Autosomal and mitochondrial adaptation following admixture: a case study on the honey bees of Reunion Island

David Wragg, Maéva Angélique Techer, Kamila Canale-Tabet, Benjamin Basso, Jean Pierre Bidanel, Emmanuelle Labarthe, Olivier Bouchez, Yves Le Conte, Johanna Clémencet, Hélène Delatte, et al.

Autosomal and Mitochondrial Adaptation Following Admixture: A Case Study on the Honeybees of Reunion Island

David Wragg[1,2], Maéva Angélique Techer[3,4,5], Kamila Canale-Tabet[1], Benjamin Basso[6], Jean-Pierre Bidanel[7], Emmanuelle Labarthe[1], Olivier Bouchez[8], Yves Le Conte[9], Johanna Clémencet[4], Hélène Delatte[3], and Alain Vignal[1,*]

[1]GenPhySE, Université de Toulouse, INRA, INPT, INP-ENVT, Castanet Tolosan, France

[2]The Roslin Institute, University of Edinburgh, Midlothian, United Kingdom

[3]CIRAD, UMR PVBMT, Saint Pierre, La Réunion, France

[4]UMR PVBMT, Université de La Réunion, Saint Pierre, La Réunion, France

[5]Ecology and Evolution Unit, Okinawa Institute of Science and Technology Graduate University, Kunigami-gun, Okinawa, Japan

[6]Institut de l'abeille (ITSAP), UMT PrADE, Avignon, France

[7]GABI, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France

[8]INRA, US 1426, GeT-PlaGe, Genotoul, Castanet-Tolosan, France

[9]INRA, UR 406 Abeilles et Environnement, UMT PrADE, Avignon, France

*Corresponding author: E-mail: alain.vignal@inra.fr.

Accepted: November 28, 2017

Data deposition: Sequences for this project have been deposited in the Sequence Read Archive (SRA) at www.ncbi.nlm.nih.gov/sra under the accession PRJNA311274, as part of the SeqApiPop dataset.

Abstract

The honeybee population of the tropical Reunion Island is a genetic admixture of the *Apis mellifera unicolor* subspecies, originally described in Madagascar, and of European subspecies, mainly *A. m. carnica* and *A. m. ligustica*, regularly imported to the island since the late 19th century. We took advantage of this population to study genetic admixing of the tropical-adapted indigenous and temperate-adapted European genetic backgrounds. Whole genome sequencing of 30 workers and 6 males from Reunion, compared with samples from Europe, Madagascar, Mauritius, Rodrigues, and the Seychelles, revealed the Reunion honeybee population to be composed on an average of 53.2 ± 5.9% *A. m. unicolor* nuclear genomic background, the rest being mainly composed of *A. m. carnica* and to a lesser extent *A. m. ligustica*. In striking contrast to this, only 1 out of the 36 honeybees from Reunion had a mitochondrial genome of European origin, suggesting selection has favored the *A. m. unicolor* mitotype, which is possibly better adapted to the island's bioclimate. Local ancestry was determined along the chromosomes for all Reunion samples, and a test for preferential selection for the *A. m. unicolor* or European background revealed 15 regions significantly associated with the *A. m. unicolor* lineage and 9 regions with the European lineage. Our results provide insights into the long-term consequences of introducing exotic specimen on the nuclear and mitochondrial genomes of locally adapted populations.

Key words: adaptation, genomics/proteomics, insects, molecular evolution.

Introduction

Apis mellifera unicolor (Latreille 1804) is a honeybee subspecies endemic to Madagascar, belonging to a haplogroup endemic to the South West Indian Ocean (SWIO) islands, which includes Madagascar, the Mascarene archipelago (Reunion, Mauritius, and Rodrigues islands), and the Seychelles and Comoros archipelagos (Techer et al. 2017). Its mitochondrial haplotype, or "mitotype," groups with other subspecies of honeybee of the African (A) evolutionary lineage. Other lineages comprising distinct subspecies include western and northern Europe (M), eastern Europe (C), the Near East and central Asia (O), and Yemen and Ethiopia (Y), which diverged

>150 thousand years ago (Ruttner 1988; Franck et al. 2001; Harpur et al. 2014; Wallberg et al. 2014). The presence of the honeybee on Reunion prior to the arrival of man is still subject to debate. In his book on traveling in Madagascar and Reunion, Du Bois (1674) mentions that honeybees had been imported to Reunion in 1666, probably from Fort Dauphin in Madagascar shortly after the first French settlements, had successfully multiplied, and that it was easy to find honey in the woods, whereas Hermann (1920) suggests honeybees were present on the island before man settled there. Since the end of the 19th century, in order to develop beekeeping, the principal honeybee subspecies imported from Europe were *A. m. ligustica* and/or *A. m. carnica*, both of which are praised for their high honey production, low levels of swarming, and their calmness (Franck et al. 2000), but also included to a lesser extent *A. m. caucasica*. However, to prevent the spread of the microsporidia *Nosema* spp. the importation of honeybees was prohibited in 1982, which had until recently (May 2017) successfully maintained Reunion honeybees free from other parasites such as *Varroa destructor*. Accordingly, since 1982, the honeybee population on Reunion is believed to have remained free from genetic flow originating from outside of the island. During this time, the genomes of the local subspecies *A. m. unicolor* and the imported European subspecies *A. m. carnica* and/or *A. m. ligustica* have had the opportunity to naturally recombine over many generations, leading to the emergence of a novel population. Prior to the introduction of hives with mobile frames to Reunion in the 1960s, honey production involving natural swarms was extensive, a practice which still continues to this day to some degree. Management practices also involve transhumance, whereby hives are moved several times per year around the island by the beekeepers as they follow the availability of resources. Reunion honeybees therefore provide an excellent opportunity to investigate natural selection in a hybrid population founded by genetic backgrounds of temperate and tropical origin.

Reunion has a humid tropical climate with two seasons: austral summer from November to April, characterized by intense rainfalls and cyclones, followed by the drier austral winter. Maps indicating monthly averages for temperature, precipitation, and wind speed are provided in supplementary figure S1, Supplementary Material online. The topography of the island includes two volcanoes, one of which remains active, that culminate at 2,636 and 3,071 m, respectively. The island has very variable climate zones within its ~2,500 km^2 surface contributing to an exceptional level of biodiversity, with 28% of plant species being endemic to Reunion, whereas 48% are exotic species having been introduced by settlers (l'index de la flore vasculaire de La Reunion; http://flore.cbnm.org; last accessed November 2017). Agriculture, secondary vegetation and urban areas have transformed almost half of the island, mostly in the west coast lowlands, whereas only one-third of indigenous habitat where invasion by introduced floral species is localized still remains (Strasberg et al. 2005).

A study of the genetic structure of the Reunion honeybee population, sampling >2,000 honeybees throughout the island and following standard methods (Evans et al. 2013) employing microsatellite markers and sequences of the *tRNAleu-cox2* hyper-variable region, revealed an absence of population structure, despite the island's various climate zones (Techer MA, in press). This contrasts for instance with the east–west distribution of genetic clusters of the melon fly *Bactrocera cucurbitae* (Jacquard et al. 2013). However, the absence of structure might be explained by the practice of transhumance.

To gain a detailed insight of the genetic structure of the Reunion honeybee population, and to investigate mechanisms of admixture and adaptation acting on the nuclear and mitochondrial genomes, we performed a study by whole genome sequencing in which honeybees from Reunion were compared with African, European, and other SWIO samples from Madagascar, Mahé of the Seychelles and Mauritius. In particular, local ancestry inference along nuclear haplotypes was performed in Reunion with reference to the indigenous A-type *A. m. unicolor* and introducted C-type *A. m. carnica* and *A. m. ligustica* subspecies, followed by tests of heterogeneity to identify regions under putative selection.

Materials and Methods

Sampling

Drones, being haploid, provide the ideal data for haplotype analyses. However, it was not feasible to source sufficient numbers of drones from all populations/subspecies to undertake a drone-specific study. An effective alternative is to sample haploids where possible, and to use their haplotypes as references to phase the genotypes of diploid workers. Honeybees sampled from Reunion comprised individual haploid drones ($n = 6$) and diploid workers ($n = 30$), sampled from different managed colonies distributed throughout the island (fig. 1). These were supplemented by drones from surrounding islands of the SWIO including Rodrigues ($n = 2$), Mahé in the Seychelles ($n = 2$), Mauritius ($n = 2$), and Madagascar ($n = 6$), and by a number of reference honeybees representing *A. m. carnica* (Germany, $n = 3$ haploid; Slovenia, $n = 3$ haploid; Croatia, $n = 4$ diploid), *A. m. ligustica* (Italy, $n = 10$ haploid); *A. m. caucasia* (France, imported from Georgia, $n = 10$ haploid); *A. m. mellifera* (France, $n = 6$ haploid; Poland, $n = 2$ diploid; Spain, $n = 2$ diploid); *A. m. scutellata* (South Africa, $n = 10$ diploid); and *A. m. jemenitica* (Saudi Arabia, $n = 7$ diploid; Yemen, $n = 3$ diploid). It is worth noting that in Spain there is a well-documented north-east to south-west geographical cline of M to A lineage genetic background in honeybees (Chávez-Galarza et al. 2015). The two *A. m. mellifera* samples from Spain were reported by

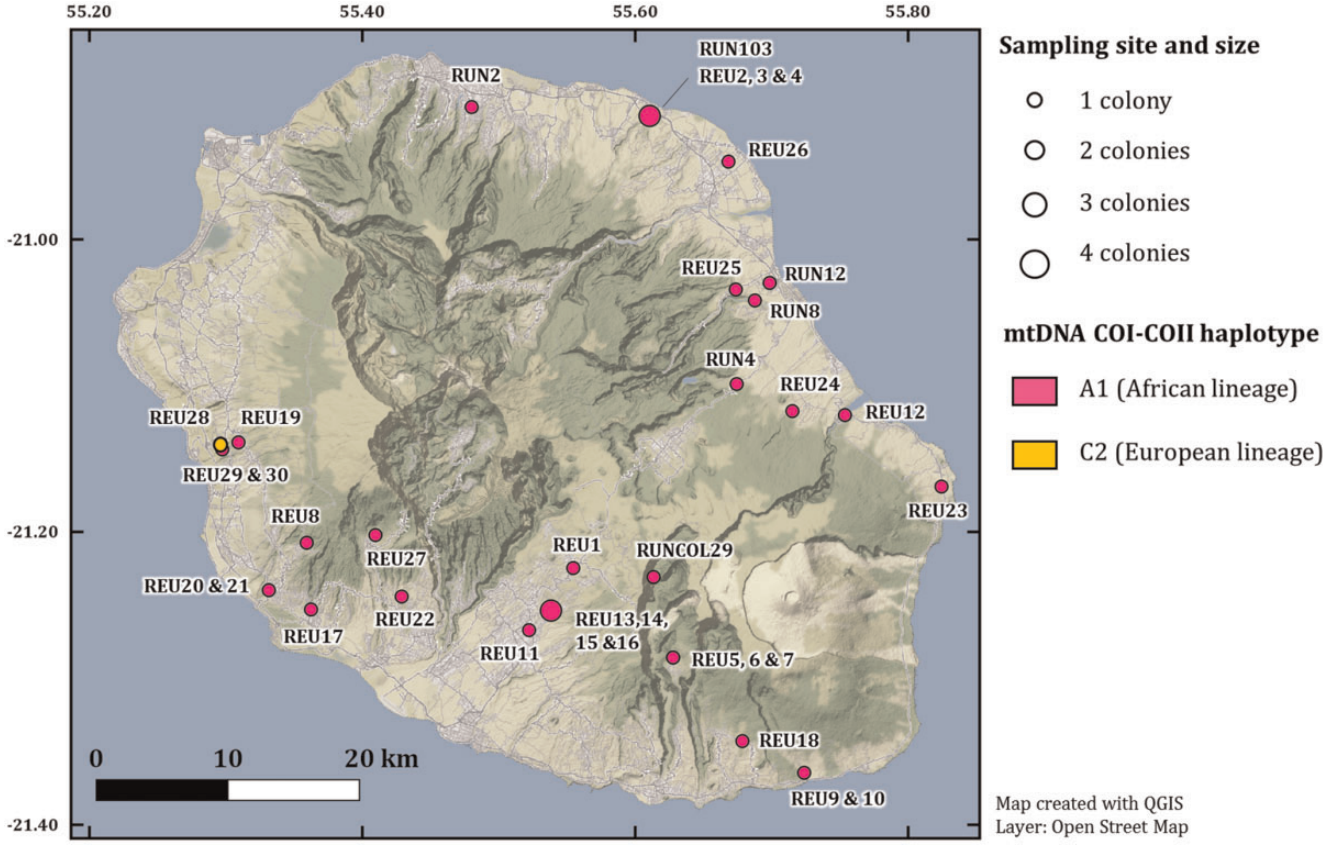

Fig. 1.—Map of Reunion Island indicating sampling locations. The size of the circle indicates the number of colonies sampled. Pink: A1 (African lineage) mitochondrial DNA; yellow: C2 (European lineage) mitochondrial DNA.

