Supplementary Materials for

**Thrice out of Asia and the adaptive radiation of the western honey bee**

Kathleen A. Dogantzis, Tanushree Tiwari, Ida M. Conflitti, Alivia Dey, Harland M. Patch, Elliid M. Muli, Lionel Garnery, Charles W. Whitfield, Eckart Stolle, Abdulaziz S. Alqarni, Michael H Allsopp, Amro Zayed*

*Corresponding author. Email: zayed@yorku.ca

Published 3 December 2021, *Sci. Adv.* 7, eabj2151 (2021)
DOI: 10.1126/sciadv.abj2151

**The PDF file includes:**

- Supplementary Methods
- Supplementary Text
- Figs. S1 to S13
- Tables S1 to S10
- Legends for data S1 to S3
- References

**Other Supplementary Materials for this manuscript include the following:**

- Data S1 to S3
Supplementary Methods

Sample collection, DNA extraction, and sequencing

We composed a population genomic dataset of 251 individual Apis mellifera samples, of which 160 samples are newly sequenced, with the remaining samples downloaded from the Sequence Read Archive (SRA) (Data S1), including 15 Apis cerana genomes (54). New samples were collected across several years and wide geographic areas to ensure we were sampling one bee per colony. New samples were received as either whole bees, partial bees, or already extracted DNA. We extracted DNA from bee tissues using a Mag-Bind® Blood & Tissue DNA HDQ 96 Kit (Omega Bio-tek Inc., USA) optimised for the KingFisher™ Flex Purification System (Thermo Fisher Scientific Inc., USA). For tissue lysis, either half or whole bee heads or thoraces were flash frozen in liquid nitrogen and finely ground using a pestle. We then added 350µl Tissue Lysis Buffer, 20µl Proteinase K, and heated samples overnight at 50°C. After processing with the KingFisher System, samples were eluted using the Mag-Bind Kit Elution Solution (Thermo Fisher Scientific Inc., USA) to a final volume ranging from 70–85µl. DNA was quantified using NanoDrop™ 2000 Spectrophotometer (Thermo Fisher Scientific Inc., USA). DNA quality was assessed on 1.0% agarose gel using electrophoresis. Genome sequencing of new samples was carried out at The Centre for Applied Genomics (The Hospital for Sick Children, Ontario, Canada) using Illumina HiSeqX to generate 150bp pair-end reads.
Sequence alignment, variant discovery, and filtration

Sequencing reads of *Apis mellifera* and *Apis cerana* were trimmed of adapters and low quantity bases (<20) using Trimmomatic v0.36 (55). Trimmed reads were retained for downstream assembly if >50bps and >35bps in length from 100-150bp or 50bp read length Illumina data respectively. Reads were then aligned to the *Apis mellifera* reference genome (Amel 4.5)(56) using default parameters of NextGenMap aligner v0.4.12 (57). Midway through our analysis, a new assembly of the honey bee genome was released (88); we elected to continue to use Amel 4.5 to be able to utilize the large transcriptomic and functional genomic datasets generated using Amel 4.5. The average percentage of mapped reads across samples was 96.6%, and we are confident that there was little information loss. BAM files were sorted using SAMTools v1.3.1 (89) and duplicate reads were marked using Picard v2.1.0 (https://broadinstitute.github.io/picard/). Base quality scores were recalibrated using GATK v3.7 BaseRecalibrator (59) using previously identified variants as reference (17, 90). Variants were identified by constructing intermediate gVCF files for each genome using HaplotypeCaller GATK v3.7 (58, 59), and then samples were aggregated using GenotypeGVCFs GATK v3.7. Variants were filtered using VariantRecalibrator GATK v3.7 using previously identified variants as reference (17, 90) in addition to the following hard filter thresholds: MQ < 40.0, QD < 5.0, FS > 11.0, MQRankSum -2.0 < x > 2.0, and ReadPosRankSum -2.0 < x > 2.0. In addition, we excluded variants located within five base pairs of an indel and within five base pairs of areas with low complexity (90), we excluded loci with greater than 20% missing data, and excluded variants from the unmapped scaffolds using GATK v3.7.

Supplementary Text
Sample inclusion and population classification

The ADMIXTURE and PCA analyses revealed seven genetically distinct clusters. The hierarchical structure analysis grouped subspecies into their respective lineages, except four samples that were labeled as outliers. *A. m. intermissa* samples were grouped into their own cluster, mirroring results of the admixture analysis, likely the result of high admixture. The A lineage was divided into two clusters, one composed of putatively known *A. m. monticola* samples, and the second composed of remaining African subspecies and *A. m. scutellata*. Because there is little definition of subspecies groupings within the A lineage cluster, the two aforementioned A lineage groups were maintained for subspecies level analyses. Based on results from the preceding structure analyses, all seven genetically distinct lineages, and the following subspecies classifications were retained: *A. m. mellifera* (M lineage), *A. m. iberiensis* (M lineage), *A. m. sinisxinyuan* (M lineage), *A. m. carnica* (C lineage), *A. m. ligustica* (C lineage), *A. m. anatoliaca* (O lineage), *A. m. caucasica* (O lineage), *A. m. syriaca* also known as *A. m. yemenitica* (Y lineage), *A. m. lamarckii* (L lineage), *A. m. scutellata* (A lineage), *A. m. monticola* (A lineage), and *A. m. unicolor* (U lineage). The following samples were excluded from analyses: samples that ~50% of ancestry attribute to an unexpected lineage (n=3), *A. m. intermissa* (n=16), genetic outliers identified in the hierarchical structure analysis (n=4), and haploid samples (n=8) (Data S1). For analyses that are lineage focused, we retained samples that had ≥90% assignment to one ancestral lineage to reduce the effects of admixture. At the subspecies level, ancestry exceptions were made for uniformly admixed subspecies; *A. m. syriaca*, *A. m. pomonella*, and *A. m. sinisxinyuan*. Additionally, we exclude individual samples that, while grouped with the expected
lineage, did not clearly cluster with a putatively identified subspecies or population in the phylogenetic analyses, which made classification ambiguous (Data S1).

