



Mitochondrial SNP markers to monitor evolutionary lineage ancestry in *Apis mellifera mellifera* conservation programs

Dora HENRIQUES¹, Melanie PAREJO^{2,3}, Ana R. LOPES¹, Maria Alice PINTO¹

¹Centro de Investigação de Montanha, Instituto Politécnico de Bragança, Campus de Santa Apolónia, 5300-253, Bragança, Portugal

²Agroscope, Swiss Bee Research Centre, Bern, Switzerland

³Department of Genetics, Physical Anthropology and Animal Physiology University of the Basque Country (UPV/EHU), Leioa, Spain

Received 20 February 2019 – Revised 17 May 2019 – Accepted 13 June 2019

Apis mellifera mellifera / mtDNA / maternal lineage / SNPs / conservation

The European dark honey bee, *Apis mellifera mellifera*, is threatened in most of its native range, in part, due to introgressive hybridization with bees from the highly divergent C-lineage, mainly *Apis mellifera carnica* and *Apis mellifera ligustica* (De la Rúa et al. 2009; Pinto et al. 2014). Yet, the maintenance of locally adapted genetic diversity is critical for the population long-term survival and sustainability (De la Rúa et al. 2009; Meixner 2010). The growing awareness that genetic diversity is important for sustainable beekeeping led to implementation of different conservation and breeding programs throughout Europe, which are in need of reliable and cost-efficient molecular tools to accurately monitor C-lineage introgression into *A. m. mellifera* (De la

Rúa et al. 2009; Henriques et al. 2018a, b; Meixner 2010). The large mating flight distances and the polyandrous mating system make it challenging to preserve honey bee subspecies in an open conservation area where intruders can fly in (Neumann et al. 1999). It is therefore necessary to regularly control the genetic ancestry of new or superseded colonies.

A number of studies have designed diagnostic tools with nuclear (nc) SNP markers capable of discriminating different honey bee lineages (Chapman et al. 2015, 2017; Muñoz et al. 2015; Parejo et al. 2016). Recently, a set of highly informative ncSNPs for estimating C-derived introgression into *A. m. mellifera* (Muñoz et al. 2015) were combined into four assays, which were tested and validated for genotyping in the Agena BioScience™ iPLEX MassARRAY platform (Henriques et al. 2018a). The number of ncSNPs included in each of the four assays (M1 = 34, M2 = 32, M3 = 28, M4 = 23) is lower than the maximum capacity of the iPLEX protocol (40 SNPs) allowing additional markers, such as mitochondrial SNPs (mtSNPs). By combining SNPs of both mitochondrial and nuclear

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13592-019-00666-z>) contains supplementary material, which is available to authorized users.

Corresponding author: M. Pinto, apinto@ipb.pt. Pinto, apinto@ipb.pt

Dora Henriques and Melanie Parejo contributed equally to this work.

Manuscript editor: Marina Meixner

compartments, a more complete identification is achieved in a single genotyping step.

Analyzing mtDNA, in addition to ncDNA, enables better-informed decision-making in *A. m. mellifera* conservation and breeding programs. Due to independent transmission of nuclear and mitochondrial DNA, a colony carrying C-derived mtDNA can be identified as pure *A. m. mellifera* at the nuclear level, after several generations of backcrossing with *A. m. mellifera* drones (Jensen et al. 2005; Meixner et al. 2013). Then, whether a colony exhibiting cytonuclear incongruence is maintained in a conservation area is a decision to be made by the conservation manager.

The aim of this study was to improve the SNPs assays developed by Henriques et al. (2018a) adding informative mitochondrial SNPs. To this end, we used whole mtDNA sequence data of 155 drones, each representing a single colony, collected across Europe (Table S1) (Henriques et al. 2019; Parejo et al. 2016). Mapping and variant calling were performed following best practices detailed in Henriques et al. (2018b) (see Table S1 for further details). A total of 397 SNPs were identified in the 16,343 bp reference mitochondrial genome (Crozier and Crozier 1993). These were distributed across the 13 protein-coding genes with ATP8 containing the lowest (5) and COX1 the highest number (46) of SNPs (Table S2). Maternal ancestry of the 155 individuals was assessed by a neighbor-joining (NJ) tree constructed in MEGA7 (Tamura et al. 2013). A total of 104 individuals were placed in the M-lineage cluster and 51 in the C-lineage cluster (Table S1; Fig. S1).

