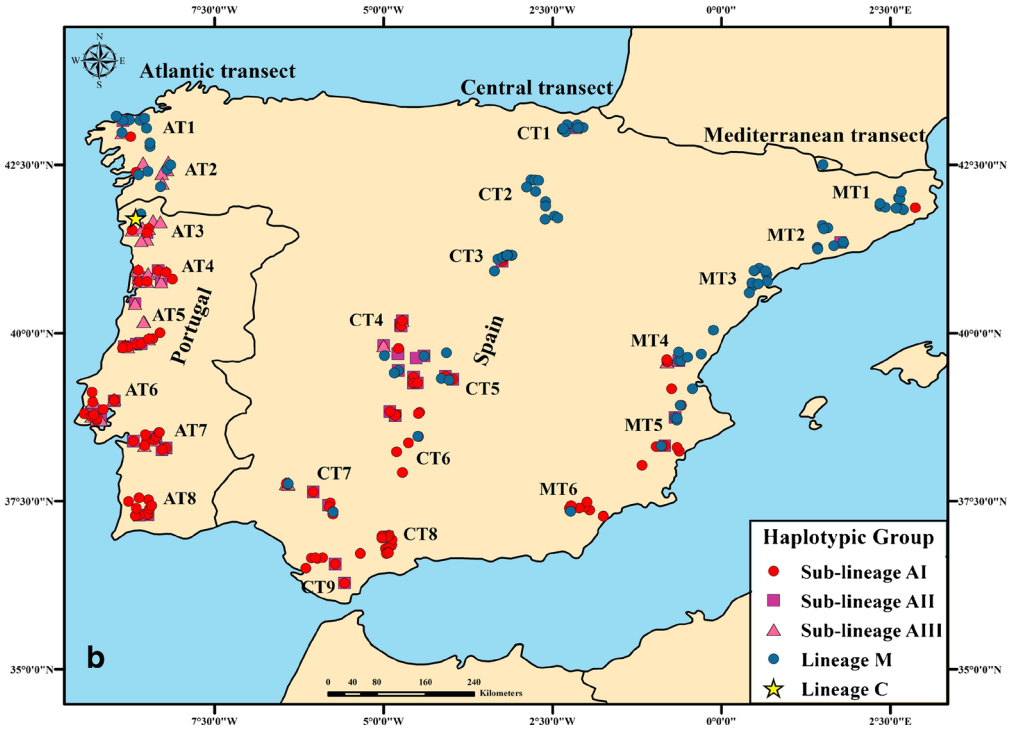
To identify lineages and sub-lineages to which the novel haplotypes belong, a phylogenetic analysis of haplotypes was performed in PHYLIP 3.65c (Felsenstein 1993) using maximum parsimony, with 77 reference haplotypes of A, M, and C ancestry downloaded from GenBank (Supplementary Material Table S2). The phylogenetic analyses included the polymorphic sites of the P element, the first Q element, and the fragment of cox2. The remaining Q elements, which varied between two and three, were considered as single mutational steps. Gaps were considered as a fifth character. Positions indicating large deletions within the P element (P, P_1, and absence of P) and the first two *Dra* I restriction sites were upweighted. The tRNA^{leu}-cox2 region of *Apis cerana* was used as an outgroup. The absence of the P element in lineage C and the absence of the Q element in *A. cerana* were coded as a single mutational step and the remaining positions as missing data.

Relationships among haplotypes and their relative proportions in Iberia were inferred using the median-joining network algorithm (Bandelt et al. 1999), as implemented in NETWORK 4.6.1.1 (Fluxus Engineering, Clare, UK), with the polymorphic sites within the P and Q elements used to infer the network. Deletions from 2 to 35 bp were considered as a single mutational step and coded as a 1-bp gap. As was done for phylogenetic analysis, large deletions within the P element (P, P_1, and absence of P) and the first two *Dra* I restriction sites were upweighted.

3. RESULTS

Sequence data of the tRNA^{leu}-cox2 intergenic region generated from 812 honey bee individuals collected in Iberia (N = 711), Western Europe and north of the Pyrenees (N = 34), Eastern Europe (N = 36), and North Africa (N = 31) revealed 212 haplotypes, of which 182 are reported here for the first time (Table S1). These novel haplotypes


7°30'0"W
5°0'0"W
2°30'0"W
0°0'0"
2°30'0"E
42°30'0"N
40°0'0"N
37°30'0"N
35°0'0"N
Atlantic transect
Central transect
Mediterranean transect
AT1
AT2
AT3
AT4
AT5
AT6
AT7
AT8
CT1
CT2
CT3
CT4
CT5
CT6
CT7
CT8
CT9
MT1
MT2
MT3
MT4
MT5
MT6
Portugal
Spain
a
0 45 90 180 270
Kilometers
Haplotypic Group
Lineage A
Lineage M
Lineage C