**Identification of outlier loci**

The A lineage has the least amount (2 SNPs) of outlier SNPs among all lineages. To account for the large sample size in the A lineage, we repeated the outlier detection analysis with ten representative samples but were only able to identify ~80 outlier SNPs. It is hypothesized that low outlier SNPs within the A lineage may be attributed to demographic effects resulting in substantially lower than expected F<sub>ST</sub> values (91). Additionally, considering the longer evolutionary time between the A, L, and U lineages, outlier markers may be ‘shared’ among lineages and thus not represented as unique to the A lineage. Finally, it is also postulated that local adaption among the subspecies that make up the A lineage has disrupted or prevented congruent shifts in allele frequency that would result in high F<sub>ST</sub> at the lineage level.
Fig. S1. Genetic clustering of *Apis mellifera* samples using nuclear SNPs. Patterns of ancestry for all *Apis mellifera* samples as estimated by the program ADMIXTURE using 1M loci that were selected among a pool of SNPs pruned for bi-allelic loci with a MAF > 0.05. Vertical bars represent individual bees and coloured segments represent the proportion of ancestry estimated to K=3-10 genetic clusters. When the ADMIXTURE analysis is conducted with K predictive clusters 3-6, there is evidence for common ancestry between the O and C lineage, and between the A, L, U, and Y lineages. When K is increased above 7, we begin to see the division of subspecies into separate clusters.
Fig. S2. Genetic clustering of *Apis mellifera* samples using unlinked nuclear SNPs. Patterns of ancestry for all *Apis mellifera* samples as estimated by the program ADMIXTURE using unlinked SNPs (38,493). Vertical bars represent individual bees and coloured segments represent the proportion of ancestry estimated to K=6-8 genetic clusters.
Fig. S3. Principal component analysis of diploid *Apis mellifera* samples using nuclear SNPs.
The first two principal components grouped *Apis mellifera* samples into seven distinct clusters representative of the M, C, O, Y, L, A, and U lineages. The PCA analysis also clearly illustrates the effects of admixture on the genetic relatedness among lineages. For example, *A. m. internissa*, which is highly admixed with M lineage ancestry, is situated between the A and M lineage clusters. Similarly, the putatively identified O lineage *A. m. pomonella* samples from Kyrgyzstan, which have high C lineage ancestry, are adjacent to the C lineage cluster. These relationships correlate with clustering patterns detected in the ADMIXTURE analysis.
Fig. S4. Neighbor-joining phylogeny of *Apis mellifera* samples using SNPs located genome wide. The entire SNP dataset was pruned of ambiguous loci, as implemented by RAxML v8.2.12 (63), and loci with low coverage (<0.8) in *Apis cerana*. The neighbor-joining tree were constructed using SNPs located genome wide (2,126,091) using allele-sharing distance and was rooted with *Apis cerena*. Major nodes are labeled by bootstrap support.
Fig. S5. Maximum-likelihood Phylogeny of *Apis mellifera* samples using SNPs located genome wide. The program TreeMix v1.13 (66) was used to produced maximum-likelihood trees using SNPs located genome wide that were further pruned for biallelic loci (1,884,783). The analysis was performed with SNPs formatted as allele frequencies with samples grouped into their respective lineages (A) and subspecies (B) grouping and was rooted using *Apis cerana*. 
**Fig. S6. Phylogeny of* Apis mellifera using protein coding SNPs.** The entire SNP dataset was pruned of ambiguous loci, as implemented by RAxML v8.2.12 (63), and loci with low coverage (<0.8) in *Apis cerana*. (A) A neighbor-joining tree constructed with SNPs located within protein coding regions (276,602) using allele-sharing distance and rooted with *Apis cerana*. Major nodes are labeled by bootstrap support. (B) A maximum-likelihood tree constructed with SNPs located within protein coding regions (276,602) using the gamma model of rate heterogeneity (ASC_GTRGAMMA) with the Lewis ascertainment bias correction in RAxML v8.2.12 (63). The tree was rooted with *Apis cerana*. Major nodes are labeled by bootstrap support.
**Fig. S7. Phylogeny of *Apis mellifera* using a subsample of SNPs.