Harpur et al. (2014) to have admixture proportions equivalent to *A. m. mellifera* samples from Poland in a data set which also comprised African samples. However, given that they were sampled from the south of Spain (Cordoba and Murcia) there is the possibility that they are in fact *A. m. iberiensis*. For the purposes of this study, we have grouped them with the *A. m. mellifera* samples as both subspecies are derived from the M lineage (Ruttner 1988). With the exception of the samples from Reunion, all other diploid samples were sequenced by Harpur et al. (2014), and downloaded from NCBI. The DNA of samples sequenced in this study was extracted from the thorax as described in Wragg et al. (2016).

Sequencing, Mapping, and Variant Detection

Pair-end sequencing was performed on Illumina HiSeq platforms, with 20 samples per lane, following the manufacturer's protocols for library preparations. The data downloaded from NCBI were generated from single-end libraries on the Illumina HiSeq 2000, and so all sequence data used in this study were produced using the same sequencing by synthesis technology. Sequencing reads were mapped to the reference genome (Amel 4.5) using BWA-MEM v0.7.9a (Li H. 2013: Aligning sequence reads, clone sequences and assembly contigs with BWA MEM. ArXiv13033997 Q-Bio. http://arxiv.org/abs/1303.3997; last accessed November 2017), and processed with regards to marking duplicates, local realignment, base quality score recalibration, and variant detection as described in Wragg et al. (2016). Sequencing depth (DP) ranged from 6 to 48× across the data sets with a median of 17, full details per sample are provided in supplementary table S1, Supplementary Material online.

Variant sites identified across the samples from the SWIO were merged and filtered on depth of coverage to generate a list of autosomal sites for that data set, with $9 \leq DP \leq 3\mu DP$ in which $\mu =$ average DP across samples on a site-by-site basis calculated at run-time, for which these samples were regenotyped with the GATK UnifiedGenotyper (McKenna et al. 2010); base quality BQ ≥ 20), resulting in 9,077,645 SNPs. Plink (v1.9; https://www.cog-genomics.org/plink2; last accessed November 2017; Chang et al. 2015) was used to filter these SNPs on minor allele frequency (MAF) ≥ 0.01, individual and genotyping call rates ≥ 0.9, resulting in 3,704,546 autosomal SNPs which were used as a panel for genotyping the non-SWIO samples. A subset of "unlinked" SNPs were generated from these 3.7 M SNPs, by filtering to remove SNPs < 1 kb apart, and those with pairwise linkage

disequilibrium (LD) $r^2 > 0.1$ for SNPs spaced <50 kb apart, resulting in 19,888 SNPs. This subset of unlinked SNPs was used to investigate the autosomal population genetics of Reunion honeybees so as to avoid overstratification of the data.

Mitochondrial Analyses

Consensus mitochondrial whole-genome sequences were generated for each individual using SAMTOOLS and BCFTOOLS. Sequence reads aligned to the mitochondrial genome were extracted, and together with unmapped reads were aligned to the mitochondrial genome of *A. m. scutellata* (GenBank Acc. No. KJ601784). The best reference from which to generate a consensus sequence for each sample was determined as that with the least error rate for high quality aligned reads (mapping quality ≥ 20) as reported by Picard's CollectAlignmentSummaryMetrics tool (https://broadinstitute.github.io/picard; last accessed November 2017). Sequences were concatenated into a multiple fasta and aligned with ClustalW (Larkin et al. 2007) to generate a nexus file for plotting a network in PopArt (Leigh et al. 2015). As described in Wragg et al. (2016), to reconstruct *tRNAleu-cox2* sequence mitotypes, sequencing reads were extracted from the BAM alignment for the region, aligned against 226 mitotype reference sequences downloaded from GenBank, and the reference mitotype to which the reads aligned best was then used as a reference for de novo assembly with MITObim v1.8; (Hahn et al. 2013). The resulting fasta files were manually aligned in AliView (Larsson 2014). In the resulting alignment, at positions where insertions/deletions (indels) were present, "G" nucleotides were recoded to an otherwise absent nucleotide (e.g., "C") and "missing" nucleotides at these positions recoded as "G" to facilitate their inclusion in network analysis with PopArt, which would otherwise exclude sites with 5% missing rate. Individual mitochondrial gene trees were made using FastTree 2.1.8 (Price et al. 2010) with default parameters and drawn with ggtree (Yu et al. 2017). McDonald–Kreitman tests were performed for each of the 13 coding mitochondrial genes separately, and Tajima's *D* tests were performed on the whole mitochondrial genome of all Reunion bees having *A. m. unicolor* mitochondrial DNA using PopGenome (Pfeifer et al. 2014). Mitotype sequences were verified for 42 SWIO honeybees by Sanger sequencing of PCR amplifications using primers E2 and H2 (Garnery et al. 1993). Sanger sequences were processed using FinchTV (https://digitalworldbiology.com/FinchTV; last accessed November 2017) and aligned manually in AliView against the de novo assembled mitotypes. Whole mitochondrial DNA sequences were assigned to groups using the partitioning around medoids clustering method as implemented in the fpc R package (Hennig C. 2015: fpc: Flexible Procedures for Clustering. https://CRAN.R735project.org/package=fpc; last accessed November 2017.). This method was chosen for its robustness toward the presence of outliers, reducing the likelihood of overestimating the number of clusters.

Autosomal Population Genetics Analyses

To assess the general population structure of the data, principal components analysis (PCA) was conducted in R using the glPca function of the adegenet package (Jombart 2008; Jombart and Ahmed 2011) from a distance matrix, where distance is expressed as genomic proportions (1-identity by state) as generated in Plink. Using the same distance matrix, a complimentary neighbor-joining (NJ) tree was constructed using the NJ function of the phangorn package for R (Schliep 2011), which performs the NJ tree estimation of Saitou and Nei (Saitou and Nei 1987).

A more detailed assessment of population structure was performed using ADMIXTURE v1.23 (Alexander et al. 2009), assuming no prior knowledge of population origin, testing $2 \leq K \leq 16$. The likelihood of the results was determined from the cross-validation (CV) error values, and Q estimates plotted in R. Network analyses by *k*-nearest neighbor (*k*NN) were conducted using NetView (Neuditschko et al. 2012) for R (https://github.com/esteinig/netview; last accessed November 2017). *k*NN graphs were generated assuming $2 \leq k \leq 100$, from which networks were constructed for each *k* where the number of communities (*n*) reached a temporary plateau, indicating the most visually informative network at $k = 41$. This network was plotted incorporating the Q data from the $K = 7$ ADMIXTURE analysis, which does not impact on the structure of the network but simply replaces the nodes with pie charts of the ADMIXTURE results for each sample. TreeMix (Pickrell and Pritchard 2012) analyses were conducted on stratified allele frequency data generated by Plink. This software is designed for the inference of patterns of population splitting and mixing from genome-wide allele frequency data. It calculates the maximum likelihood tree for the set of populations, and optionally attempts to infer a number of admixture events. Based on the results of the PCA and kNN analyses, *A. m. caucasia* was specified as the root population, and the TreeMix analysis conducted in 100 SNP windows with sample size correction, assuming a single migration event, for 100 iterations. The results were parsed and plotted in R.

Autosomal Signatures of Selection Analyses

To investigate whether or not the distribution of the principal genetic backgrounds identified (*A. m. unicolor* and C lineage) in the Reunion population was heterogenous, the data were preprocessed for PCAdmix (Brisbin et al. 2012) and Chromosomal Ancestry Differences (CAnD) analyses (McHugh et al. 2016). The initial 3.7-M SNP data were first filtered to retain only samples from Reunion, together with the haploid *A. m. carnica, A. m. ligustica*, and *A. m. unicolor* (Madagascar and Mauritius) samples. The data were further

filtered to retain only SNPs with 100% call rate, resulting in 1,441,078 SNPs. By using SNPs with 100% call rate, we avoid the risk of introducing erroneous genotypes through imputation which can influence the results of downstream analyses. The diploid Reunion samples were subsequently phased using shapeit2 (Delaneau et al. 2011), using the haploid data as reference haplotypes. Shapeit2 is very insensitive to the effective population size parameter carried forward by the imputation algorithm incorporated from impute2, and so an arbitrary value of 100 k was used which is lower than the *A. mellifera* subspecies range of estimates (157–457 k) from Wallberg et al. (2014). Phasing was performed per chromosome using the minimum permitted window size of 100 kb, and a genetic map estimated from the crossover data generated by Liu et al. (2015). Briefly, Liu et al. sequenced haploid drones from three colonies together with their diploid queen. For each colony, by comparing the linkage of markers across all drones, queens could be phased into haplotypes at the chromosome level. This enabled recombination events in each drone to be identified, from which 3,505 crossovers (spanning >10 kb) were detected. To generate the genetic map, each chromosome was partitioned into 1-Mb intervals and for each interval the recombination rate was calculated from the sum of crossovers observed by Liu et al. divided by the number of drones they sequenced ($n = 43$) multiplied by 100 due to there being 100 meioses per centiMorgan (cM). For each physical position along a chromosome the genetic position was estimated in cM as the recombination rate of the interval in which the position was located, divided by 1 Mb.

PCAdmix classifies blocks of SNPs by ancestry through PCA, projecting the loadings of admixed individuals based on the loadings of putative ancestors. It employs a Hidden Markov Model (HMM) to smooth the results, and returns the posterior probabilities of ancestry affiliation for each block from the HMM. The analysis was run in 1-cM windows using the aforementioned genetic map, with the *A. m. carnica* and *A. m. ligustica* haplotypes as a reference for the C lineage and the *A. m. unicolor* haplotypes from Madagascar and Mauritius as a reference for the A lineage, whereas all Reunion samples were provided as the admixed population. The results were parsed and plotted using R. To detect heterogeneity in ancestry across the individual haplotypes of the genomes of the admixed Reunion population, the posterior probabilities obtained from the HMM for each of the two reference backgrounds were subject to CAnD analysis. As the analysis was performed on probabilities assigned to haplotypes, and not genotypes, the ploidy of the bees is of no consequence. CAnD tests for systematic differences in genetic contributions from ancestral populations to chromosomes in admixed individuals. Regions were considered significant if Bonferonni-corrected $P < 0.001$ and where the mean posterior probability of ancestry inference across samples for the region >0.75. Genes and their orthologs for *Drosophila melanogaster* were identified for significant regions using the *A. mellifera* Amel4_5 and *D. melanogaster* BDGP6 genes data sets of EnsemblMetazoa's BioMart. Unless elaborated upon further, gene functions are broadly summarized according to their gene ontologies, sourced through BioMart and flybase (flybase.org).

Results

Maternal Origins of Reunion Island Honeybees

Analysis of *tRNAleu-cox2* mitotype sequences is presently the most practical and cost-effective means of characterizing sample origin and honeybee subspecies in large numbers of samples. Mitotype sequences were generated from the resequencing data of each of the SWIO samples, associated with previously identified mitotypes by BLAST, and a network constructed (fig. 2*A*). All but one of the Reunion samples (REU28) and the two samples from Rodrigues were found to possess an A1 mitotype, with the exceptions possessing C1 or C2 mitotypes. This lack of mitotype diversity was subsequently verified by Sanger sequencing 42 of the 48 SWIO samples (supplementary table S2, Supplementary Material online). The results support previous findings concerning the near fixation of the A1 mitotype on Reunion (Franck et al. 2001; Techer et al. 2017), and the presence of C1 and C2 mitotypes on Rodrigues (Techer et al. 2015).