7°30'0"W
5°0'0"W
2°30'0"W
0°0'0"
2°30'0"E
42°30'0"N
40°0'0"N
37°30'0"N
35°0'0"N
Atlantic transect
Central transect
Mediterranean transect
AT1
AT2
AT3
AT4
AT5
AT6
AT7
AT8
CT1
CT2
CT3
CT4
CT5
CT6
CT7
CT8
CT9
MT1
MT2
MT3
MT4
MT5
MT6
Portugal
Spain
b
0 40 80 160 240
Kilometers
Haplotypic Group
Sub-lineage AI
Sub-lineage AII
Sub-lineage AIII
Lineage M
Lineage C

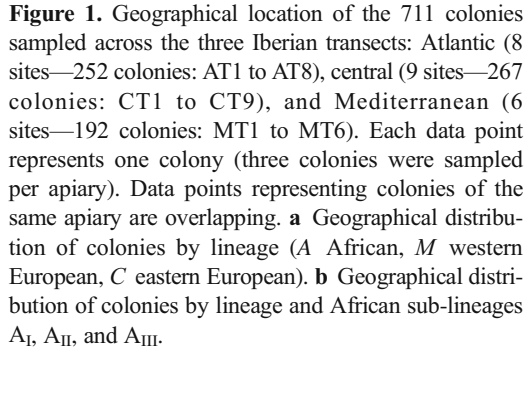


Figure 1. Geographical location of the 711 colonies sampled across the three Iberian transects: Atlantic (8 sites—252 colonies: AT1 to AT8), central (9 sites—267 colonies: CT1 to CT9), and Mediterranean (6 sites—192 colonies: MT1 to MT6). Each data point represents one colony (three colonies were sampled per apiary). Data points representing colonies of the same apiary are overlapping. **a** Geographical distribution of colonies by lineage (*A* African, *M* western European, *C* eastern European). **b** Geographical distribution of colonies by lineage and African sub-lineages A_I, A_{II}, and A_{III}.

(27 band patterns, 137 sequence variants) were harbored by 164 *A. m. iberiensis* and 18 *A. m. intermissa* individuals (1 band pattern, 17 sequence variants). Total detected variation was grouped into lineages A, M, and C. Haplotypes belonging to three lineages were identified in IHBs (128 A, 59 M, 1 C) and in *A. m. ligustica* (1 C, 1 M). In contrast, haplotypes in the other subspecies belonged to a single lineage (18 A in *A. m. intermissa* , 8 M in *A. m. mellifera* , and 6 C in *A. m. carnica*) (Table S1).

3.1. Phylogenetic analysis of novel haplotypes

The 182 novel haplotypes reported herein, together with 88 (including 30 renamed; see revision of haplotype names in Supplementary Material) previously published haplotypes, are shown in the phylogenetic topology of Figure 2. The novel haplotypes, identified among Iberian and reference honey bees, were grouped within lineages A (131) and M (51) (Table S1). In lineage A, 72 haplotypes belong to sub-lineage A_I, 28 to sub-lineage A_{II}, 31 to sub-lineage A_{III}, and none was identified as sub-lineage Z (see Supplementary Material for the rationale of mtDNA lineages and African sub-lineages discrimination). Lineage A and M formed a group, with sub-lineage Z as a basal group. Sub-lineages A_I, A_{II}, and A_{III} were grouped within one clade, although A_I is not differentiated. The phylogeny supported three African sub-lineages forming a sister group of lineage M. Lineage M formed a well-differentiated group, with novel haplotypes M79 and M79a forming a new sub-group. The main distinctive feature of this sub-group is a P element that is similar to the P_0 which is typical of A haplotypes, with additional nucleotide substitutions and restriction sites characteristic of M haplotypes (positions 14, 20, 24, 32, 34, 35, 38 for substitutions; e, k, l, x, ab, ad for deletions; and 15, 17, 26, 29, 36, 37 for *Dra* I sites in Figure S2). The group formed by C haplotypes was clearly differentiated from the remaining haplotypes.