** The entire SNP dataset was pruned of ambiguous loci, as implemented by RAxML v8.2.12 (63), and loci with low coverage (<0.8) in *Apis cerana*. (A) A neighbor-joining tree constructed with randomly selected SNPs located among intrageneric and intergenic regions (276,602) using allele-sharing distance and rooted with *Apis cerana*. Major nodes are labeled by bootstrap support. (B) A maximum-likelihood tree constructed with randomly selected SNPs located among intrageneric and intergenic regions (276,602) using the gamma model of rate heterogeneity (ASC_GTRGAMMA) with the Lewis ascertainment bias correction in RAxML v8.2.12 (63). The tree was rooted with *Apis cerana*. Major nodes are labeled by bootstrap support.
Fig S8. Divergence dating applied to both topologies resolved by the phylogenetic reconstruction of *Apis mellifera*. (A) Divergence time estimates applied to the phylogeny resolved using SNPs located genome wide (2,126,091). Estimated divergence times are labeled at the nodes, and purple bars at the nodes illustrate the 95% confidence interval of divergence times. (B) Divergence time estimates applied to the phylogeny resolved using SNPs located within coding regions (276,602). Estimated divergence times are labeled at the nodes, and purple bars at the nodes illustrate the 95% confidence interval of divergence times.
**Fig. S9. Biogeographic range estimation applied to Apis mellifera.** We applied a biogeographic range estimation to the topology (A) resolved using SNPs located genome wide (2,126,091) and to the topology (B) resolved using SNPs located within protein coding regions (276,602). We defined three biogeographic areas based on the current *Apis mellifera* distribution: Europe [E], Africa [F], and Asia [A] and tested all six biogeographic models provided by BioGeoBEARS (69, 70). We used the Akaike Information Criterion (AIC) and the Log of the likelihood scores (LnL) to compare models and determine the best fit to the phylogeny. The best fit model for each phylogeny (A-B) is highlighted in Table S1. Pie charts at nodes indicate the marginal maximum likelihood probabilities for the estimated ancestral range. Probabilities for each labeled node (1-23) in topologies (A-B) are outlined in Table S2.
Fig. S10. Microregional biogeographic range estimation applied to *Apis mellifera*. We applied a biogeographic range estimation to the topology (A) resolved using SNPs located genome wide (2,126,091) and to the topology (B) resolved using SNPs located within protein coding regions (276,602). We defined six biogeographic areas based on the current *Apis mellifera* distribution: East Asia [EA], West Asia [WA], East Europe [EE], West Europe [WE], North Africa [NAf] and Africa [Af] and tested all six biogeographic models provided by BioGeoBEARS (69, 70). We restricted the analysis to only include states that had adjacent ranges, and limited the ranges occupied by the species to three (such as in the three-continent model; Fig S9). We used the Akaike Information Criterion (AIC) and the Log of the likelihood scores (LnL) to compare models and determine the best fit to the phylogeny, which was the DEC +J model for both topologies. The most likely ancestral regions are indicated at the nodes by letters corresponding to the defined biogeographic areas or combined states.
Fig. S11. Linkage disequilibrium decay among lineages. Linkage disequilibrium (LD) was measured for each lineage as the average squared correlation coefficient ($r^2$). LD was graphed as the average measure of $r^2$ per increasing intervals of 20bp up to a distance of 5000bp. LD decays by half at an average distance of 701bps among lineages, and decays the quickest among A lineage samples, reflective of a large effective population size.
Fig. S12. Linkage disequilibrium decay among subspecies. Linkage disequilibrium (LD) was measured for each lineage as the average squared correlation coefficient ($r^2$). LD was graphed as the average measure of $r^2$ per increasing intervals of 20bp up to a distance of 5000bp. LD decays by half at an average distance of 716bp bps among lineages. (A) LD decay of M lineage subspecies, (B) LD decay of C lineage subspecies, (C) LD decay of O lineage subspecies, (D) LD decay of A, L, Y and U lineage subspecies.
Fig. S13. Measures of shared genetic drift between lineages and subspecies. We calculated outgroup $f_3$ statistics, which is robust to lineage specific drift, to quantify the genetic distance between populations relative to an outgroup, *Apis cerana*. Higher $f_3$ values between lineage suggests greater shared genetic drift. (A) $f_3$ values estimates between lineages, (B) $f_3$ values estimates between subspecies.
Table S1: Models of ancestral range estimation applied to Fig. S9 (A-B). The best fit model to both topologies was DEC+J. LnL: Log-likelihood, k: number of parameters fitted to the model, d: rate of anagenetic range expansion, e: rate of anagenetic range contraction, j: weight of jump dispersal, AIC: Akaike Information Criterion.