To select informative mtDNA markers, we randomly chose 78 M- and 38 C-lineage individuals as training set and 26 M- and 13 C-lineage as holdout set (following Anderson (2010)). F_{ST} values between M- and C-derived individuals were inferred for each mtSNP from the training set using PLINK 1.9 (Chang et al. 2015), and those showing F_{ST} values > 0.75 were selected. Sequence information of the flanking regions (250 bp) of the selected mtSNPs was then used to supplement the four ncSNP assays (Henriques et al. 2018a) using the *Replex* option of the software AssayDesign 4.0 (Agena BioScience™). The software searched for optimal areas within

the flanking regions to design forward, reverse, and iPLEX extension primers compatible with those published in Henriques et al. (2018a).

From a total of 397 mtSNPs, 193 were highly informative ($F_{ST} > 0.75$, Table S2). Of these, the assay design tool was able to incorporate five into the M4 ncSNP assay. The five mtSNPs are located in COX1, ND4, and ls-rRNA (Table I). To evaluate their power for distinguishing M- and C-lineage colonies, a NJ tree was constructed using the holdout data set. The NJ tree shows that the five mtSNPs are sufficient to discriminate colonies of C- from M-lineage maternal ancestry (Fig. S2).

The five informative mtSNPs could only be added to the M4 ncSNP assay designed previously (Henriques et al. 2018a), which contained the lowest number of ncSNPs (23). Henriques et al. (2018a) showed that assay M1 (34 ncSNPs) has the best individual performance in estimating nuclear C-lineage introgression. Accordingly, a good approach for concurrent inference of nuclear and mitochondrial ancestry could be to combine M1 with the new 28-plex M4 (23 ncSNPs plus 5 mtSNPs). We therefore tested if the introgression proportions (Q values) estimated from the ncSNP assays M1 + M4 were similar to those obtained from the four assays together (Henriques et al. 2018a). From the 117 SNPs included in the four assays, we were able to extract 113 from the 155 whole-genome (WG) sequences. We run ADMIXTURE (Alexander et al. 2009) in the 155 individuals using the four assays combined (113 ncSNPs) and M1 + M4 (55 ncSNPs, as 2 of them were not in the WG sequences). Q values were estimated for $K = 2$ using 10,000 iterations in 20 independent runs. Q values estimated with M1 + M4 were highly correlated with those estimated with the 4 assays ($r = 0.995$, P value < 0.001), and the differences between them were negligible (mean absolute difference = 0.027; Table S3).

There is a need of accurate identification of C-lineage introgressed colonies for an efficient management of *A. m. mellifera* conservation programs. A more complete identification requires informative markers of both nuclear and mitochondrial compartments (Meixner et al. 2013). For example, in this study, cytonuclear incongruence is observed in two *A. m. ligustica* (C-lineage) individuals that carry mitochondria

Table 1. Genomic information of the five selected SNPs to distinguish C and M maternal lineages

Locus name	Position	Gene	Annotation	F_{ST} (total*)	F_{ST} (training*)
MT_2447_CT	2447	COX1	Synonymous	1.000	1.000
MT_2687_TC	2687	COX1	Synonymous	1.000	1.000
MT_13513_GA	13513	ls-rRNA	–	0.732	0.764
MT_9361_CT	9361	ND4	Nonsynonymous	0.987	0.982
MT_2939_TA	2939	COX1	Synonymous	1.000	1.000

*Estimated using the total and training sets

of M-lineage ancestry (Table S3). Here, we improved the recently designed M4 ncSNP assay (Henriques et al. 2018a) by adding five mtSNPs and provided PCR and iPLEX primers for genotyping in the MassARRAY platform (Table S4). We showed that C-lineage introgression into *A. m. mellifera* can be accurately estimated by combining the new cytonuclear M4 assay improved here with the best performing M1 nuclear assay developed previously (Henriques et al. 2018a). The combined M1 + M4 assays, containing 57 nuclear and five mitochondrial SNPs, provide a robust molecular tool for assisting management decisions toward protection and reestablishment of the endangered *A. m. mellifera*.

ACKNOWLEDGMENTS

We thank João Costa (Instituto Gulbenkian de Ciência, Oeiras, Portugal) for redesigning the M4 assay. Analyses were performed using computational resources at the Instituto Politécnico de Bragança.

Les marqueurs SNP mitochondriaux pour surveiller l'ascendance de la lignée évolutive dans les programmes de conservation de *Apis mellifera mellifera*

Wissenschaftliche Notiz: Mitochondriale SNP Marker zur Überwachung der evolutionären Herkunftslinie in Erhaltungsprogrammen von *Apis mellifera mellifera*

AUTHOR CONTRIBUTIONS

MAP and DH conceived the ideas and designed the methodology. DH and MP performed most of the analyses with assistance of ARL. MP and DH wrote the manuscript with input from MAP. All the authors critically reviewed the manuscript for important intellectual content.