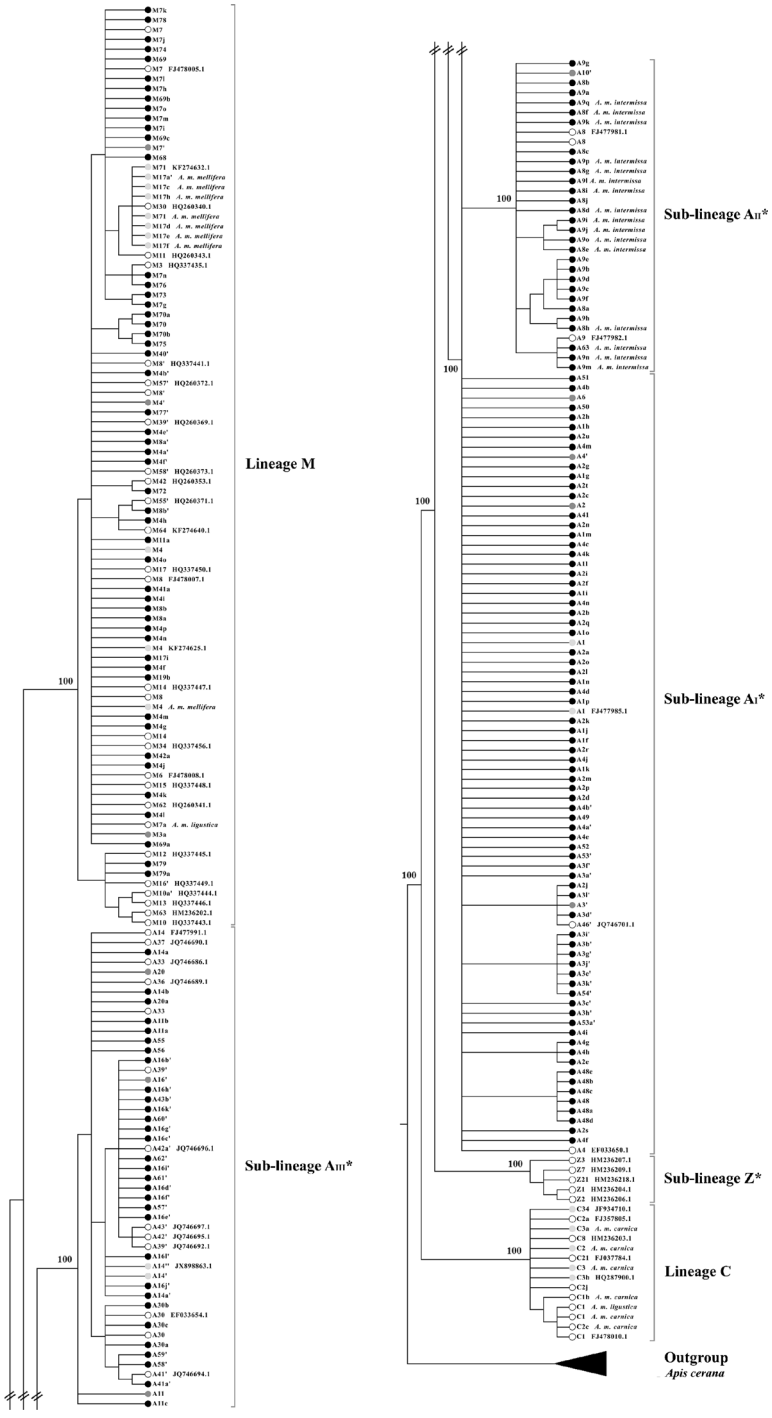
3.2. Distribution of haplotypes in Iberia

A total of 188 haplotypes was observed among the 711 colonies sampled across the three Iberian transects (AT-Atlantic, CT-central, MT-Mediterranean), of which 128 belong to lineage A, 59 to lineage M, and 1 to lineage C (Table S1). Within lineage A, 74 haplotypes belong to sub-lineage A_I, 15 to sub-lineage A_{II}, and 39 to sub-lineage A_{III}. Colonies of A ancestry were predominant (481, 67.5%), as compared with those of M (229, 32.2%) and C ancestry (1, 0.3%). Within the 481 colonies belonging to lineage A, those of sub-lineage A_I ancestry were more ubiquitous (333, 69.2%) than either A_{II} (68, 14.2%) or A_{III} (80, 16.6%).

As previously reported, haplotype A2 was the most frequent African haplotype (6.6%) in Iberia, occurring in virtually all sampling sites where lineage A is found, followed by A2a (6.5%), A1 (5.9%), A8 (5.6%), and A1h (3.8%). Among M haplotypes, M4 was the most frequent (9.7%), occurring mostly in the northeastern part of Iberia, followed by M7 and M7n with 2.1% each. All these haplotypes were observed in more than seven sites (Table S3). While the majority of haplotypes were found with low frequency, others were more abundant in certain sites, such as the case of M7 and A49 in AT1; A16' in AT2; A16b' in AT3; A1h in AT4; A20 in AT5; A1 from AT6 to AT8; M4 in CT1 and CT2; M79 in CT3; A8 in CT4 and CT5; A2 in CT6, CT7, and CT9; A2 and A2c in CT8; M4 from MT1 to MT3; and A2a from MT4 to MT6 (Table S3).

3.3. Genetic diversity in Iberia

Genetic diversity estimated for each Iberian sampling site is shown in Table I. Na ranged



from 7 in AT8 to 20 in AT4. Sites AT1 and AT3 exhibited the highest values of Ne with 13.235 effective haplotypes. For Np, the highest value was observed in AT4 (16) and the lowest in AT8 (1). The unbiased haplotype diversity (uh) ranged from 0.956 (AT1, AT3)

◀ **Figure 2.** Phylogeny of haplotypes based on the tRNA^{leu}-cox2 intergenic region for Iberian (*A. m. iberiensis*) and reference (A. m. *mellifera*, *A. m. ligustica*, and *A. m. carnica*) honey bee subspecies. Strict consensus (level 95%) from over 10,000 equally parsimonious trees. Values indicate bootstrap support of 1000 pseudoreplicates. *Sub-lineages A_I, A_{II}, A_{III}, and Z belong to the African lineage. *Black circles* indicate novel haplotypes, *gray circles* indicate haplotypes (band patterns) reported by others but sequenced here for the first time, *light gray circles* indicate renamed haplotypes, and *white circles* indicate haplotypes reported by others with a complete description (*Dra* I band patterns and sequence data).

to 0.464 (AT8). Globally, the highest diversity estimates were found across the Atlantic transect (except for AT8).