Consensus mitochondrial genome sequences were generated for each individual against the mitochondrial sequences of the Amel 4.5 reference and that of an *A. m. scutellata* individual, and the best alignment in each case retained. With the exception of the three honeybees having C1 and C2 mitotypes, all SWIO samples aligned better to the *A. m. scutellata* reference. A network (fig. 2*B*) was constructed from aligned sequences, representing 812 segregating sites. The samples from Reunion, Madagascar, Mauritius, and Mahé (Seychelles) formed a cluster branching out from the other A lineage cluster comprising *A. m. scutellata* samples. Within this cluster, the samples from Madagascar branched out from the Reunion, Mauritius, and Seychelles samples, albeit with a very low number of sites separating the two populations, suggesting the divergence to be much more recent than the one between the SWIO *A. m. unicolor* and the mainland African *A. m. scutellata* honeybees (fig. 2*C*). A clustering analysis by partitioning around medoids confirmed this grouping of samples from the Reunion, Madagascar, Mauritius, and Mahé (Seychelles), and revealed three other clusters: one containing *A. m. scutellata* honeybees, another corresponding to *A. m. mellifera*, and a final cluster containing all remaining honeybees. It should be noted that some exchange of mitochondrial DNA has occurred in Europe, involving *A. m. ligustica* from Italy and some *A. m. mellifera* samples (fig. 3).

The sequences of the 13 mitochondrial protein coding, the 2 rRNA, and the 22 tRNA genes were also analyzed separately, to look for substantive differences within the genes between the different populations. For the 13 protein-

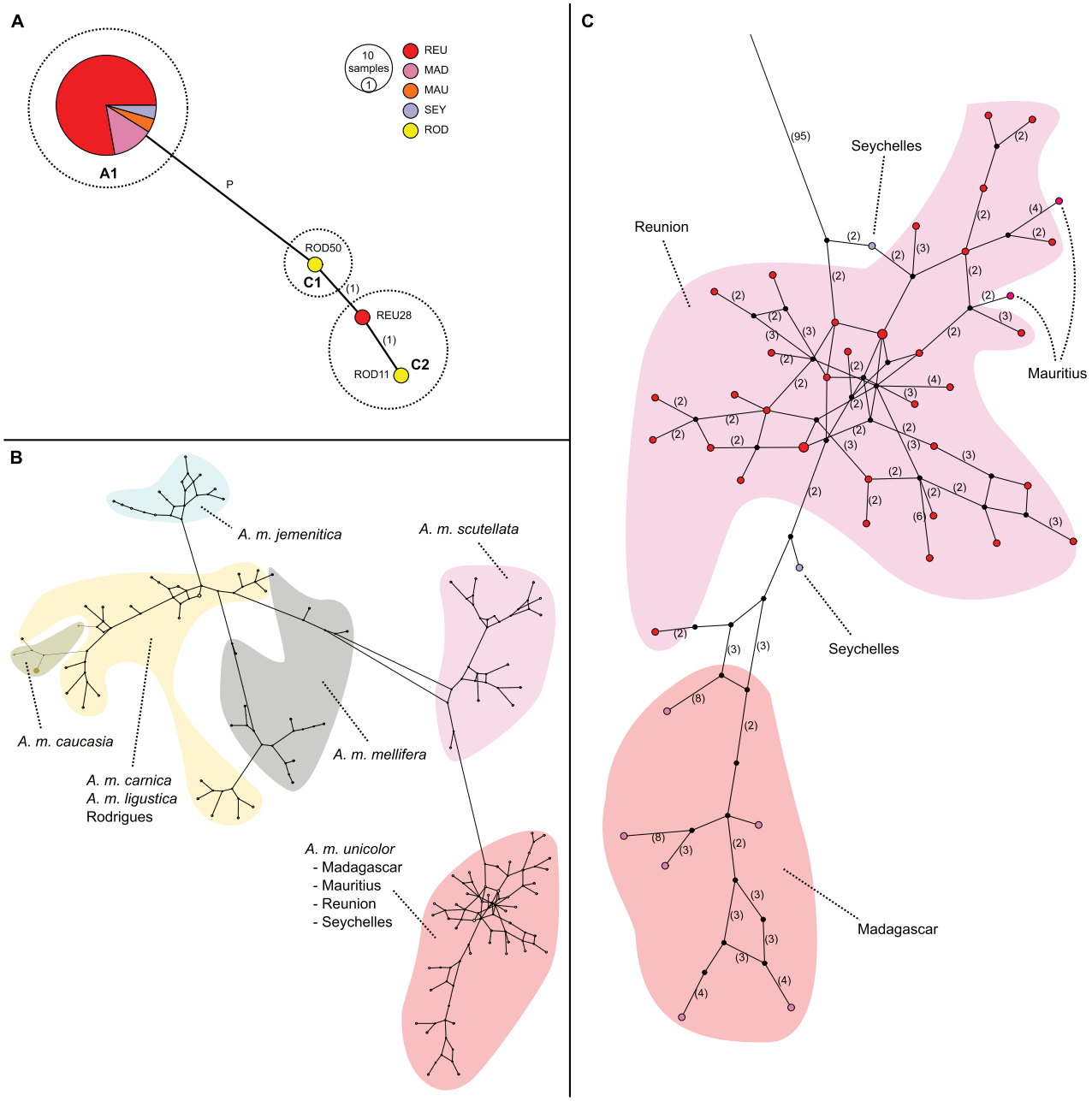

Fig. 2.—Global mitochondrial analyses. (*A*) Barcode sequencing of the *tRNAleu-cox2* mitotype in honeybees from the SWIO. C1 and C2 sequences diverge notably from the A1 sequence, respectively, by the presence or absence of the P element. (*B*) Network of mitochondrial genome sequences of all honeybees in the study. (*C*) Magnified region of network indicating the *Apis mellifera unicolor* cluster. Unless otherwise indicated, network vertices indicate a single point mutation.

coding genes, both DNA and protein sequences were analyzed and network diagrams drawn for all 37 genes (supplementary fig. S2, Supplementary Material online). The results recapitulate the global analysis and indicate the existence of nonsynonymous changes between subspecies. The A1 SWIO mitochondrial gene sequences have a high similarity, suggesting a close link between samples and a relatively low diversity at the scale of the SWIO population, with the notable exceptions of the *cox3* and *nd4l* genes, for which Madagascar has different haplotypes as compared with Reunion, Mauritius, and the Seychelles, both at the DNA and amino-acid levels (fig. 4 and supplementary fig. S2, Supplementary Material online). The *cox3* gene shows a specifically high level of diversity in Madagascar, with four different DNA and amino acid haplotypes observed in the six samples sequenced. As was shown by the *tRNAleu-cox2* mitotype results, the REU28 sample and the two samples from Rodrigues are closer to C-type honeybees and more

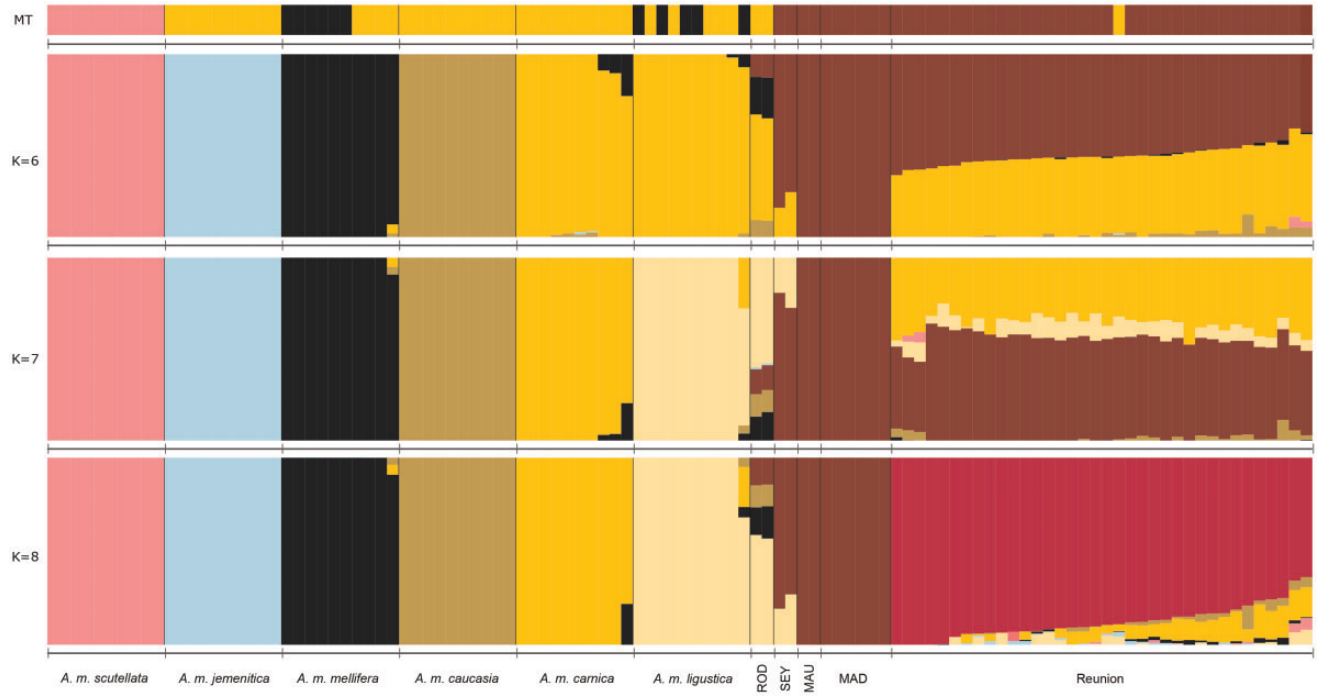

Fig. 3.—ADMIXTURE analysis. The top lane "MT" illustrates mitochondrial DNA assignments to the four clusters identified by partitioning around medoids of mitochondrial genome sequences (see Materials and Methods). Following which are admixture plots for K=6 to K=8 genetic backgrounds. ROD, Rodrigues; SEY, Seychelles; MAU, Mauritius; MAD, Madagascar (*Apis mellifera unicolor*).

specifically to *A. m. carnica* in the case of REU28. Nonsynonymous changes were found in all 13 mitochondrial coding genes, with 35 separating the *A. m. unicolor* population from all other populations (supplementary table S3, Supplementary Material online). To test whether the fraction of fixed nonsynonymous sites was higher than expected, which can be indicative of positive selection, a McDonald–Kreitman test was performed on the 13 mitochondrial DNA coding genes, between the Reunion and Madagascar honeybees having a typical *A. m. unicolor* mitotype (all except REU28) and the C-type *A. m. ligustica* and *A. m. carnica*, the results of which were not significant (supplementary table S4, Supplementary Material online). Many nucleotide differences, both SNPs and indels were also found for the tRNA genes. For instance, nucleotide differences in the *tRNA-pro, tRNA-thr*, and one of the *tRNA-ser* genes separate the SWIO samples from the other subspecies (fig. 4 and supplementary fig. S2, Supplementary Material online). A Tajima's *D* test on the mitochondrial DNA of the Reunion samples having *A. m. unicolor* mitotypes gave a value of −0.85, only slightly suggestive of a recent selective sweep or a population expansion after a bottleneck.