FUNDING INFORMATION

This work was financed by FEDER (Fundo Europeu de Desenvolvimento Regional) through the program COMPETE 2020–POCI (Programa Operacional para a Competitividade e Internacionalização) and by Portuguese funds through FCT (Fundação para a Ciência e a Tecnologia) in the framework of the project BeeHappy (POCI-01-0145-FEDER-029871). Melanie Parejo was supported by a mobility fellowship awarded from the Swiss National Science Foundation (SNSF).

COMPLIANCE WITH ETHICAL STANDARDS

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

REFERENCES

Alexander, D. H., Novembre, J., Lange, K. (2009) Fast model-based estimation of ancestry in unrelated individuals *Genome research* 19:1655–1664

- Anderson, E. (2010) Assessing the power of informative subsets of loci for population assignment: standard methods are upwardly biased *Molecular Ecology Resources* 10:701–710
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., Lee, J. J. (2015) Second-generation PLINK: rising to the challenge of larger and richer datasets *Gigascience* 4:7
- Chapman, N. C., Harpur, B. A., Lim, J., Rinderer, T. E., Allsopp, M. H., Zayed, A., Oldroyd, B. P. (2015) A SNP test to identify Africanized honeybees via proportion of ‘African’ ancestry *Molecular Ecology Resources* 15:1346–1355
- Chapman, N. C., Bourgeois, A. L., Beaman, L. D., Lim, J., Harpur, B. A., Zayed, A., Allsopp, M. H., Rinderer, T. E., Oldroyd, B. P. (2017) An abbreviated SNP panel for ancestry assignment of honeybees (*Apis mellifera*) *Apidologie* 48:776–783
- Crozier, R., Crozier, Y. (1993) The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization *Genetics* 133:97–117
- De la Rúa, P., Jaffé, R., Dall’Olio, R., Muñoz, I., Serrano, J. (2009) Biodiversity, conservation and current threats to European honeybees *Apidologie* 40:263–284
- Henriques, D., Browne, K. A., Barnett, M. W., Parejo, M., Kryger, P. et al. (2018a) High sample throughput genotyping for estimating C-lineage introgression in the dark honeybee: an accurate and cost-effective SNP-based tool *Scientific Reports* 8:8552
- Henriques, D., Parejo, M., Vignal, A., Wragg, D., Wallberg, A., Webster, M. T., Pinto, M. A. (2018b). Developing reduced SNP assays from wholegenome sequence data to estimate introgression in an organism with complex genetic patterns, the Iberian honeybee (*Apis mellifera iberiensis*). *Evolutionary applications*, 11(8), 1270–1282.
- Henriques, D., Chávez-Galarza, J., Quaresma, A., Neves, C. J., Lopes, A. R., Costa, C., Costa, F. O., Rufino, J., Pinto, M. A. (2019) From the popular tRNA leu-COX2 intergenic region to the mitogenome: insights from diverse honey bee populations of Europe and North Africa. *Apidologie*, 50(2), 215–229.
- Jensen, A. B., Palmer, K. A., Boomsma, J. J., Pedersen, B. V. (2005) Varying degrees of *Apis mellifera ligustica* introgression in protected populations of the black honeybee, *Apis mellifera mellifera*, in northwest Europe *Molecular Ecology* 14:93–106
- Meixner, M. D. (2010) A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them *Journal of Invertebrate Pathology* 103:S80–S95
- 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* *Journal of Apicultural Research* 52:1–28
- Muñoz, I., Henriques, D., Johnston, J. S., Chávez-Galarza, J., Kryger, P., Pinto, M. A. (2015) Reduced SNP panels for genetic identification and introgression analysis in the dark honey bee (*Apis mellifera mellifera*) *PLoS ONE* 10:e0124365
- Neumann, P., Moritz, R. F., van Praagh, J. (1999) Queen mating frequency in different types of honey bee mating apiaries *Journal of Apicultural Research* 38:11–18
- Parejo, M., Wragg, D., Gauthier, L., Vignal, A., Neumann, P., Neuditschko, M. (2016) Using whole-genome sequence information to foster conservation efforts for the European dark honey bee, *Apis mellifera mellifera* *Frontiers in Ecology and Evolution* 4:140
- Pinto, M. A., Henriques, D., Chávez-Galarza, J., Kryger, P., Garnery, L. 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 *Journal of Apicultural Research* 53:269–278 doi:<https://doi.org/10.3896/IBRA.1.53.2.08>
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0 *Molecular Biology and Evolution* 30:2725–2729

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.