3.4. Relationships among haplotypes in Iberia

A median-joining network that illustrates the frequencies and relationships among the haplotypes found in Iberia is shown in Figure 3. Two highly divergent clusters, corresponding to lineages A and M, were observed. For lineage A, the three African sub-lineages were mainly linked by haplotypes A1 and A2 (sub-lineage A_I); A8 and A9a (sub-lineage A_{II}); and A11, A20 and A30 (sub-lineage A_{III}). Haplotypes A1, A2, A8, and A16' exhibit the greatest number of links with other haplotypes. For lineage M, haplotypes M79 and M79a present an almost intermediate position between haplotypes M and A, as observed in the phylogenetic tree (Figure 2). Haplotypes M4 and M7 occupy a central position between the two divergent groups and show more links than the other M haplotypes.

3.5. Relationships among lineages

The neighbor-joining tree of Iberian sampling sites and reference subspecies showed three well-supported groups (Figure 4a). The first group is formed by populations with a high proportion of C haplotypes (*A. m. ligustica* and *A. m. carnica*). The second group includes mostly M haplotypes carried by the reference *A. m. mellifera* and IHBs of the northeastern half of Iberia (AT1, CT1 to CT3, MT1 to MT3). The third group contains the reference *A. m. intermissa* and IHBs of the southwestern half of Iberia (AT2 to AT8, CT4 to CT9, MT4 to MT6). Within the third group, populations of CT4, CT5, and *A. m. intermissa* formed a well-supported group dominated by sub-lineage A_{II}. Sub-lineage A_{III} occurred in high proportions in AT3 and AT4, while sub-lineage A_I was the most frequent from AT7 to AT8, CT6 to CT9, and MT5 to MT6.

Analysis of PCo grouped the Iberian sites and the reference subspecies similarly to the neighbor-joining tree (Figure 4b). The main axis explained 78.5% of the genetic variation, separating the three aforementioned groups. The second axis explained 14.7% of the genetic variation and allowed a better differentiation of lineage C.

4. DISCUSSION

Here, we provide the most comprehensive survey of sequence data of the tRNA^{leu}-cox2 region ever reported for a honey bee subspecies. Over 889 sequences, 77 from GenBank and 812 sequenced in this study, were examined and corresponding haplotype names were revised, following the nomenclature criteria established earlier (Garnery et al. 1993; Franck et al. 1998, 2000; Rortais et al. 2011) and further expanded here (see Supplementary Material).

Sequence analysis of the 742 honey bee colonies sampled in Iberia and North Africa revealed a total of 164 novel haplotypes belonging to A (113) and M (51) lineages, which represents a 188.3% (of 60 previously reported) and 46.8% (of 109 previously reported) increase, respectively (Garnery et al. 1993, 1995, 1998; De la Rúa et al. 1998, 2005; Franck et al. 2001; Collet et al. 2006; Magnus and Szalanski 2010; Szalanski and Magnus 2010; Rortais et al. 2011; Pinto et al. 2012, 2014; Muñoz et al. 2013; Bertrand et al. 2015). Within lineage A, a greater number of novel haplotypes was detected in sub-lineage A_I (69) than in sub-lineages A_{II} (13) and A_{III} (31), representing an increase of 246.4% (of 28 previously reported), 260.0% (of 5 previously reported), and 114.8% (of 27 previously reported), respectively (Collet et al. 2006; De la Rúa et al.

Table I. Diversity measures for each of the 23 Iberian sites. Sampling sites are indicated from AT1 to MT6

Sampling sites	N	Na	Np	Ne	uh
AT1	30	18	11	13.235	0.956
AT2	30	16	8	10.714	0.938
AT3	30	18	13	13.235	0.956
AT4	30	20	16	12.857	0.954
AT5	33	14	5	8.712	0.913
AT6	39	18	10	12.168	0.942
AT7	30	14	6	9.783	0.929
AT8	30	7	1	1.815	0.464
CT1	30	11	5	6.618	0.878
CT2	30	8	2	3.913	0.770
CT3	30	12	8	7.377	0.894
CT4	30	10	2	4.545	0.807
CT5	30	13	3	7.627	0.899
CT6	33	13	4	7.949	0.902
CT7	30	16	4	12.500	0.952
CT8	30	13	4	6.000	0.862
CT9	24	11	4	4.721	0.822
MT1	36	14	8	7.624	0.894
MT2	30	13	6	5.556	0.848
MT3	30	16	9	7.627	0.899
MT4	33	13	5	5.261	0.835
MT5	33	16	8	6.444	0.871
MT6	30	8	3	4.737	0.816

AT Atlantic transect, *CT* central transect, *MT* Mediterranean transect, *N* number of individuals, *Na* mean number of haplotypes, *Np* number of private haplotypes, *Ne* number of effective haplotypes, *uh* unbiased haplotype diversity

1998, 2005; Garnery et al. 1993, 1995, 1998; Franck et al. 2001; Magnus and Szalanski 2010; Szalanski and Magnus 2010; Pinto et al. 2012; Muñoz et al. 2013).