Global Ancestry Inference along Nuclear Genome

An initial assessment of population structure was performed by principal components analysis (PCA) on a distance matrix of the 19,888 unlinked SNPs (fig. 5*A*). Samples from Reunion form an intermediate group half way between the *A. m. unicolor* group of Madagascar and Mauritius, and the C lineage group of *A. m. carnica* and *A. m. ligustica*, along each of the four principal components plotted. An unrooted phylogeny based on genetic distance illustrates the branching of samples by subspecies, sampling location, and finally by ploidy (fig. 5*B*). The two honeybees from Rodrigues having C mitotypes group close to the *A. m. ligustica* and *A. m. carnica* honeybees, whereas REU28 having the C1 mitotype cannot be distinguished from the other Reunion samples. Concerns about potential issues arising when analyzing together honeybees having different ploidy were raised by Wragg et al. (2016) and are further demonstrated here in the phylogeny, where the overall picture shows a correct grouping of geographical origins but with a separation by ploidy within groups. Cross-validation (CV) error rate from ADXMITURE analyses indicates the most likely number of genetic backgrounds within the data to be $K = 6$ (CV = 0.48; supplementary table S5, Supplementary Material online), clearly differentiating *A. m. scutellata*, *A. m. jemenitica*, *A. m. mellifera*, *A. m. caucasia*, and *A. m. unicolor*, with a further background represented by the combined *A. m. carnica* and *A. m. ligustica* samples (fig. 3). At $K = 7$ (CV = 0.49), the next most likely number of backgrounds, *A. m. carnica* and *A. m. ligustica* separate, whereas at $K = 8$ (CV = 0.52), Reunion emerges as an independent genetic background. Where $K = 6$ or 7, Reunion samples exhibit an admixed

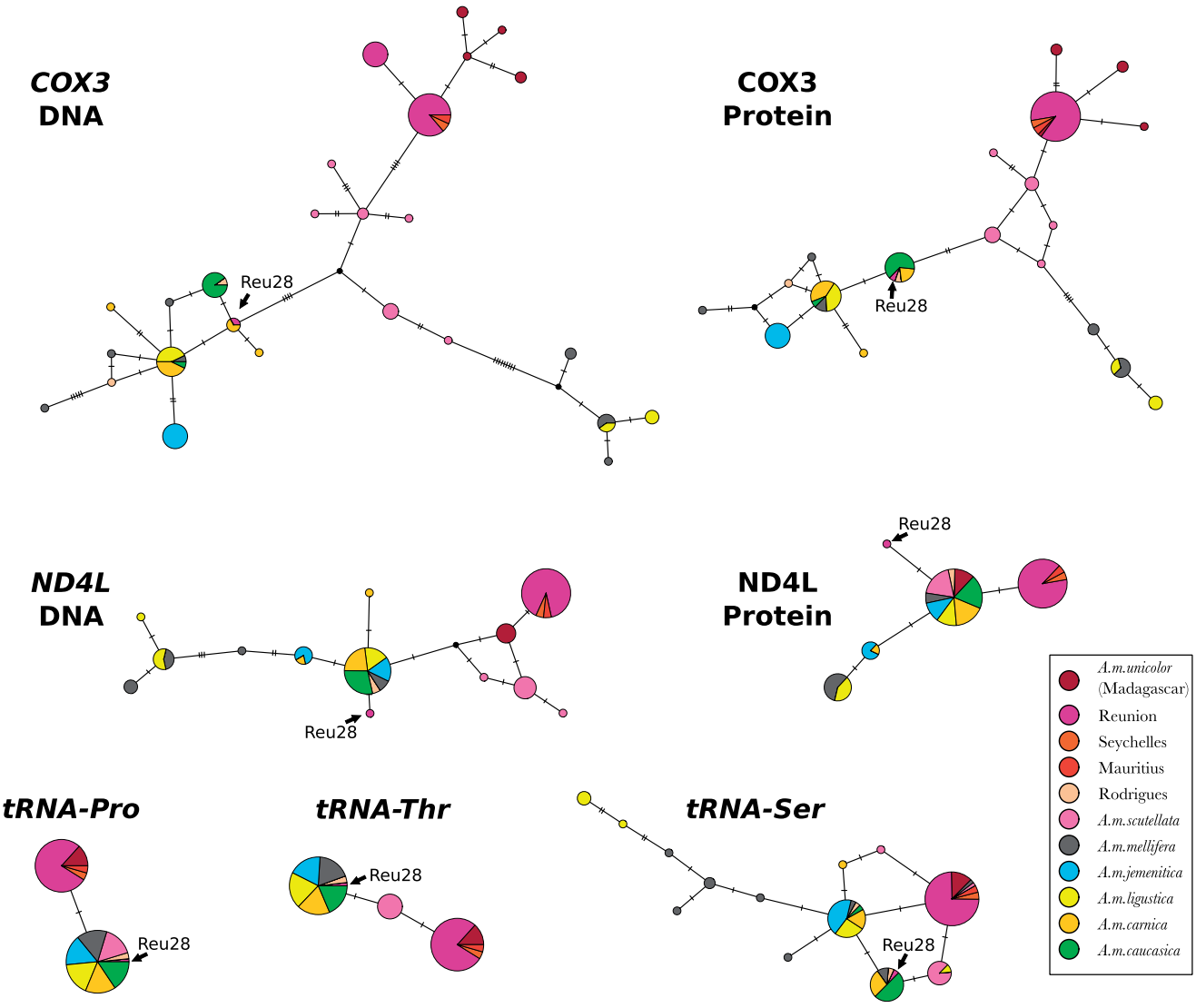

Fig. 4.—Haplotype networks for mitochondrial genes differentiating the SWIO samples. The *cox3* and *nd4l* coding genes and the *tRNA-pro*, *tRNA-thr*, and *tRNA-ser* genes each have haplotypes which are specific to the samples from the SWIO islands. In the case of the the *cox3* and *nd4l* genes, Madagascar samples have different haplotypes than the other SWIO samples. Reference samples are indicated by their subspecies names. For the SWIO samples, the name of the island is indicated. The REU28 sample having a European C lineage haplotype distinct from all other Reunion samples is indicated.

background of *A. m. carnica* and *A. m. ligustica* C-mitotype bees (43.2 ± 3.5%) and of *A. m. unicolor* (53.2 ± 5.9%), which is in accordance with the PCA results placing them half way between these two groups. Interestingly, the only Reunion sample (REU28) having the C mitotype does not stand out at all, having proportions of A-type and C-type backgrounds of 42.8% and 52.8%, respectively. Admixture is also evident in the samples from Rodrigues, and to a lesser extent Seychelles, whereas those from Madagascar and Mauritius possess only the *A. m. unicolor* background. More specifically, considering $K = 7$ where all reference populations separate, admixed honeybees from Seychelles exhibit background averages of 77.2% *A. m. unicolor* and 22.7% *A. m. ligustica*, whereas those from Rodrigues are a complex mix of 57.7% *A. m. ligustica*, 16.2% *A. m. mellifera*, 13.3% *A. m. unicolor*, and 11.9% *A. m. caucasia*. As only two samples are available for each of these populations additional sampling would be required to better interpret these results. However, the absence of the A mitotype has previously been reported on Rodrigues (Techer et al. 2015) and supports the low levels of *A. m. unicolor* background detected in our samples. At $K = 7$ Reunion honeybees comprise background averages of 53.2% *A. m. unicolor*, 35.6% *A. m. carnica*, 7.7% *A. m. ligustica*, 2.5% *A. m. caucasica*, and <1% *A. m. scutellata*, *A. m. jemenitica*, and *A. m. mellifera*.

A *k*-nearest neighbor (*k*NN) network, with pie-charts representing the Q values from the ADMIXTURE analysis at

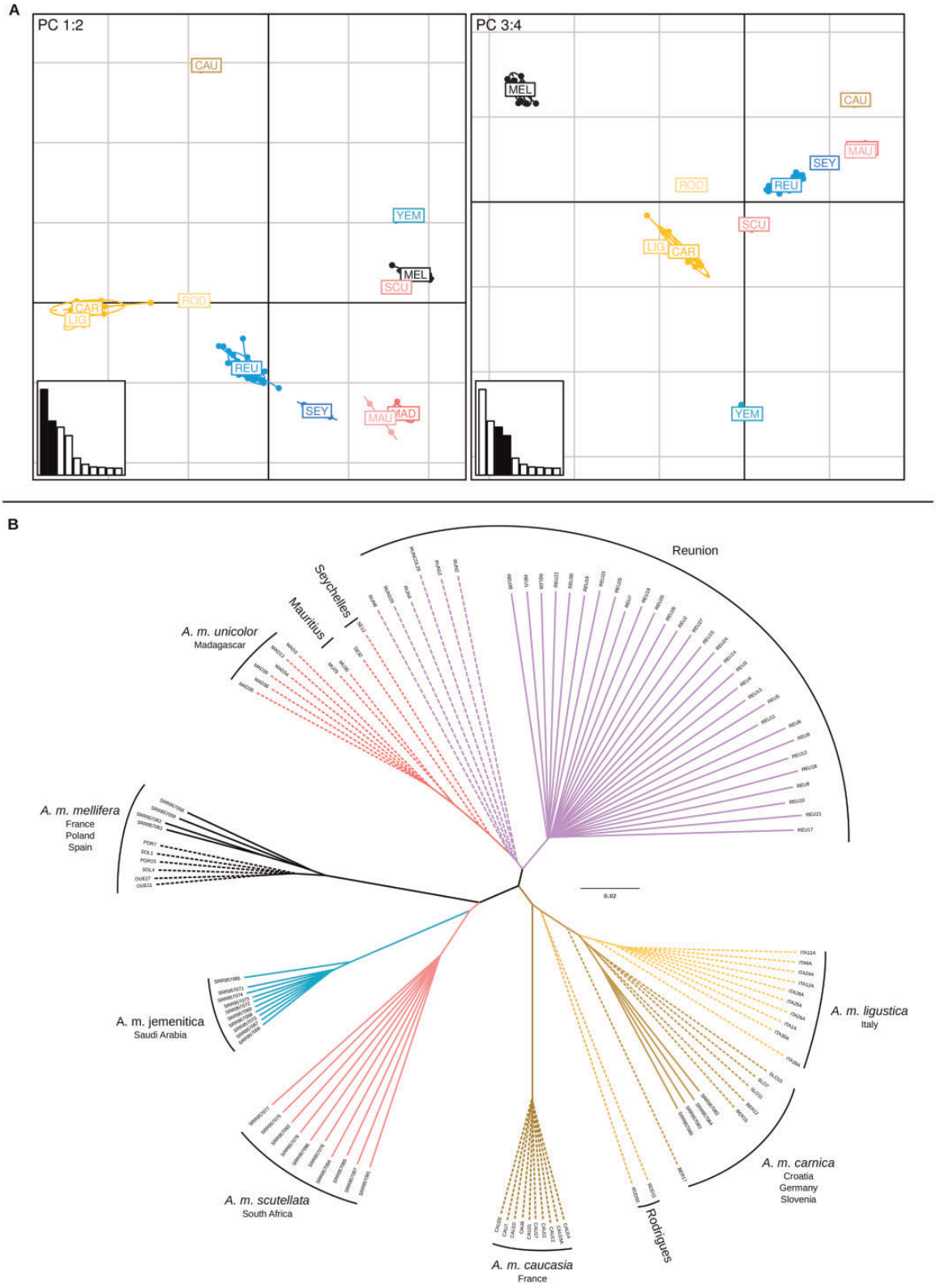

Fig. 5.—Autosomal analyses of genetic distance, based on 19,888 unlinked SNPs. (*A*) Principal components analysis. CAU, *Apis mellifera caucasia*; CAR, *A. m. carnica*; LIG, *A. m. ligustica*; ROD, Rodrigues; REU, Reunion; SEY, Seychelles; YEM, *A. m. jemenitica*; MEL, *A. m. mellifera*; SCU, *A. m. scutellata*; MAU, Mauritius; MAD, Madagascar (*A. m. unicolor*). (*B*) Unrooted neighbor-joining tree on which haploid individuals are indicated with dashed branches while diploid individuals are indicated with solid branches.