| Figure  | Model | LnL  | k  | d            | e            | j            | AIC  |
|---------|-------|------|----|--------------|--------------|--------------|------|
| Fig. S9. A. | DEC   | -15.18 | 2  | 2.83E+00     | 1.00E-12     | 0.00E+00     | 34.37 |
|         | DEC+J | -9.81  | 3  | 1.00E-12     | 1.00E-12     | 1.16E-01     | **25.61** |
|         | DIVA  | -15.84 | 2  | 1.21E+00     | 4.76E+00     | 0.00E+00     | 35.67 |
|         | DIVA+J| -9.99  | 3  | 1.00E-12     | 4.06E-01     | 1.83E-01     | 25.99 |
|         | BAYAREA | -21.70 | 2  | 3.33E+00     | 5.00E+00     | 0.00E+00     | 47.39 |
|         | BAYAREA+J | -11.23 | 3  | 1.06E+00     | 2.26E-01     | 1.42E-01     | 28.46 |
| Fig. S9. B. | DEC   | -15.15 | 2  | 2.21E+00     | 1.00E-12     | 0.00E+00     | 34.29 |
|         | DEC+J | -9.81  | 3  | 1.00E-12     | 1.00E-12     | 1.16E-01     | **25.63** |
|         | DIVA  | -15.48 | 2  | 2.47E+00     | 5.00E+00     | 0.00E+00     | 34.66 |
|         | DIVA+J| -9.85  | 3  | 1.00E-12     | 1.00E-12     | 1.22E-01     | 25.7  |
|         | BAYAREA| -39.54 | 2  | 9.58E-02     | 2.22E-01     | 0.00E+00     | 83.09 |
|         | BAYAREA+J| -10.09 | 3  | 1.00E-12     | 8.12E-08     | 1.07E-01     | 26.18 |
Table S2: Estimated probabilities of ancestral range for topologies in Fig. S9 (A-B). Node numbers refer to locations indicated on Fig. S10 (A-B). Each geographic location represents the three defined areas (Europe, Asia, Africa), or combined states.

| Figure | Node | Europe | Asia | Africa | Europe + Asia | Europe + Africa | Asia + Africa | Europe + Asia + Africa |
|--------|------|--------|------|--------|--------------|----------------|-------------|-----------------------|
| Fig. S9. A | 1    | 1.000  | 0.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 2    | 1.000  | 0.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 3    | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 4    | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 5    | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 6    | 1.000  | 0.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 7    | 1.000  | 0.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 8    | 0.000  | 0.000| 1.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 9    | 0.000  | 0.000| 1.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 10   | 0.000  | 0.000| 1.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 11   | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 12   | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 13   | 0.015  | 0.675| 0.015  | 0.121        | 0.012          | 0.013       | 0.002                 |
|         | 14   | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 15   | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 16   | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 17   | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 18   | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 19   | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 20   | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 21   | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 22   | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 23   | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
| Fig. S9. B | 1    | 1.000  | 0.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 2    | 1.000  | 0.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 3    | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 4    | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 5    | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 6    | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 7    | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 8    | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 9    | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|         | 10   | 0.000  | 1.000| 0.000  | 0.000        | 0.000          | 0.000       | 0.000                 |
|   |     |     |     |     |     |     |     |
|---|-----|-----|-----|-----|-----|-----|-----|
| 11| 0.000| 1.000| 0.000| 0.000| 0.000| 0.000| 0.000|
| 12| 0.000| 1.000| 0.000| 0.000| 0.000| 0.000| 0.000|
| 13| 0.008| 0.736| 0.015| 0.058| 0.001| 0.147| 0.035|
| 14| 0.030| 0.714| 0.060| 0.043| 0.004| 0.117| 0.031|
| 15| 0.123| 0.524| 0.246| 0.013| 0.020| 0.058| 0.015|
| 16| 0.266| 0.705| 0.000| 0.029| 0.000| 0.000| 0.000|
| 17| 0.317| 0.671| 0.000| 0.011| 0.000| 0.000| 0.000|
| 18| 1.000| 0.000| 0.000| 0.000| 0.000| 0.000| 0.000|
| 19| 0.317| 0.671| 0.000| 0.011| 0.000| 0.000| 0.000|
| 20| 0.000| 1.000| 0.000| 0.000| 0.000| 0.000| 0.000|
| 21| 1.000| 0.000| 0.000| 0.000| 0.000| 0.000| 0.000|
| 22| 0.000| 0.000| 1.000| 0.000| 0.000| 0.000| 0.000|
| 23| 0.000| 0.000| 1.000| 0.000| 0.000| 0.000| 0.000|
Table S3. Genetic variation of lineages. Genetic variation and measures of effective population size ($N_e$) for lineages using two different estimates of mutation rate ($\mu$). Genetic variation measured by: $\pi$: nucleotide diversity, $\theta_w$: Watterson’s estimator, $S$: segregating sites, $S_{\text{singletons}}$: singletons.

| Lineage | $N$ | $\pi$  | $\theta_w$ | $S$     | $S_{\text{singletons}}$ | $N_e$ where $\mu = 5.27 \times 10^{-9}$ | $N_e$ where $\mu = 3 \times 10^{-9}$ |
|---------|-----|--------|------------|---------|-------------------------|------------------------------------------|---------------------------------------|
| M       | 17  | 0.00158| 0.00183    | 1182374 | 413493                  | 115620                                   | 203105                                |
| C       | 15  | 0.00138| 0.00185    | 1158888 | 421866                  | 116960                                   | 205459                                |
| O       | 13  | 0.00181| 0.0017     | 1028854 | 281121                  | 107802                                   | 189372                                |
| Y       | 13  | 0.00187| 0.00196    | 1181707 | 331483                  | 123819                                   | 217508                                |
| L       | 9   | 0.00277| 0.00252    | 1369138 | 283530                  | 159153                                   | 279579                                |
| A       | 111 | 0.00354| 0.01012    | 9557653 | 3250486                 | 640249                                   | 1124704                               |
| U       | 8   | 0.00233| 0.0017     | 890294  | 140872                  | 107277                                   | 188451                                |
Table S4. Genetic variation of subspecies. Genetic variation and measures of effective population size \((N_e)\) for subspecies using two different estimates of mutation rate \((\mu)\). Genetic variation measured by: \(\pi\): nucleotide diversity, \(\theta_w\): Watterson’s estimator, \(S\): segregating sites, \(S_{\text{singletons}}\): singletons.