While the PCR-RFLP *Dra* I test has proved to be efficient in mtDNA identification of colonies collected around the world (Meixner et al. 2013), sequencing is required to fully describe *Dra* I haplotypes. Despite the numerous mtDNA surveys that have been carried out in Iberia (Smith et al. 1991; Garnery et al. 1995, 1998; Franck et al. 1998, 2001; De la Rúa et al. 2005; Cánovas et al. 2008; Pinto et al. 2012, 2013; Chávez-Galarza et al. 2015) and North Africa (Franck et al. 2001; Chahbar et al. 2013; Achou et al. 2015), analysis of sequence data revealed 28 novel *Dra* I band patterns (1 from North African samples and 27 from Iberian samples). Haplotype variants carrying very short indels in large fragments and nucleotide substitutions undergo undetected in polyacrylamide or high resolution agarose gels. In these cases, sequencing is recommended because it better describes haplotypes and resolves cryptic diversity patterns. For example, previous *Dra* I surveys described A2 as the most frequent haplotype in Iberia, which was mainly observed in the South of Spain (Franck et al. 1998; Miguel et al. 2007; Cánovas et al. 2008). However, our sequence data revealed the presence of two haplotypes, A2 and A2a, with similar frequencies in this region. Of these two, haplotype A2 predominates in the South of the central transect, and A2a in the South of the Mediterranean transect (see Table S3). Further, while A2a is as abundant as A2, the latter exhibits a central position in the network and displays more connections with other haplotypes (Figure 3) suggesting that it is a more ancestral haplotype. In addition to finding 28 novel *Dra* I band patterns (haplotypes defined by an Arabic numeral), the sequencing approach uncovered 154 novel variants (haplotypes defined by a lower-case letter following the Arabic numeral), which represents a substantial addition to the diversity levels reported for Iberia and North Africa.

Our results reinforce the hypothesis of a hybrid origin for *A. m. iberiensis* originating from a process of secondary contact (Smith et al. 1991; Chávez-Galarza et al. 2015). The haplotypes form a well-defined M-A cline (Figure 1) and a complex network (Figure 3), which can be explained by multiple origins of haplotypes (Crandall and Templeton 1993) and by a more recent history of diversification of haplotypes belonging to African sub-lineages. The links with large genetic distances that separate the most frequent M haplotypes represent several mutational steps that might have accumulated during the climatic oscillations produced by glacial episodes. Contrary to what is seen in lineage M, lineage A exhibits shorter links for the most frequent haplotypes, suggesting that A haplotypes evolved in a more stable climate and

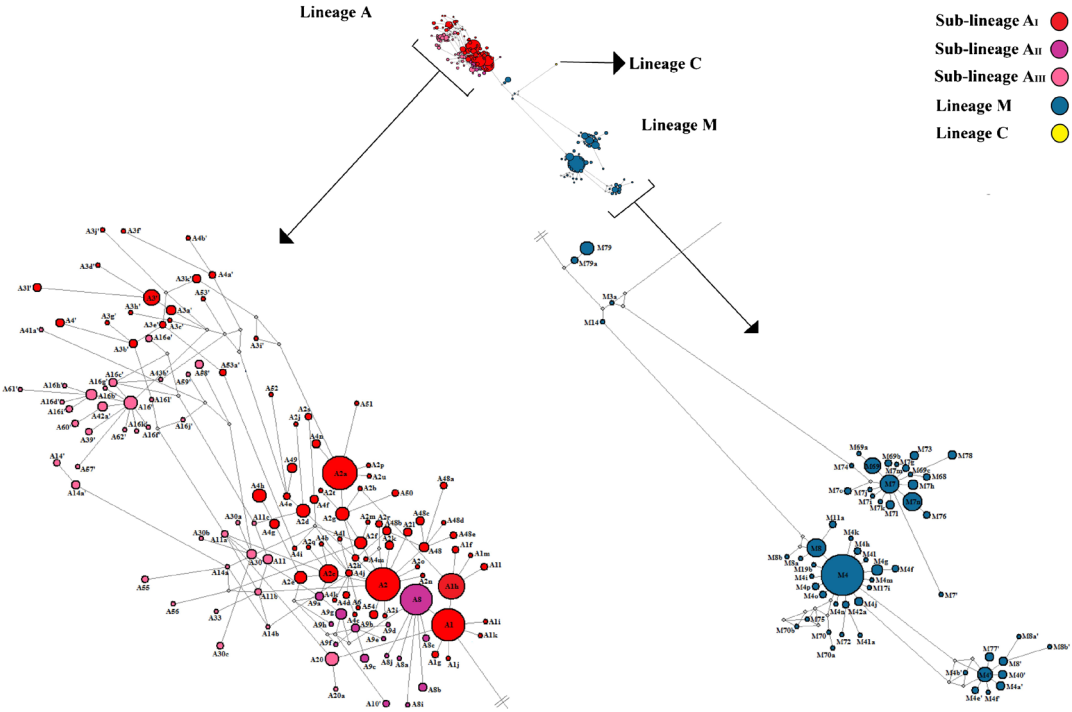
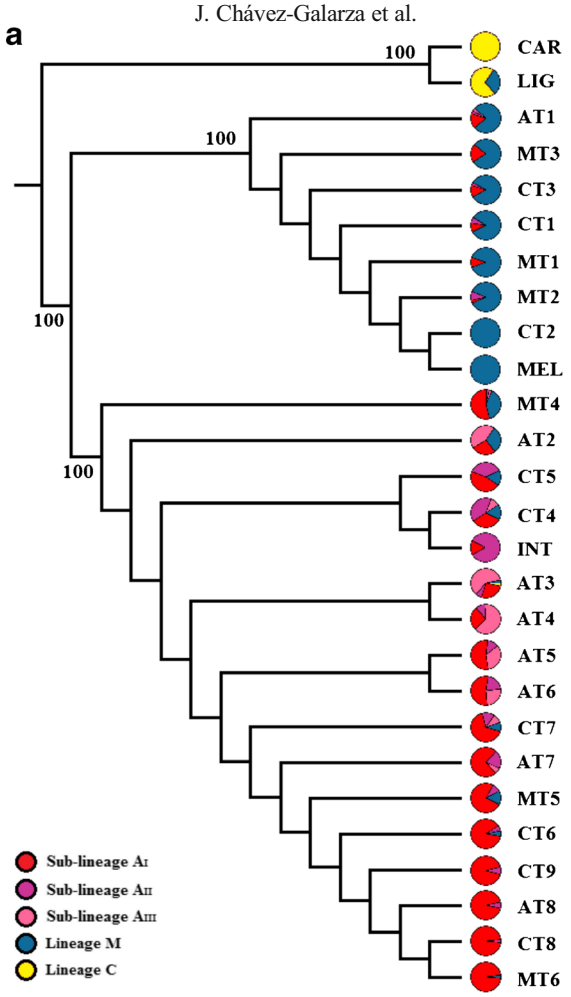