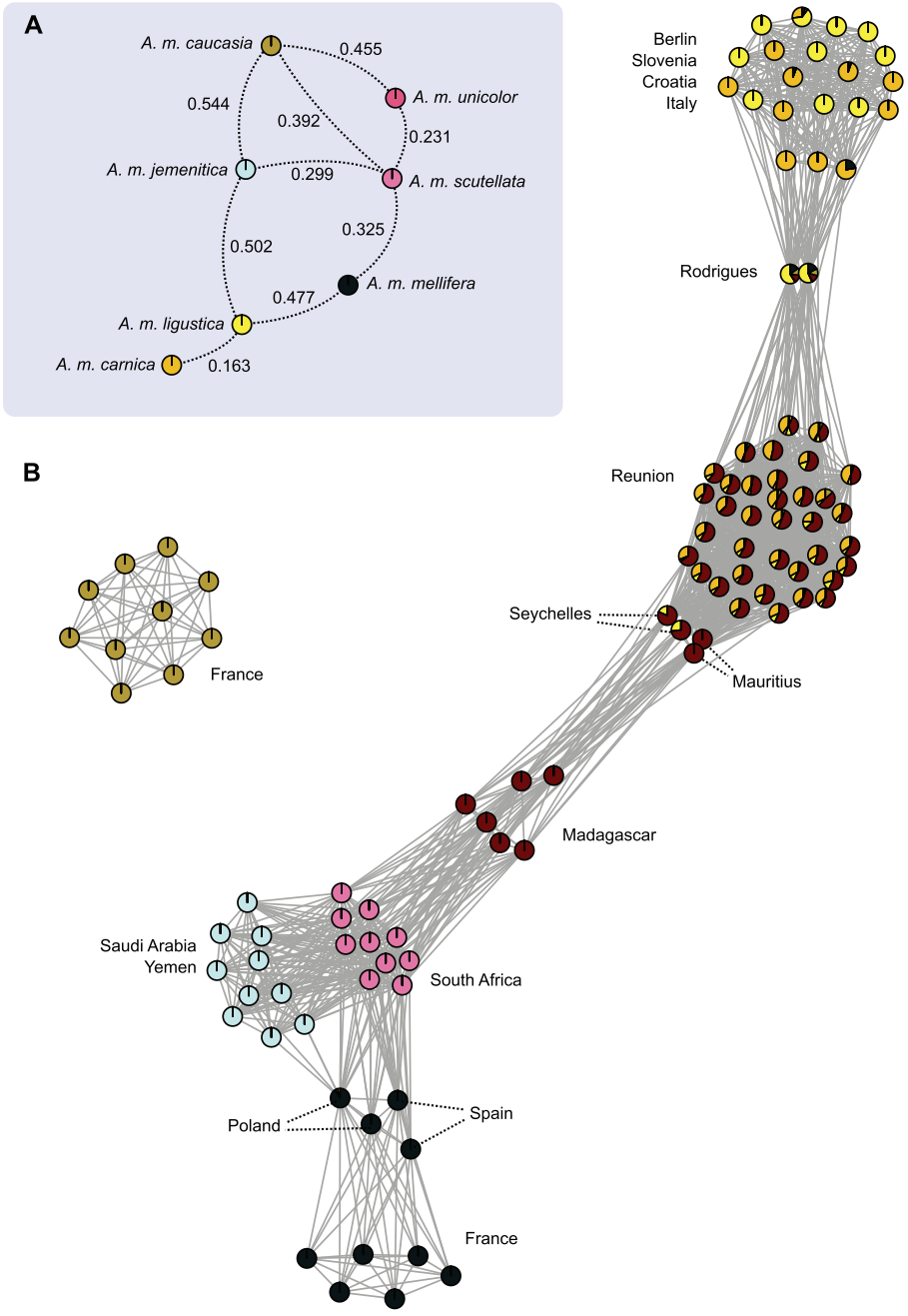

Fig. 6.—*k*-Nearest neighbor (*k*NN) network. (*A*) The lowest and highest F_{ST} values connecting each population from ADMIXTURE at K=7. (*B*) *k*NN network for 41 nearest neighbors with nodes indicating the results from ADMIXTURE at K=7.

$K = 7$ (supplementary table S6, Supplementary Material online), indicate the Reunion samples to be almost equidistant between *A. m. unicolor* from Madagascar and the *A. m. ligustica* and *A. m. carnica* samples from the C lineage (fig. 6). The most distant population, also evident by PCA, is *A. m. caucasia* which had no connectivity to the other populations. *A. m. caucasica* also returned the highest F_{ST} values in the ADMIXTURE analysis (fig. 6 and supplementary table S7, Supplementary Material online).

To further investigate the admixture observed in Reunion samples, the data were analyzed using TreeMix, where the most frequent migration observed was from *A. m. carnica* → Reunion (supplementary fig. S3, Supplementary Material online), followed by [*A. m. carnica*, *A. m. ligustica*] → Reunion (supplementary figs. S3 and S4, Supplementary Material online). TreeMix also allows a three-population test to be conducted. The test are in the form *f*3(A; B, C), where a significantly negative value of the *f*3 statistics ($Z < -4$) implies

that population *A* is admixed. Where *A* = Reunion, the most significant result strongly indicates the population to be an admixture of *A. m. unicolor* and *A. m. carnica* ($Z = -60.9$, supplementary table S8, Supplementary Material online).

Local Ancestry Inference along Nuclear Genome

To investigate whether or not the distribution of *A. m. unicolor* and C lineage genetic backgrounds in Reunion individuals was heterogeneous, the data were analyzed using PCAdmix. This analysis was performed using the previously constructed genetic map, and ancestry determined in 1 cM blocks. As such the physical block sizes were variable along chromosomes, returning a mean size of 19.8 ± 18.7 kb (supplementary fig. S5, Supplementary Material online) and a mean of 34 ± 25 SNPs per block. A contiguous run of blocks of the same ancestry is herein referred to as a haplotype block, the size distributions of which for each ancestral background are presented in supplementary figure S6, Supplementary Material online, accompanied by a summary of haplotype blocks per sample (supplementary table S9, Supplementary Material online) and the individual distribution of haplotype blocks along the chromosomes (supplementary fig. S7, Supplementary Material online). There was little difference in median haplotype block size between the A (4.5 cM) and C (4.4 cM) lineages, or in the frequency of blocks irrespective of contiguity assigned to either the A (0.51) or C (0.49) lineage. Pearson's *r* as a measure of the linear correlation between the PCAdmix and ADMIXTURE ancestry inferences indicates strong correlations for both the A ($r = 0.961$) and C ($r = 0.9$) lineage backgrounds. The heterogeneity of local ancestry inference was tested using the CAnD method, which aims to test for systematic differences in ancestry of admixed populations which could for instance arise due to preferential selection of one or the other background. *A. m. carnica* and *A. m. ligustica* are the least divergent pair of subspecies in our data set ($F_{ST} = 0.163$) and moreover, the contribution of *A. m. ligustica* to Reunion bees is low (7.7%). We therefore treated these two subspecies as a single C lineage reference background, contrasted against the A lineage *A. m. unicolor* background represented by bees from Madagascar and Mauritius. A loss of heterogeneity is observed at a number of genomic regions, indicating possible selection for either the A or C lineage background (fig. 7). Of these, 15 regions were significantly ($P < 0.001$) associated with the A lineage while 9 were significantly associated with the C lineage (supplementary table S10, Supplementary Material online). However, none of the regions identified had a mean posterior probability across all samples >0.91, indicating none of them to have reached fixation. Annotation of the SNPs present in the selected regions with Variant Effect Predictor (VEP) (McLaren et al. 2016) showed that 103 have potential direct effects on protein structure, including stop gain or loss, missense, and splice region variants (supplementary table S11, Supplementary Material online). The genotypes of SNPs having alleles alternatively fixed in the *A. m. unicolor* honeybees from Madagascar and Mauritius as opposed to the European *A. m. ligustica* and *A. m. carnica* were plotted in supplementary fig. S8, Supplementary Material online, for all regions detected, to visually confirm the prevalence of haplotypes from one or the other background. An example of two regions very close in chromosome 13 and showing preferential selection for the A background is shown on figure 8. Genome-wide annotation of 146,292 SNPs alternatively fixed in *A. m. unicolor* and European samples performed using VEP revealed 1,795 of them to have potential direct effects on protein structure, including stop gain or loss, missense, and splice region variants (supplementary table S12, Supplementary Material online).

The most significant region identified to be associated with A lineage ancestry (chromosome 13: 4.69–4.73 Mb) spans the four exons of *GB40069* and the last two exons of *GB40148* (fig. 8). The first is orthologous to *dpr13*, member of the defective proboscis extension response gene family in *Drosophila*, whose protein has a number of immunoglobulin features and is involved in the sensory perception of chemical stimuli. The latter has three orthologs in *Drosophila*, all of which encode proteins with Cytochrome b561 features and are thus likely to be involved in ascorbate regeneration. A region in very close proximity (chromosome 13: 4.78–4.83 Mb) (fig. 8) contains the forkhead protein *FoxP* (*GB40150*), which is of interest as another forkhead box protein (*foxo*) has previously been linked to differentiation between African and European lineages (Wallberg et al. 2014), and also between Savannah and Desert populations of bees from Kenya (Fuller et al. 2015). The same region also contains *GB40066*, an uncharacterized gene having two missense mutations in its second exon (SNPs at positions 4,790,007 and 4,790,053) whose allele frequencies in the Reunion population are much closer to *A. m. unicolor* than to C-type bees (supplementary table S11, Supplementary Material online).

In addition, identified in region (chromosome 5: 6.88–6.90) were an ortholog (*GB42054*) to the *Drosophila Na pump α subunit* (*ATPα*) gene, which has been inferred from mutant phenotypes to be involved in response to mechanical, temperature, and auditory stimuli; *GB48639* (chromosome 5: 8.52–8.56 Mb) which is orthologous to *serrate* whose protein contains 14 epidermal growth factor (EGF)-like domains and plays a role in morphogenesis; and *GB43497* (chromosome 11: 6.62–6.77 Mb) which is orthologous to *neuromusculin* and whose protein has a number of immunoglobulin-like features. A further region identified (chromosome 2: 12.72–12.76 Mb) hosts a cluster of six genes, and includes *GB55370* which is orthologous to *wrapper* whose protein also includes a number of immunoglobulin-like features, and which plays a role in axon development; *GB55367* an ortholog to *rtf1* which plays a role in transcription-coupled histone modification,