| Subspecies           | N   | \(\pi\) | \(\theta_w\) | \(S\)      | \(S_{\text{singletons}}\) | \(N_e\) where \(\mu = 5.27 \times 10^{-9}\) | \(N_e\) where \(\mu = 3 \times 10^{-9}\) |
|----------------------|-----|---------|-------------|------------|-----------------------------|----------------------------------|----------------------------------|
| A. m. monticola      | 35  | 0.00349 | 0.00768     | 5840100    | 2409655                     | 486049                           | 853826                           |
| A. m. scutellata     | 76  | 0.00352 | 0.00934     | 8264280    | 3047949                     | 590975                           | 1038147                          |
| A. m. carnica        | 7   | 0.00148 | 0.00155     | 779953     | 384561                      | 98070                            | 172276                           |
| A. m. ligustica      | 6   | 0.00167 | 0.00135     | 645534     | 267854                      | 85470                            | 150143                           |
| A. m. lamarckii      | 9   | 0.00277 | 0.00252     | 1369138    | 283530                      | 159153                           | 279579                           |
| A. m. iberiensis     | 8   | 0.00172 | 0.00147     | 773919     | 231068                      | 93255                            | 163818                           |
| A. m. melifera       | 5   | 0.00185 | 0.00138     | 618442     | 223847                      | 87412                            | 153554                           |
| A. m. sinisxinyuan   | 10  | 0.00177 | 0.00158     | 876114     | 185977                      | 100243                           | 176094                           |
| A. m. anatolicaia    | 7   | 0.00197 | 0.00173     | 872109     | 272238                      | 109649                           | 192617                           |
| A. m. caucasica      | 5   | 0.0018  | 0.00125     | 560203     | 195786                      | 79179                            | 139091                           |
| A. m. pomonellia     | 5   | 0.00218 | 0.00211     | 944590     | 392631                      | 133521                           | 234552                           |
| A. m. syriaca        | 9   | 0.00254 | 0.00299     | 1628875    | 574300                      | 189345                           | 332617                           |
| A. m. unicolor       | 8   | 0.00233 | 0.0017      | 890294     | 140872                      | 107277                           | 188451                           |
| A. m. jemenitica     | 13  | 0.00187 | 0.00196     | 1181707    | 331483                      | 123819                           | 217508                           |
Table S5. Measures of pairwise genetic differentiation calculated with weighted $F_{ST}$ between lineages.

|    | M    | C    | O    | Y    | L    | A    | U    |
|----|------|------|------|------|------|------|------|
| M  | 0.000|      |      |      |      |      |      |
| C  | 0.664| 0.000|      |      |      |      |      |
| O  | 0.660| 0.553| 0.000|      |      |      |      |
| Y  | 0.583| 0.668| 0.625| 0.000|      |      |      |
| L  | 0.657| 0.609| 0.557| 0.533| 0.000|      |      |
| A  | 0.373| 0.379| 0.352| 0.335| 0.196| 0.000|      |
| U  | 0.655| 0.686| 0.635| 0.614| 0.481| 0.274| 0.000|
Table S6. Measures of pairwise genetic differentiation calculated with weighted $F_{ST}$ between subspecies.

|       | A. m. carnica | A. m. ligustica | A. m. monticola | A. m. scutellata | A. m. unicolor | A. m. caucasica | A. m. anatolica | A. m. pomonella | A. m. syriaca | A. m. iberiensis | A. m. sinisxinyuan | A. m. lamarcki | A. m. jementica |
|-------|---------------|-----------------|-----------------|-----------------|--------------|----------------|----------------|----------------|-------------|----------------|-------------------|----------------|----------------|
| A. m. carnica | 0.000         |                 |                 |                 |              |                |                |                |              |                |                   |                |                |
| A. m. ligustica | 0.129         | 0.000           |                 |                 |              |                |                |                |              |                |                   |                |                |
| A. m. monticola | 0.407         | 0.403           | 0.000           |                 |              |                |                |                |              |                |                   |                |                |
| A. m. scutellata | 0.371         | 0.368           | 0.038           | 0.000           |              |                |                |                |              |                |                   |                |                |
| A. m. unicolor | 0.670         | 0.665           | 0.307           | 0.295           | 0.000        |                |                |                |              |                |                   |                |                |
| A. m. caucasica | 0.638         | 0.658           | 0.385           | 0.352           | 0.656        | 0.000          |                |                |              |                |                   |                |                |
| A. m. anatolica | 0.556         | 0.578           | 0.376           | 0.336           | 0.621        | 0.160          | 0.000          |                |              |                |                   |                |                |
| A. m. pomonella | 0.517         | 0.534           | 0.326           | 0.283           | 0.595        | 0.068          | 0.147          | 0.000          |              |                |                   |                |                |
| A. m. syriaca | 0.456         | 0.467           | 0.285           | 0.249           | 0.518        | 0.263          | 0.200          | 0.210          | 0.000        |                |                   |                |                |
| A. m. mellifera | 0.724         | 0.711           | 0.383           | 0.351           | 0.651        | 0.716          | 0.687          | 0.605          | 0.567        | 0.000          |                   |                |                |
| A. m. iberiensis | 0.701         | 0.690           | 0.404           | 0.367           | 0.647        | 0.719          | 0.675          | 0.624          | 0.569        | 0.108          | 0.000             |                |                |
| A. m. sinisxinyuan | 0.594       | 0.581           | 0.382           | 0.345           | 0.613        | 0.644          | 0.604          | 0.539          | 0.510        | 0.207          | 0.260             | 0.000          |                |
| A. m. lamarckii | 0.572         | 0.565           | 0.242           | 0.205           | 0.481        | 0.553          | 0.525          | 0.490          | 0.415        | 0.551          | 0.556             | 0.526          | 0.000          |
| A. m. jementica | 0.650         | 0.650           | 0.379           | 0.343           | 0.614        | 0.641          | 0.609          | 0.589          | 0.530        | 0.654          | 0.666             | 0.631          | 0.533          |
Table S7. Summary of outlier SNPs among lineages. Number of outlier SNPs per lineage based on pairwise measures of $F_{ST}$. Number of genes associated with outlier SNPs for each lineage.