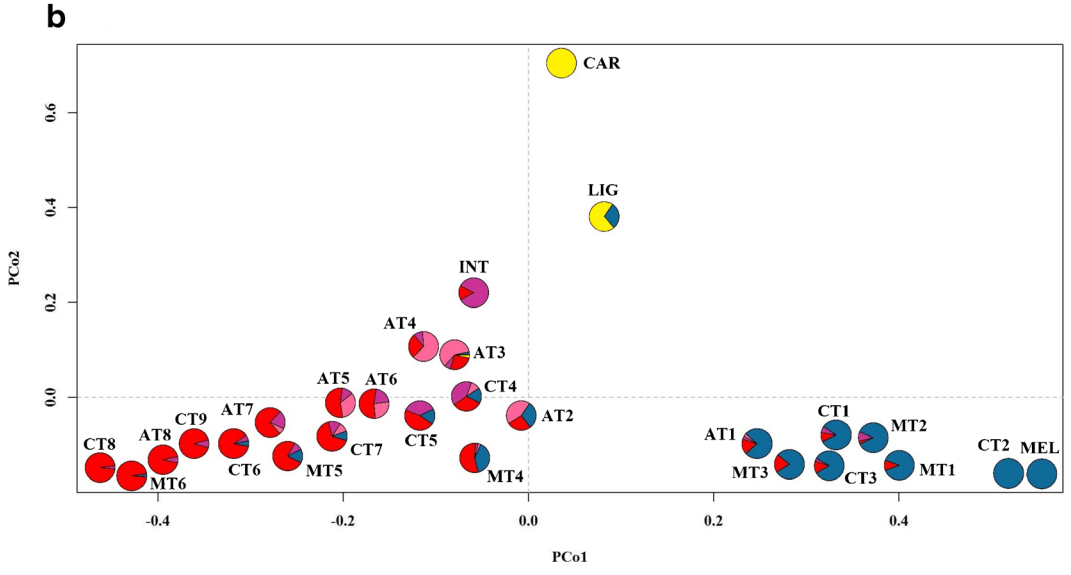
Figure 3. Median-joining network of haplotypes observed in the Iberian Peninsula. The network is zoomed in to better visualize the relationships among haplotypes. Hypothetical (unsampled or extinct) haplotypes are indicated as *gray filled squares*. The size of circles is proportional to haplotype frequencies. Links between haplotypes are proportional to genetic distances between them.

were less affected by the Pleistocenic climatic changes.

We report for the first time haplotypes (M79 and M79a) exhibiting a striking architecture combining features of both lineages M and A (Figure S2). The median-joining network suggests that these haplotypes could have appeared earlier ("ancestral") than the remaining haplotypes (Figure 3). In much the same way, the three African sub-lineages and lineage M could have diverged from an ancestral haplotype represented today by haplotypes of sub-lineage Z origin (Figure 2), a result also supported by *Cyt* b sequences (Alburaki et al. 2011). Within lineage M, the "ancestral" haplotypes formed a well-defined group separated from the other typical M haplotypes (Figure 2). These findings, together with Z haplotypes carried by *A. m. lamarckii* and *A. m. syriaca* populations from Egypt, Syria, and Lebanon (Alburaki et al. 2011), suggest that lineage M could have a more ancestral African origin, with a center located between Northeast Africa and Near East, congruent with the hypothesis of Ruttner et al. (1978). This finding deserves, however, further investigation using coding regions, as the faster evolving intergenic region is more appropriate for discriminating haplotypes (Garnery et al. 1993).