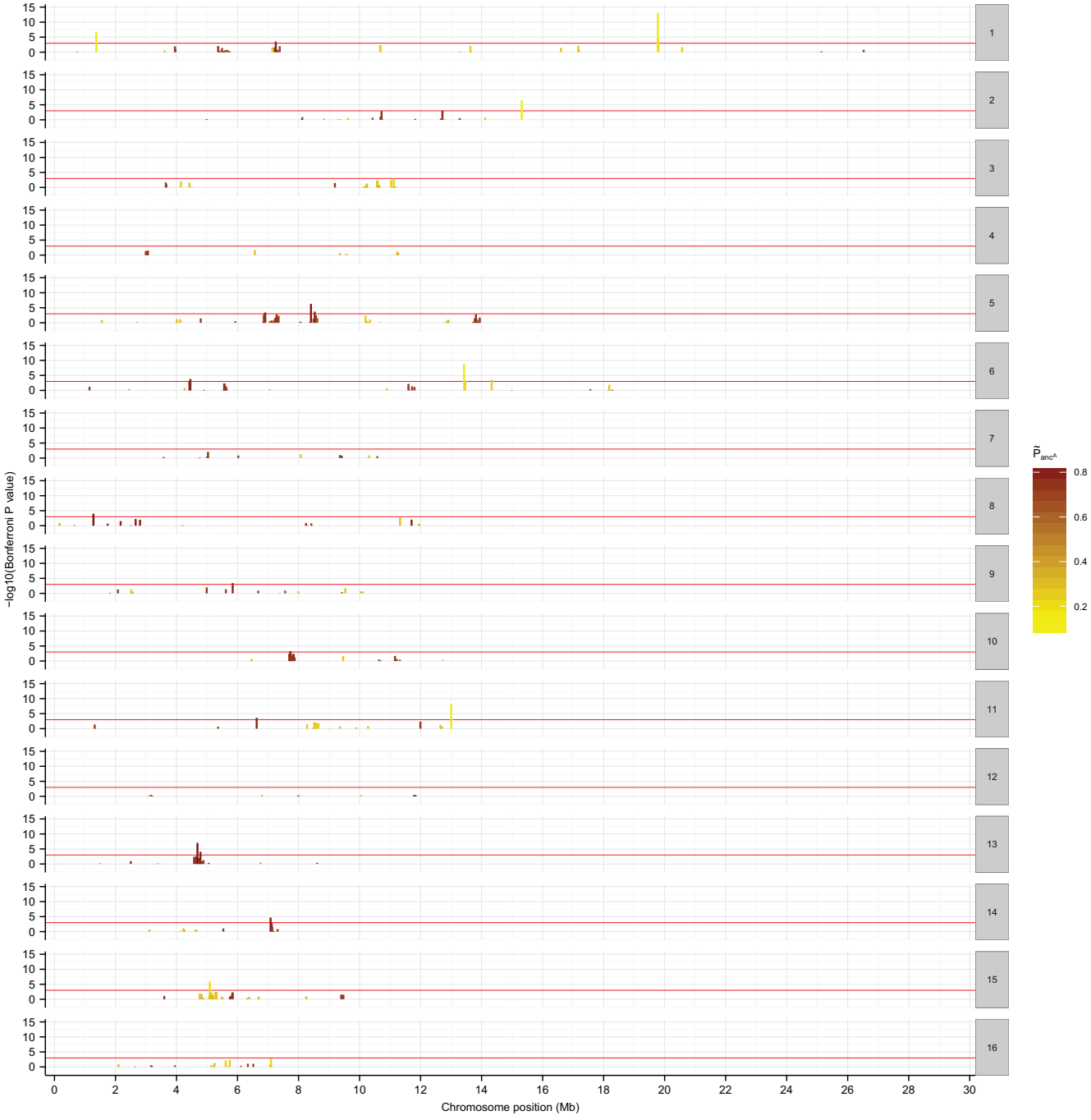

Fig. 7.—Heterogeneity (CAnD) test of local ancestry on Reunion. The color of the bar indicates the mean posterior probability of A-type ancestry in 1-cM windows across Reunion haplotypes. The horizontal lines indicates a significance threshold of $P<0.001$.

Notch signaling in the wing margins, and is required for maximal induction of heat-shock genes. Another region identified (chromosome 5: 8.40–8.44 Mb) spans the first exon and upstream region of an ortholog (*GB48653*) to the *homothorax* gene in *Drosophila*. This gene encodes a protein with a homeobox DNA-binding domain required for morphogenesis of organs within the peripheral nervous systems and antennal identity.

The most significant region identified to be associated with C lineage ancestry (chromosome 1: 19.78–19.79 Mb) hosts a gene (*GB55328*) orthologous to the *tropomyosin* gene, which plays a central role in the calcium dependent regulation of muscle contraction. In addition, within this region is a *tropomyosin-2-like* gene (*GB55329*). Tropomyosin has been found to be significantly overexpressed in insecticide-resistant *Aedes aegypti* moquitoes after exposure to deltamethrin

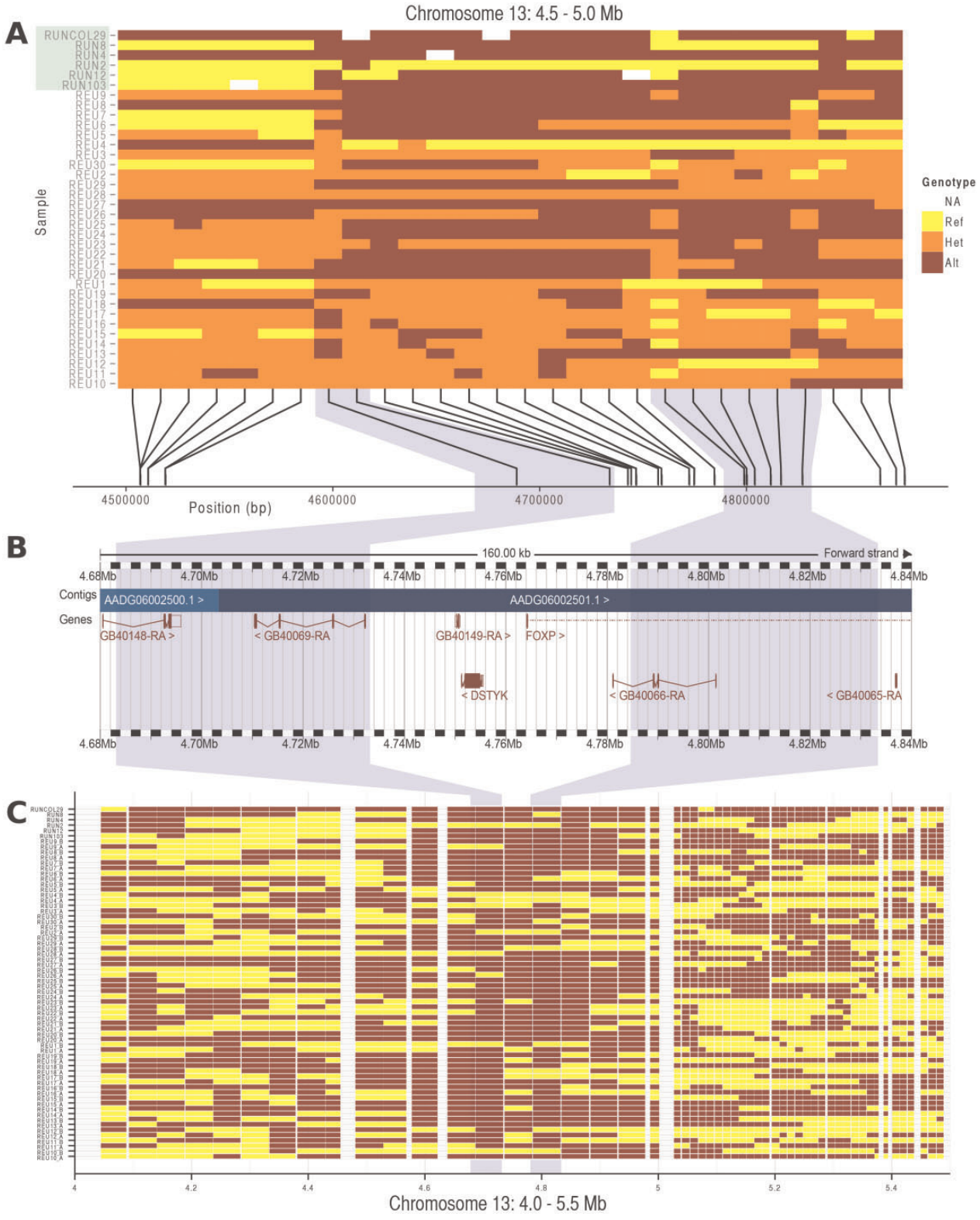

Fig. 8.—Region on chromosomes 13 with a preferential selection of *Apis mellifera unicolor* haplotypes. The two regions on chromosome 13 having a preferential selection of *A. m. unicolor* haplotypes are highlighted in blue, *unicolor* haplotype (−log10 *P*=7 and 4 for the left and right regions, respectively, see supplementary table S9, Supplementary Material online). (*A*) Genotypes in the Reunion samples (haploid drones highlighted in green), of SNPs having alleles alternatively fixed in the *A. m. unicolor* samples from Madagascar and Mauritius and in the *A. m. caucasica* and *A. m. ligustica* samples from Europe.

(Lertkiatmongkol et al. 2010). A further gene identified (*GB55492*) on chromosome 2 (15.31–15.33 Mb) is orthologous to *importin 9* whose predicted function is Ran GTPase binding. Ran GTPase has been demonstrated to regulate hemocytic phagocytosis against virus infection in shrimp (Liu et al. 2009). The Ran protein plays an important role in the innate immune system of invertebrates via phagocytosis (Ye and Zhang 2013), and is expressed at elevated levels in *Drosophila* embryonic Kc cells following exposure to deltamethrin (Liu et al. 2016). A further link to immune pathways defence mechanisms is the ubiquitin-related modifier 1 ortholog (*GB42195*) on chromosome 1 (1.37–1.37 Mb), demonstrated to be involved in the regulation of JNK signaling and response against oxidative stress in *Drosophila* (Khoshnood et al. 2016). The JNK pathway has been shown to be involved in the humoral immune response of greater wax moth larvae (Wojda et al. 2004). In addition, of note, is an ortholog to the *fringe* gene (*GB44913*) on chromosome 11 (13.00–13.02 Mb) which, like the *serrate* gene, is involved in morphogenesis, and which accommodates a long noncoding RNA that is overexpressed in honeybee queen ovaries (Humann et al. 2013). Interestingly, this significant region on chromosome 11 is situated within a larger region identified recently in a study on Africanized honeybees sampled in Brazil (Nelson et al. 2017).

Given the high levels of mitochondrial DNA of *A. m. unicolor* origin in the Reunion population, we further looked at nuclear genes having annotations indicating that they interact with the mitochondria, none of which emerged as significant in the selection analysis at $P < 0.001$. We searched for genes containing the Gene Ontology (GO) annotation keywords *mitoch* and *respir* (see complete list of keywords in supplementary table S13, Supplementary Material online) in the *A. mellifera* Amel4_5 Ensembl database. Of the 8,835 genes having GO annotations in the database, 165 (1.86%) could be linked to the mitochondria or to the respiratory chains. Two of these, *NDUFA5* and *NDUFA10*, are located in regions exhibiting a weak signal of selection toward the *A. m. unicolor* background, with *P* values of 0.073 and 0.006, respectively. These two genes are essential components of the mitochondrial respiratory chain complex (Garcia et al. 2017).

Discussion

The mitochondrial and nuclear genomes of the honeybee *A. mellifera* on Reunion were studied by whole genome sequencing, giving the highest genome-wise resolution achieved to date for this population. The results obtained can be interpreted both in terms of human management and of adaptation to the environment. Imports of European honeybees to the island for a duration of at least a century followed by a 30-year import ban, together with transhumance of colonies, has led to the emergence of an admixed population with balanced proportions of *A. m. unicolor* and C-type nuclear ancestry. Although the proportions of ancestries are homogenous between colonies, their genome-wide distribution is variable.

Influence of Ploidy on Sequencing Results

Having both haploid and diploid samples in our data set, we first checked the influence of ploidy using standard population genetics approaches. Substructure arising from differences in ploidy could be observed in a neighbor-joining phylogeny based on genetic distance, although the effect is not so pronounced as to disturb the phylogeny where proportional numbers of samples with different ploidy have been included for each population. However, there is a clear effect in the Reunion data where the number of diploids outnumbers the haploids by 5 to 1. Future studies seeking to integrate samples with differing ploidy should be mindful of this when processing their data. Further substructure is evident by sampling location within lineage, for instance within *A. m. mellifera* the diploid samples originate from Poland (SRR957058, SRR957059) and Spain (SRR957061, SRR957062), and branch accordingly, as do the haploid *A. m. carnica* samples from Slovenia (SLO7, SLO11, SLO15) and Berlin (BER12, BER15). It is worth recalling that these analyses were based on a subset of SNPs that had been filtered to remove those in strong LD ($r^2 > 0.1$) and spaced <1,000 bp apart. More stringent filtering criteria might remove the ploidy effect, but may also impact on the ability to detect population substructure arising from sampling location, which is clearly of interest to the population geneticist. Additional work in this area would be of interest to maximize the benefits that can be derived from integrating haploid and diploid sequence data. Nonetheless, the impact of ploidy in this regard does not confound this study as it is explicitly focused on the analyses of haplotypes, wherein haploid data make a significant contribution by acting as reference haplotypes for phasing diploid data.

Fig. 8.—Continued

White: missing, yellow: homozygous European allele, orange: heterozygous, brown: homozygous *A. m. unicolor* allele. Distances between SNPs are not to scale and their positions along the chromosome are indicated by the black lines. (*B*) Gene models from Ensembl (http://metazoa.ensembl.org/; last accessed November 2017) in the selected region. (*C*) Ancestral backgrounds of 1 cM haplotype blocks in the region as determined by PCAdmix analysis: yellow: European haplotype, brown: *A. m. unicolor* haplotype. Physical block sizes vary according to the local recombination rate along the chromosome. For each diploid sample, both chromosomes are represented and labeled with the sample name followed by _A and _B.