| Lineage | Number of outlier SNPs | Genes |
|---------|------------------------|-------|
| M (Europe) | 6479                    | 1538  |
| M (Asia)  | 3885                    | 990   |
| C        | 10428                   | 2539  |
| O        | 7105                    | 2012  |
| Y        | 10854                   | 2784  |
| L        | 12148                   | 2457  |
| A        | 2                       | 1     |
| U        | 14434                   | 2466  |
Table S8. Enrichment of outlier SNPs among genic and promoter regions. Chi-square ($\chi^2$) test for the enrichment of outlier SNPs concentrated within three functional annotation categories (promoter, protein coding, and intronic) when compared to the expected genomic distribution. The percent change represents the percent increase or decrease of observed values compared to expected values. P-values were corrected for false discovery rate (FDR) using the Benjamini-Hochberg for each functional category.

| Lineage | Annotation Region | Observed number of Outlier SNPs | Expected number of Outlier SNPs | $\chi^2$ | P-value | FDR | Percent Change |
|---------|-------------------|---------------------------------|-------------------------------|---------|---------|-----|---------------|
| M (Europe) | Promoter          | 319                             | 257                           | 15.23   | 9.50E-05| 1.66E-04| 24.12         |
| M (Asia) | Promoter          | 151                             | 154                           | 0.05    | 8.22E-01| 8.22E-01| -1.95         |
| C       | Promoter          | 550                             | 414                           | 46.33   | 9.98E-12| 3.49E-11| 32.85         |
| O       | Promoter          | 319                             | 282                           | 4.91    | 2.67E-02| 3.12E-02| 13.12         |
| Y       | Promoter          | 544                             | 431                           | 30.74   | 2.96E-08| 6.91E-08| 26.22         |
| L       | Promoter          | 654                             | 482                           | 63.54   | 1.57E-15| 1.10E-14| 35.68         |
| U       | Promoter          | 649                             | 573                           | 10.39   | 1.26E-03| 1.77E-03| 13.26         |
| M (Europe) | Protein Coding   | 606                             | 468                           | 43.34   | 4.59E-11| 6.43E-11| 29.49         |
| M (Asia) | Protein Coding   | 302                             | 281                           | 1.64    | 2.01E-01| 2.01E-01| 7.47          |
| C       | Protein Coding   | 1087                            | 754                           | 158.69  | 2.18E-36| 1.53E-35| 44.16         |
| O       | Protein Coding   | 692                             | 514                           | 66.52   | 3.46E-16| 6.06E-16| 34.63         |
| Y       | Protein Coding   | 1008                            | 785                           | 68.43   | 1.32E-16| 3.08E-16| 28.41         |
| L       | Protein Coding   | 1176                            | 878                           | 108.87  | 1.74E-25| 6.09E-25| 33.94         |
| U       | Protein Coding   | 1097                            | 1044                          | 2.91    | 8.78E-02| 1.02E-01| 5.08          |
| M (Europe) | Intron          | 2572                            | 2668                          | 5.81    | 1.59E-02| 2.79E-02| -3.60         |
| M (Asia) | Intron           | 1507                            | 1600                          | 9.05    | 2.62E-03| 6.12E-03| -5.81         |
| C       | Intron           | 4125                            | 4294                          | 11.28   | 7.85E-04| 2.75E-03| -3.94         |
| O       | Intron           | 2884                            | 2926                          | 0.99    | 3.21E-01| 3.21E-01| -1.44         |
| Y       | Intron           | 4397                            | 4469                          | 1.97    | 1.60E-01| 1.87E-01| -1.61         |
| L       | Intron           | 4816                            | 5002                          | 11.77   | 6.02E-04| 2.75E-03| -3.72         |
| U       | Intron           | 5858                            | 5944                          | 2.08    | 1.49E-01| 1.87E-01| -1.45         |
Table S9. Significance of overlap among genes under selection between lineages. A one-way Fisher’s Exact test to evaluate whether the overlap of genes with outlier SNPs between lineages is greater than expected by chance. *P*-values were corrected for false discovery rate (FDR) using the Benjamini-Hochberg correction.