This study expands our knowledge of the complex architecture of the $tRNA^{leu}$-cox2 intergenic region, re-emphasizing that Iberia is an important source of honey bee mtDNA diversity especially of African ancestry (Franck et al. 1998; Cánovas et al. 2008; Pinto et al. 2013). Furthermore, this study supports the utility of sequencing as a complement to the *Dra* I test, not only for characterization of novel band-pattern haplotypes, but also for revealing cryptic variation both of which enable a fuller description of diversity patterns within and among honey bee subspecies.


a
100
100
100
100
CAR
LIG
AT1
MT3
CT3
CT1
MT1
MT2
CT2
MEL
MT4
AT2
CT5
CT4
INT
AT3
AT4
AT5
AT6
CT7
AT7
MT5
CT6
CT9
AT8
CT8
MT6
Sub-lineage AI
Sub-lineage AII
Sub-lineage AIII
Lineage M
Lineage C


b
CAR
LIG
INT
AT4
AT3
AT5
AT6
CT4
AT2
CT7
CT5
AT7
CT9
AT8
CT8
CT6
MT5
MT6
MT4
AT1
CT1
MT2
MT3
CT3
MT1
CT2
MEL
PCo2
PCo1
0.6
0.4
0.2
0.0
-0.4
-0.2
0.0
0.2
0.4

◀ **Figure 4.** Genetic relationships among Iberian sampling sites (*AT* Atlantic transect, *CT* central transect, *MT* Mediterranean transect) and reference subspecies (*CAR A. m. carnica*, *LIG A. m. ligustica*, *MEL A. m. mellifera*, *INT A. m. intermissa*) using ΦPT values obtained with frequency data for lineages and African sub-lineages. Pie charts display the proportions of lineages and sub-lineages. **a** Neighbor-joining tree with support of 1000 bootstraps. **b** PCo presenting 78.5 and 14.7% of total variation for each axis.

ACKNOWLEDGMENTS

We thank numerous beekeepers, technicians of beekeepers' associations, and of government institutions, A. Brandão, I. Moura, I. Muñoz, and M. Neto for assisting with sampling in Iberia. The reference samples were provided by A. Abrahams, B. Dahle, G. Soland-Reckeweg, G. Fert, N. Carreck, P. de la Rúa, R. Dall'Olio, R. Van der Zee, S. Aboulfaraj, and S. Saenz, to whom we are deeply grateful. An earlier version of the manuscript was improved by the constructive comments made by two anonymous reviewers. JC-G and DH were supported by PhD scholarships (SFRH/BD/68682/2010 and SFRH/BD/84195/2012, respectively) from the Portuguese Science Foundation (FCT). MAP is a member of and receives support from the COST Action FA1307 (SUPER-B). Financial support for this research was provided by FCT and COMPETE/QREN/EU through the project PTDC/BIA-BEC/099640/2008 and through the 2013-2014 BiodivERsA/FACCE-JPI joint call for research proposals (138573 - BiodivERsA/0002/2014), with the national funders FCT and the French National Research Agency.

Author contributions MAP conceived the study. JC-G performed the analyses assisted by DH and CJN. WLA provided African samples. JC-G, MAP, and LG interpreted the results. JC-G and MAP wrote the paper with input from JSJ.

Variation de l'ADN mitochondrial d'*Apis mellifera iberiensis* : nouvelles perspectives obtenues à partir d'une étude à grande échelle utilisant les données de séquences de la région intergénique de l'ARNt^{leu}-cox2

Abeille ibérienne / Apidae / haplotype M ancestral / test *Dra* I

Variation der mitochondrialen DNA in *Apis mellifera iberiensis* : neue Erkenntnisse aus einer umfangreichen Untersuchung von Sequenzdaten der tRNAleu-cox2 Zwischengenregion

Iberische Honigbiene / ancestraler Haplotyp M / *Dra* I Test

REFERENCES

Achou, M., Loucif-Ayad, W., Legout, H., Hmidan, H., Alburaki, M., Garnery, L. (2015) An insightful molecular analysis reveals foreign honeybees among Algerian honeybee populations (*Apis mellifera* L.). J. Data Mining Genomics Proteomics **6**, 166.

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.

Arias, M.C., Rinderer, T.E., Sheppard, W.S. (2006) Further characterization of honey bees from the Iberian Peninsula by allozyme, morphometric and mtDNA haplotype analyses. J. Apic. Res. **45**, 188–196.

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

Bertrand, B., Alburaki, M., Legout, H., Moulin, S., Mougel, F., Garnery, L. (2015) MtDNA COI-COII marker and drone congregation area: An efficient method to establish and monitor honeybee (*Apis mellifera* L.) conservation centers. Mol. Ecol. Res. **15**, 673–683.

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.

Cánovas, F., De la Rúa, P., Serrano, J., Galián, J. (2011) Microsatellite variability reveals beekeeping influences on Iberian honeybee populations. Apidologie **42**, 235–251.

Chahbar, N., Muñoz, I., Dall'Ollio, R., De la Rúa, P., Serrano, J., Doumandji, S. (2013) Population structure of North African honey bees is influenced by both biological and anthropogenic factors. J. Insect Conserv. **17**, 385–392.