Autosomal and Mitochondrial Accounts on the History of Honeybee Subspecies Introduction in Reunion

Statistically, the most likely result ($K = 6$) from the ADMIXTURE analyses suggests the Reunion population to be a mixture of *A. m. unicolor* originating from Madagascar and European honeybees from the C lineage. The distribution of admixture proportions is homogenous, with each sample having on an average just over 50% of *A. m. unicolor* background (fig. 3). Increasing levels of K differentiates *A. m. carnica* and *A. m. ligustica* ($K = 7$), and Reunion ($K = 8$) populations. Each of these results might be considered plausible, particularly given that *A. m. carnica* and *A. m. ligustica* are different subspecies, and with Reunion being an island population there is strong potential for genetic drift due to reduced gene flow. Considering $K = 8$, samples from the Seychelles and Mauritius are dominated by the Malagasy background, with some admixture from European honeybees in the former, which suggests two possible scenarios. The first being that honeybees from Madagascar arrived on Reunion well before man settled in 1665, allowing sufficient time to differentiate by genetic drift, and subsequently did not reach Mauritius which is again further east, until it was introduced artificially—probably from Madagascar based on our results. The second scenario would be that honeybees from Madagascar arrived on all islands at around the same time period, and that the import of European lineages to Reunion occurred sufficiently long ago to facilitate the emergence of a homogenous island population through panmixia.

Analysis of the mitochondrial DNA *tRNAleu-cox2* hypervariable region illustrates the dominance of the *A. m. unicolor* maternal heritage, with only 1 of the 36 Reunion honeybees possessing a C mitotype. This suggests either that imports from Europe were limited or that European mitochondrial DNA was selected against, and favors the emergence of a specific Reunion background due to drift. However, whole mitochondrial genome analysis suggests a closer proximity between samples from Reunion, Mauritius, and Seychelles on one side and those from Madagascar on the other (fig. 2). Moreover, Mauritius, Seychelles, and Reunion consistently possess the same major haplotypes for the coding genes, whereas the samples from Madagascar differ from this group both at the DNA and protein levels for the *cox3* and *nd4l* genes (fig. 4 and supplementary fig. S2, Supplementary Material online). For *cox3*, four distinct DNA haplotypes are found for the Madagascar samples, all of them different to the two haplotypes found in the A-type Reunion samples, whereas the samples from Seychelles and Mauritius share one of the two Reunion haplotypes. This remains true at the amino acid level, where all Reunion A-type, Mauritius, and Seychelles samples have the same sequence together with one of the Madagascar samples, whereas the remaining samples from Madagascar return three distinct protein sequences. For *nd4l*, all Reunion A-type, Mauritius, and Seychelles samples share the same DNA haplotypes and protein sequences, distinct from those of the Madagascar samples. All this suggests a common maternal inheritance for Mauritius, Seychelles, and Reunion populations, which is in disagreement with the hypothesis of the genetic background of Reunion samples being different from the other SWIO samples by genetic drift alone, as suggested by the nuclear ADMIXTURE analysis at $K = 8$.

One might expect given the dominance of the mitochondrial A lineage background on Reunion, that PCA of nuclear genetic distance would result in its clustering with samples from Madagascar, as in the case of Mauritius (fig. 5*A*). However, their intermediate placement suggests either an independent genetic background, or an intermediate state between the C and A lineages, as might arise through admixture. This intermediate placing of Reunion samples midway between *A. m. ligustica* and *A. m. carnica* from the C lineage and *A. m. unicolor* from the A lineage corresponds to the proportions found in the ADMIXTURE analyses for $K = 6$ or 7, and is further exemplified in the *k*NN network analysis in which Reunion is clearly a link between the two lineages (fig. 6*B*). Furthermore, the population splitting and mixing inferred by Treemix with seven populations, allowing for one migration event, indicates the most significant migrations to be from *A. m. carnica* and *A. m. ligustica* into the Reunion population (supplementary figs. S3 and S4, Supplementary Material online). Considering the A and C lineages as reference genetic backgrounds, a detailed examination of admixture through local ancestry inference of haplotype blocks revealed the average fraction of ancestry for A and C lineage backgrounds to be strongly correlated with the initial genome-wide estimates by ADMIXTURE. The results also demonstrated the two genetic backgrounds to be present in a large number of small haplotypes, consistent with extensive recombination over time coupled with the high recombination rate observed in honeybees (Liu et al. 2015).

To summarize the genetic accounts on the history of the Reunion population, our results support a common maternal origin for honeybees in Reunion, Mauritius, and the Seychelles, most likely originating from Madagascar. This is in agreement with the account of Du Bois (1674), that honeybees imported from Fort Dauphin had acclimatized well on Reunion island, rather than the alternative hypothesis of their presence before the arrival of man (Hermann 1920). However, more extensive sampling is required to establish their precise geographical origin on Madagascar, and the exact timing of the dissemination of honeybees in the SWIO. A more affordable option in this regard than whole genome sequencing or even *tRNA-cox2* sequencing, would be to genotype a few diagnostic SNPs, such as those found in the *cox3* gene and others segregating in the SWIO samples.

Selection on the Mitochondrial Genome

Most intriguing are the completely different narratives given by the mitochondrial and nuclear genomes. Only 1 out of 36 Reunion samples exhibited a European mitotype, consistent with the results of another study (Techer et al. 2017), which is in stark contrast with the 43% European autosomal DNA found on an average in the samples. Similar observations of differences between autosomal and mitochondrial proportions in an admixed population have been documented on Puerto Rico (Rivera-Marchand et al. 2012; Galindo-Cardona et al. 2013), where Africanized honeybees and European honeybees have been hybridizing since the 1990s. Puerto Rico has a uniform geographic distribution of Africanized honeybee morphology and mitochondria, whereas analysis of microsatellites indicates a heterogeneous population comprising two genetic backgrounds of near equal proportions.

Clearly, although large numbers of hives of European origin were imported between the late 19th century and 1982, they must have at first been outnumbered by the local *A. m. unicolor* colonies, but one would expect beekeepers to propagate imported colonies thus leading to a spread of the mitochondrial DNA through the queens. Until the 1960s, when hives with mobile frames were introduced in Reunion, honeybee production was extensive and today, a non-negligible proportion of beekeepers still gather natural swarms, so it is expected that both colonies with *A. m. unicolor* and European maternal lineages would be found. One possibility is that on a whole, the genome of the European honeybees was favored on the island by man, either through repeated imports or by multiplication of European colonies with higher productivity, but that in the wild, natural colonies having *A. m. unicolor* mitochondrial DNA might have been favored by nature due to adaptation to the tropical climate. However, discrepancies between nuclear and mitochondrial DNA have been documented in several cases (Toews and Brelsford 2012). Simulation studies of spatial expansion with interbreeding have shown that introgression goes preferentially from local species toward invading ones, suggesting massive introgression in an invading species as the null expectation for neutral genes (Currat et al. 2008). Moreover, this study also shows that the rate of introgression is often negatively correlated with the rate of intraspecific gene flow and in cases where dispersal is male-biased, as in honeybees (Johnstone et al. 2012), the lower gene flow associated with the maternally inherited mitochrondrial DNA could account for a higher rate of introgression (Petit and Excoffier 2009).

Our data on mitochondrial DNA do not allow to conclude between adaptation and the neutral model, and although a few nonsynonymous changes were found in the coding genes, the McDonald–Kreitman tests performed between the two subspecies involved here were inconclusive on the possible action of positive selection. Other studies have shown that positive and negative selection can act on the mitochondrial genome, for instance on the *cox* and *nd* genes, suggesting a physiological connection between mitochondrial DNA sequence and fitness (Meiklejohn et al. 2007). We also found SNPs and indels in tRNA genes differentiating subspecies, and some of these may have an impact on general metabolism, as has been demonstrated for mutations in human tRNA genes that have been associated with diseases (Tuppen et al. 2010; Suzuki et al. 2011).

A temporal study of honeybee populations of the Macaronesian achipelago demonstrated clear impacts of European imports on the mitochondrial genetic diversity of local populations (Muñoz et al. 2013). In that study Muñoz et al. analyzed mitochondrial data from a 10-year period revealing different patterns of change in haplotype diversity. High frequencies of C lineage haplotypes were detected, consistent with earlier studies on the Macaronesian islands (De la Rúa et al. 2001, 2006), however contrary to the earlier studies there was a notable absence of C lineage haplotypes on some of the Canary islands. This lead the authors to speculate a change in importation practices on these islands—suggesting that the C lineage imports did not succeed in establishing their presence on these islands, at least at the mitochondrial level, possibly reflecting adaptation problems. Conservation policies implemented across the Canaries in 2001 in an effort to preserve the black Canarian honeybee may account for possible changes in importation practices. It would be interesting to investigate the nuclear genomes of these populations at high resolution, as in our study, to establish the extent of hybridization that has occurred to date, and to revisit the populations over time to observe how admixture levels change. Such a study might present an early window into the processes that appear to have occurred on Reunion.

Selection on the Nuclear Genome

The dominance of the *A. m. unicolor* lineage mitotype on Reunion despite European lineage imports is suggestive of a selective advantage of the *A. m. unicolor* type. If this is considered to be the case then one would expect to observe similar signals on the nuclear genome, as appears to be the case in our study. The climate and endemic flora of Reunion are likely to pose challenges to exotic subspecies adapted to different ecoclimates, and although the introduction and success of exotic floral species on the island may alleviate that particular challenge, most originate from Asia or South America whereas the introduced honeybees are from Europe. It could be speculated that several of the genes identified in regions associated with selection for the A lineage background might be linked to the ecoclimate. These include genes with orthologs in *Drosophila* annotated to be involved in sensory perception of chemical, mechanical, thermal, and auditory stimuli, morphogenesis, and the nervous system. One ortholog in particular, *rtf1*, is reported to be required

for maximal induction of heat-shock genes, and could therefore be implicated in response to stressful conditions. Another gene of potential interest, *foxp*, is a member of the FOX proteins family to which another gene, *foxo*, has been identified in previous studies comparing African and European lineages (Wallberg et al. 2014) and populations inhabiting different ecoclimates in Kenya (Fuller et al. 2015).

Similar speculations concerning regions associated with selection for the C lineage might concern genes linked to xenobiotic responses or having an immune function. Several such genes were identified, which include several orthologs to genes that have been shown to be overexpressed following exposure to the pyrethroid pesticide deltamethrin. In addition, of note, is an ortholog to *fringe*, in which a long noncoding RNA has previously been shown to be overexpressed in ovaries of honeybee queens (Humann et al. 2013). This gene is located within a quantitative trait locus underlying the transgressive ovary phenotype identified in workers from crosses between Africanized and European honeybees (Linksvayer et al. 2009). A significant but negative correlation between ovary size and the concentration of nectar collected by foragers was reported in these bees. It has also been suggested that hybridization may disrupt coevolved components of honeybee development, resulting in females with less strictly canalized queen and worker phenotypes (Linksvayer et al. 2009). It is noteworthy that the region containing *fringe* at 13.00–13.02 Mb on chromosome 11 is in the center of a larger block recently described by Nelson et al. (2017), whose study follows the introduction of African *A. m. scutellata* honeybees to South America where they spread rapidly and replaced the honeybees of European descent to the point at which the average proportion of African ancestry is now 0.84 (Nelson et al. 2017). By contrast, in our study, it is the European honeybees that have been introduced to an African landscape, and the proportion of African (*A. m. unicolor*) ancestry observed is lower (0.53), although *A. m. scutellata* is known to be more aggressive than *A. m. unicolor* which may be a contributory factor to its success. What both studies have in common, is that the populations are a mix of honeybees from tropical and temperate climates having different behavioral responses, for instance in terms of foraging and brood development.