| Population 1             | Population 2         | Observed Overlap | *P*-value   | FDR       |
|--------------------------|----------------------|------------------|-------------|-----------|
| M lineage (Europe)       | C lineage            | 640              | 1.61E-100   | 1.69E-100 |
| M lineage (Asia)         | C lineage            | 556              | 3.38E-157   | 5.92E-157 |
| M lineage (Europe)       | L lineage            | 660              | 1.90E-120   | 2.34E-120 |
| M lineage (Asia)         | L lineage            | 475              | 4.28E-104   | 4.74E-104 |
| C lineage                | L lineage            | 991              | 4.95E-174   | 1.15E-173 |
| O lineage                | L lineage            | 864              | 9.41E-165   | 1.80E-164 |
| U lineage                | L lineage            | 995              | 3.37E-172   | 7.08E-172 |
| Y lineage                | L lineage            | 1110             | 8.31E-194   | 2.49E-193 |
| M lineage (Europe)       | M lineage (Asia)     | 283              | 3.57E-50    | 3.57E-50  |
| M lineage (Europe)       | O lineage            | 610              | 1.23E-135   | 1.84E-135 |
| M lineage (Asia)         | O lineage            | 450              | 1.02E-121   | 1.34E-121 |
| C lineage                | O lineage            | 914              | 1.11E-201   | 5.07E-201 |
| M lineage (Europe)       | U lineage            | 701              | 1.35E-144   | 2.18E-144 |
| M lineage (Asia)         | U lineage            | 485              | 3.45E-110   | 4.02E-110 |
| C lineage                | U lineage            | 1036             | 1.21E-201   | 5.07E-201 |
| O lineage                | U lineage            | 915              | 3.25E-198   | 1.14E-197 |
| M lineage (Europe)       | Y lineage            | 816              | 2.13E-188   | 5.60E-188 |
| M lineage (Asia)         | Y lineage            | 536              | 2.64E-122   | 3.69E-122 |
| C lineage                | Y lineage            | 1135             | 1.46E-213   | 1.02E-212 |
| O lineage                | Y lineage            | 1043             | 1.95E-241   | 2.05E-240 |
| U lineage                | Y lineage            | 1186             | 3.45E-243   | 7.24E-242 |
Table S10. Simulations of the expected gene overlap between lineages relative to the observed overlap.

| Population 1            | Population 2            | Expected Overlap (Min) | Expected Overlap (Mean) | Expected Overlap (Max) | Observed Overlap |
|-------------------------|-------------------------|------------------------|-------------------------|------------------------|-----------------|
| M lineage (Asia)        | C lineage               | 262                    | 302                     | 357                    | 640             |
| M lineage (Asia)        | C lineage               | 160                    | 195                     | 231                    | 556             |
| M lineage (Europe)      | L lineage               | 242                    | 291                     | 343                    | 660             |
| M lineage (Asia)        | L lineage               | 155                    | 189                     | 224                    | 475             |
| C lineage               | L lineage               | 429                    | 482                     | 546                    | 991             |
| O lineage               | L lineage               | 319                    | 388                     | 438                    | 864             |
| U lineage               | L lineage               | 415                    | 469                     | 546                    | 995             |
| Y lineage               | L lineage               | 457                    | 528                     | 594                    | 1110            |
| M lineage (Europe)      | M lineage (Asia)        | 89                     | 117                     | 149                    | 283             |
| M lineage (Europe)      | O lineage               | 198                    | 240                     | 277                    | 610             |
| M lineage (Asia)        | O lineage               | 120                    | 154                     | 188                    | 450             |
| C lineage               | O lineage               | 347                    | 296                     | 444                    | 914             |
| M lineage (Europe)      | U lineage               | 247                    | 293                     | 340                    | 701             |
| M lineage (Asia)        | U lineage               | 158                    | 189                     | 223                    | 485             |
| C lineage               | U lineage               | 430                    | 484                     | 536                    | 1036            |
| O lineage               | U lineage               | 333                    | 384                     | 439                    | 915             |
| M lineage (Europe)      | Y lineage               | 279                    | 332                     | 380                    | 816             |
| M lineage (Asia)        | Y lineage               | 177                    | 213                     | 255                    | 536             |
| C lineage               | Y lineage               | 489                    | 547                     | 604                    | 1135            |
| O lineage               | Y lineage               | 372                    | 433                     | 508                    | 1043            |
| U lineage               | Y lineage               | 478                    | 531                     | 602                    | 1186            |
Data S1. (separate file)
List of new and previously sequenced samples used in this study.

Data S2. (separate file)
List of genes associated with at least one significant SNP among lineages. NA indicates genes that did not possess at least one outlier SNP for the corresponding lineage. This gene list also includes references to phenotypes associated with genes under selection among all lineages.

Data S3. (separate file)
List of significant gene ontology (GO) terms for each lineage with Benjamini-Hochberg corrected $p$-values. NA indicates the GO term was not present for the corresponding lineage.
REFERENCES AND NOTES

1. M. C. Arias, W. S. Sheppard, Phylogenetic relationships of honey bees (Hymenoptera: Apinae: Apini) inferred from nuclear and mitochondrial DNA sequence data. *Mol. Phylogenet. Evol.* **37**, 25–35 (2005).