Chávez-Galarza, J., Henriques, D., Johnston, J.S., Azevedo, J.C., Patton, J.C., Muñoz, I., De la Rúa, P., *et al*. (2013) Signatures of selection in the IHB (*Apis mellifera iberiensis*) revealed by a genome scan analysis of single nucleotide polymorphisms. Mol. Ecol. **22**, 5890–5907.

Chávez-Galarza, J., Henriques, D., Johnston, J.S., Carneiro, M., Rufino, J., Patton, J.C., Pinto, M.A. (2015) Revisiting the IHB (*Apis mellifera iberiensis*) contact zone: maternal and genome-wide nuclear variations provide support for secondary contact from historical refugia. Mol. Ecol. **24**, 2973–2992.

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 **97**, 329–335.

Cornuet, J.M., Fresnaye, J. (1989) Biometrical study of honey bee populations from Spain and Portugal. Apidologie **20**, 93–101.

Crandall, K.A., Templeton, A.R. (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. Genetics **134**, 959–969.

De la Rúa, P., Serrano, J., Galián, J. (1998) Mitochondrial DNA variability in the Canary islands honeybees (*Apis mellifera* L.). Mol. Ecol. **7**, 1543–1547.

De la Rúa, P., Hernandez-Garcia, R., Jiménez, Y., Galián, J., Serrano, J. (2005) Biodiversity of *Apis mellifera iberica* (Hymenoptera: Apidae) from northeastern Spain assessed by mitochondrial analysis. Insect Syst. Evol. **36**, 21–28.

Franck, P., Garnery, L., Solignac, M., Cornuet, J.M. (2000) Molecular confirmation of a fourth lineage in honeybees from the Near East. Apidologie **31**, 167–180.

Franck, P., Garnery, L., Loiseau, A., Oldroyd, B.P., Hepburn, H.R., Solignac, M., Cornuet, J.M. (2001) Genetic diversity of the honey bee in Africa: microsatellite and mitochondrial data. Heredity **86**, 420–430.

Franck, P., Garnery, L., Solignac, M., Cornuet, J.M. (1998) The origin of west European subspecies of honeybees (*Apis mellifera*): new insights from microsatellite and mitochondrial data. Evolution **52**, 1119–1134.

Felsenstein, J. (1993) Phylogeny Inference Package (PHYLIP). Version 3.5. University of Washington, Seattle.

Garnery, L., Franck, P., Baudry, E., Vautrin, D., Cornuet, J.M., Solignac, M. (1998) Genetic diversity of the west European honey bee (*Apis mellifera mellifera* and *A. m. iberica*). I. Mitochondrial DNA. Genet. Select. Evol. **30**, S31–S47.

Garnery, L., Mosshine, E.H., Oldroyd, B.P., Cornuet, J.M. (1995) Mitochondrial DNA variation in Moroccan and Spanish honey bee populations. Mol. Ecol. **4**, 465–471.

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.

Langella, O. (2011) Populations, version 1.2.32: Population Genetic Software. Unité Mixte de Recherches de Génétique Végétale, Institut National de la Recherche Agronomique, France.

Magnus, R., Szalanski, L. (2010) Genetic evidence for honey bees (*Apis mellifera* L.) of Middle Eastern lineage in the United States. Sociobiology **55**, 285–296.

Meixner, M.D., Pinto, M.A., Bouga, M., Kryger, P., Ivanova, E., Fuchs, S. (2013) Standard methods for characterizing subspecies and ecotypes of *Apis mellifera*. J. Apic. Res. **52**, 1–27.

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 postglacial re-colonization routes in the western Europe. Apidologie **38**, 141–155.

Muñoz, I., Pinto, M.A., De la Rúa, P. (2013) Temporal changes in mitochondrial diversity highlights contrasting population events in Macaronesian honey bee. Apidologie **44**, 295–305.

Nielsen, D., Page, R.E. Jr., Crosland, M.W. (1994) Clinal variation and selection of MDH allozymes in honey bee populations. Experientia **50**, 867–871.

Peakall, R., Smouse, P.E. (2012) GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics **28**, 2537–2539.

Pinto, M.A., Henriques, D., Chávez-Galarza, J. et al. (2014) Genetic integrity of the Dark European honey bee (*Apis mellifera mellifera*) from protected populations: a genome-wide assessment using SNPs and mtDNA sequence data. J. Apic. Res. **53**, 269–278.

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

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.

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.

Ruttner, F. (1988) Biogeography and Taxonomy of Honey Bees. Springer, Berlin.

Ruttner, F., Tassencourt, L., Louveaux, J. (1978) Biometrical-statistical analysis of the geographic variability of *Apis mellifera* L. Apidologie **9**, 363–381.

Smith, D.R., Glenn, T.C. (1995) Allozyme polymorphisms in Spanish honeybees (*Apis mellifera iberica*). J. Hered. **86**, 12–16.

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

Szalanski, L., Magnus, R. (2010) Mitochondrial DNA characterization of Africanized honey bee (*Apis mellifera* L.) populations from the USA. J. Apic. Res. **49**, 177–185.

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.