One current limitation of our study is that only 8,835 (57.7%) out of the 15,314 *A. mellifera* genes reported in the Ensembl database have GO annotations. As a result, only 25 (58%) of the 43 genes identified in significant regions following our selection analyses are annotated. This greatly limits the extent of our analysis. However, as the reference genome assembly and annotation improves, it may be possible in the future to put into context some of the genes identified that currently lack annotation.

Given the topography of the island, its diversity of microhabitats, and variations in rainfall and wind speed between the eastern and western halves, it might be interesting to consider that local adaptation can occur within its mosaic of different landscape patches. Such adaptation might be observed following the analyses presented here if clusters of samples shared haplotypes and their sampling habitats shared some if not all environmental variables. However, in the case of the Reunion honeybee, management practices involve transhumance several times a year and so there is unlikely to be constant selective pressure reinforcing adaptation to any singular microhabitat on the island.

Conclusion

Our results show that the honeybees from Reunion have a mixed ancestry of *A. m. unicolor* from Madagascar and of C-type bees *A. m. carnica* and *A. m. ligustica* from Europe. A test for preferential selection of one or the other background along the chromosomes allowed the detection of genomic regions of local Malagasy or introduced European background, suggesting adaptation to the local tropical climate or a response to management practices by the beekeepers. Interestingly, one of the regions in which the European background was favored coincides with a region recently identified by Nelson et al. (2017). Reunion honeybees exhibit balanced proportions of European and Malagasy nuclear genetic backgrounds, and preferential selection for the tropical *A. m. unicolor* mitochondrial genome, supporting previous observations and suggesting that mitochondrial genes might play an important role in adaptation.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

This work was supported by the SeqApiPop programme, funded by the FranceAgriMer grant 14-21-AT and by a grant from the INRA Département de Génétique Animale (INRA Animal Genetics division). Sequencing was performed in collaboration with the GeT platform, Toulouse (France), a partner of the National Infrastructure France Génomique, thanks to support by the Commissariat aux Grands Invetissements (ANR-10-INBS-0009). Bioinformatics analyses were performed on the GenoToul Bioinfo computer cluster. Honeybee sampling in the Indian Ocean Islands was funded through the projects e-PRPV and EPIBIO-OI funded by the European Union (ERDF), the Région Reunion and CIRAD. Drone samples were kindly provided by Dr Cecilia Costa, from the Council for Agricultura Research and Economics Honey Bee and Silkworm Unit, Bologna, Italy; Dr Ales Gregorc from the Kmetijski Inštitut Slovenije (Agricultural Institute of Slovenia); Kaspar Bienefeld from the Institute for Bee Research Hohen Neuendorf;

l'Association Conservatoire de l'Abeille Noire Bretonne and the Conservatoire de l'Abeille Noire Provençale.

Author Contributions

Y.L.C., J.-P.B., B.B., and A.V. designed the experiment. B.B., Y.L.C,. H.D., M.T., and A.V. coordinated colony selection and sampling. K.T., E.L., and O.B. performed DNA extraction, library preparation and sequencing. D.W. performed the bioinformatic analyses and cowrote the manuscript with H.D., M.T., J.C., and A.V.

Literature Cited

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19(9):1655–1664.

Brisbin A, et al. 2012. PCAdmix: principal components-based assignment of ancestry along each chromosome in individuals with admixed ancestry from two or more populations. Hum Biol. 84(4):343–364.

Chang CC, et al. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4:7.

Chávez-Galarza J, et al. 2015. Revisiting the Iberian honey bee (*Apis mellifera iberiensis*) contact zone: maternal and genome-wide nuclear variations provide support for secondary contact from historical refugia. Mol Ecol. 24(12):2973–2992.

Currat M, Ruedi M, Petit RJ, Excoffier L. 2008. The hidden side of invasions: massive introgression by local genes. Evolution 62(8):1908–1920.

De la Rúa P, Galián J, Pedersen BV, Serrano J. 2006. Molecular characterization and population structure of *Apis melliferafrom* Madeira and the Azores. Apidologie 37(6):699–708.

De la Rúa P, Galián J, Serrano J, Moritz RF. 2001. Genetic structure and distinctness of *Apis mellifera* L. populations from the Canary Islands. Mol Ecol. 10(7):1733–1742.

Delaneau O, Marchini J, Zagury J-F. 2011. A linear complexity phasing method for thousands of genomes. Nat Methods 9(2):179–181.

Du Bois. 1674. Les voyages faits par le sieur D. B. [Du Bois] aux isles Dauphine ou Madagascar, et Bourbon ou Mascarenne, ès années 1669, 70, 71 & 72. 1–260. http://gallica.bnf.fr/ark:/12148/bpt6k1027018, last accessed november 2017.

Evans JD, et al. 2013. Standard methods for molecular research in *Apis mellifera*. J Apic Res. 52(4):1–54.

Franck P, et al. 2001. Genetic diversity of the honeybee in Africa: microsatellite and mitochondrial data. Heredity 86(Pt 4):420–430.

Franck P, Garnery L, Celebrano G, Solignac M, Cornuet J-M. 2000. Hybrid origins of honeybees from Italy (*Apis mellifera ligustica*) and Sicily (*A. m. sicula*). Mol Ecol. 9(7):907–921.

Fuller ZL, et al. 2015. Genome-wide analysis of signatures of selection in populations of African honey bees (*Apis mellifera*) using new web-based tools. BMC Genomics 16:518.

Galindo-Cardona A, Acevedo-Gonzalez JP, Rivera-Marchand B, Giray T. 2013. Genetic structure of the gentle Africanized honey bee population (gAHB) in Puerto Rico. BMC Genet. 14:65.

Garcia CJ, Khajeh J, Coulanges E, Chen EI, Owusu-Ansah E. 2017. Regulation of mitochondrial complex I biogenesis in Drosophila flight muscles. Cell Rep. 20(1):264–278.

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(11):1016–1021.

Hahn C, Bachmann L, Chevreux B. 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Res. 41(13):e129.

Harpur BA, et al. 2014. Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. Proc Natl Acad Sci U S A. 111(7):2614–2619.

Hermann P. 1920. Apiculture pratique aux colonies tropicales. 90p.

Humann FC, Tiberio GJ, Hartfelder K. 2013. Sequence and expression characteristics of long noncoding RNAs in honey bee caste development – potential novel regulators for transgressive ovary size. PLoS One 8(10):e78915.

Jacquard C, et al. 2013. Population structure of the melon fly, *Bactrocera cucurbitae*, in Reunion Island. Biol Invasions 15(4):759–773.

Johnstone RA, Cant MA, Field J. 2012. Sex-biased dispersal, haplodiploidy and the evolution of helping in social insects. Proc R Soc B Biol Sci. 279(1729):787–793.

Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics Oxf Engl. 24(11):1403–1405.

Jombart T, Ahmed I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27(21):3070–3071.

Khoshnood B, Dacklin I, Grabbe C. 2016. Urm1: an essential regulator of JNK signaling and oxidative stress in *Drosophila melanogaster*. Cell Mol Life Sci. 73(9):1939–1954.

Larkin MA, et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics Oxf Engl. 23(21):2947–2948.

Larsson A. 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics Oxf Engl. 30(22):3276–3278.

Latreille PA. 1804. Notice des espèces d'abeilles vivant en grande société, ou abeilles proprement dites, et description d'espèces nouvelles. Ann Mus Natl Hist Nat. 5:161–178.

Leigh JW, Bryant D, Nakagawa S. 2015. popart: full-feature software for haplotype network construction. Methods Ecol Evol. 6(9):1110–1116.

Lertkiatmongkol P, Pethuan S, Jirakanjanakit N, Rongnoparut P. 2010. Transcription analysis of differentially expressed genes in insecticide-resistant *Aedes aegypti* mosquitoes after deltamethrin exposure. J Vector Ecol. 35(1):197–203.

Linksvayer TA, et al. 2009. The genetic basis of transgressive ovary size in honeybee workers. Genetics 183(2):693–707.

Liu H, et al. 2015. Causes and consequences of crossing-over evidenced via a high-resolution recombinational landscape of the honey bee. Genome Biol. 16:15.

Liu W, et al. 2016. Up-regulated expression of Ran reveals its potential role to deltamethrin stress in Kc cells. Gene 583(1):1–7.

Liu W, Han F, Zhang X. 2009. Ran GTPase regulates hemocytic phagocytosis of shrimp by interaction with myosin. J Proteome Res. 8(3):1198–1206.

McHugh C, Brown L, Thornton TA. 2016. Detecting heterogeneity in population structure across the genome in admixed populations. Genetics 204(1):43–56.

McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297–1303.

McLaren W, et al. 2016. The Ensembl variant effect predictor. Genome Biol. 17:1–14.

Meiklejohn CD, Montooth KL, Rand DM. 2007. Positive and negative selection on the mitochondrial genome. Trends Genet. 23(6):259–263.

Muñoz I, Pinto MA, De la Rúa P. 2013. Temporal changes in mitochondrial diversity highlights contrasting population events in Macaronesian honey bees. Apidologie 44(3):295–305.

Nelson RM, Wallberg A, Simões ZLP, Lawson DJ, Webster MT. 2017. Genomewide analysis of admixture and adaptation in the Africanized honeybee. Mol Ecol. 26(14):3603–3617.

Neuditschko M, Khatkar MS, Raadsma HW. 2012. NetView: a high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. PLoS One 7(10):e48375.

Petit RJ, Excoffier L. 2009. Gene flow and species delimitation. Trends Ecol Evol. 24(7):386–393.
Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. 2014. PopGenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol. 31(7):1929–1936.
Pickrell JK, Pritchard JK. 2012. Inference of population splits and mixtures from genome-wide allele frequency data. PLoS Genet. 8(11):e1002967.
Price MN, Dehal PS, Arkin AP. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PLoS One 5(3):e9490.
Rivera-Marchand B, Oskay D, Giray T. 2012. Gentle Africanized bees on an oceanic island. Evol Appl. 5(7):746–756.
Ruttner F. 1988. Biogeography and taxonomy of honeybees. Berlin: Springer-Verlag.
Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. 4(4):406–425.
Schliep KP. 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27(4):592–593.
Strasberg D, et al. 2005. An assessment of habitat diversity and transformation on La Réunion Island (Mascarene Islands, Indian Ocean) as a basis for identifying broad-scale conservation priorities. Biodivers Conserv. 14(12):3015–3032.
Suzuki T, Nagao A, Suzuki T. 2011. Human mitochondrial tRNAs: biogenesis, function, structural aspects, and diseases. Annu Rev Genet. 45:299–329.
Techer MA, et al. 2015. Genetic characterization of the honeybee (*Apis mellifera*) population of Rodrigues Island, based on microsatellite and mitochondrial DNA. Apidologie 46(4):445–454.
Techer MA, et al. 2017. Large-scale mitochondrial DNA analysis of native honey bee *Apis mellifera* populations reveals a new African subgroup private to the South West Indian Ocean islands. BMC Genet. 18(1):53.
Techer MA, et al. 2017. Genetic diversity and differentiation among insular honey bee populations in the South West Indian Ocean likely reflects old geographical isolation and modern introductions. Plos One 2(12):e0189234.
Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol Ecol. 21(16):3907–3930.
Tuppen HAL, Blakely EL, Turnbull DM, Taylor RW. 2010. Mitochondrial DNA mutations and human disease. Biochim Biophys Acta 1797(2):113–128.
Wallberg A, et al. 2014. A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. Nat Genet. 46(10):1081–1088.
Wojda I, Kowalski P, Jakubowicz T. 2004. JNK MAP kinase is involved in the humoral immune response of the greater wax moth larvae *Galleria mellonella*. Arch Insect Biochem Physiol. 56(4):143–154.
Wragg D, et al. 2016. Whole-genome resequencing of honeybee drones to detect genomic selection in a population managed for royal jelly. Sci Rep. 6(1):27168.
Ye T, Zhang X. 2013. Involvement of Ran in the regulation of phagocytosis against virus infection in S2 cells. Dev Comp Immunol. 41(4):491–497.
Yu G, et al. 2017. ggtree: an rpackage for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 8(1):28–36.

Associate editor: Daniel Sloan