2. N. Lo, R. S. Gloag, D. L. Anderson, B. P. Oldroyd, A molecular phylogeny of the genus *Apis* suggests that the Giant Honey Bee of the Philippines, *A. breviligula* Maa, and the Plains Honey Bee of southern India, *A. indica* Fabricius, are valid species. *Syst. Entomol.* **35**, 226–233 (2010).

3. R. Raffiudin, R. H. Crozier, Phylogenetic analysis of honey bee behavioral evolution. *Mol. Phylogenet. Evol.* **43**, 543–552 (2007).

4. K. A. Dogantzis, A. Zayed, Recent advances in population and quantitative genomics of honey bees. *Curr. Opin. Insect Sci.* **31**, 93–98 (2019).

5. F. Ruttner, *Biogeography and Taxonomy of Honeybees* (Springer Berlin Heidelberg, 1988).

6. R. A. Ilyasov, M.-l. Lee, J.-i. Takahashi, H. W. Kwon, A. G. Nikolenko, A revision of subspecies structure of western honey bee *Apis mellifera*. *Saudi J. Biol. Sci.* **27**, 3615–3621 (2020).

7. U. Kotthoff, T. Wappler, M. S. Engel, Greater past disparity and diversity hints at ancient migrations of European honey bee lineages into Africa and Asia. *J. Biogeogr.* **40**, 1832–1838 (2013).

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

9. J. M. Cridland, N. D. Tsutsui, S. R. Ramírez, The complex demographic history and evolutionary origin of the western honey Bee, *Apis Mellifera. Genome Biol. Evol.* **9**, 457–472 (2017).
10. A. Wallberg, F. Han, G. Wellhagen, B. Dahle, M. Kawata, N. Haddad, Z. L. P. Simões, M. H. Allsopp, I. Kandemir, P. De la Rúa, C. W. Pirk, M. T. Webster, A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nat. Genet.* **46**, 1081–1088 (2014).

11. L. Garnery, J. M. Cornuet, M. Solignac, Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis. *Mol. Ecol.* **1**, 145–154 (1992).

12. C. W. Whitfield, S. K. Behura, S. H. Berlocher, A. G. Clark, J. S. Johnston, W. S. Sheppard, D. R. Smith, A. V. Suarez, D. Weaver, N. D. Tsutsui, Thrice out of Africa: Ancient and recent expansions of the honey bee, *Apis mellifera*. *Science* **314**, 642–645 (2006).

13. E. O. Wilson, *The Insect Societies* (Harvard Univ. Press, 1971).

14. B. A. Harpur, S. Minaei, C. F. Kent, A. Zayed, Management increases genetic diversity of honey bees via admixture. *Mol. Ecol.* **21**, 4414–4421 (2012).

15. F. Han, A. Wallberg, M. T. Webster, From where did the Western honeybee (*Apis mellifera*) originate? *Ecol. Evol.* **2**, 1949–1957 (2012).

16. C. M. Grozinger, A. Zayed, Improving bee health through genomics. *Nat. Rev. Genet.* **21**, 277–291 (2020).

17. B. A. Harpur, C. F. Kent, D. Molodtsova, J. M. D. Lebon, A. S. Alqarni, A. A. Owayss, A. Zayed, Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 2614–2619 (2014).

18. Z. L. Fuller, E. L. Niño, H. M. Patch, O. C. Bedoya-Reina, T. Baumgarten, E. Muli, F. Mumoki, A. Ratan, J. McGraw, M. Frazier, D. Masiga, S. Schuster, C. M. Grozinger, W. Miller, Genome-wide analysis of signatures of selection in populations of African honey bees (*Apis mellifera*) using new web-based tools. *BMC Genomics* **16**, 518 (2015).
19. N. J. Haddad, W. Loucif-Ayad, N. Adjlane, D. Saini, R. Manchiganti, V. Krishnamurthy, B. AlShagoor, A. M. Batainh, R. Mugasimangalam, Draft genome sequence of the Algerian bee Apis mellifera intermissa. *Genom. Data* 4, 24–25 (2015).

20. C. Chen, Z. Liu, Q. Pan, X. Chen, H. Wang, H. Guo, S. Liu, H. Lu, S. Tian, R. Li, W. Shi, Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinisxinyuan* n. ssp. *Mol. Biol. Evol.* 33, 1337–1348 (2016).

21. A. Wallberg, C. Schöning, M. T. Webster, M. Hasselmann, Two extended haplotype blocks are associated with adaptation to high altitude habitats in East African honey bees. *PLOS Genet.* 13, e1006792 (2017).

22. D. Henriques, K. A. Browne, M. W. Barnett, M. Parejo, P. Kryger, T. C. Freeman, I. Muñoz, L. Garnery, F. Highet, J. S. Johnstn, G. P. McCormack, M. A. Pinto, High sample throughput genotyping for estimating C-lineage introgression in the dark honeybee: An accurate and cost-effective SNP-based tool. *Sci. Rep.* 8, 8552 (2018).

23. D. Henriques, M. Parejo, A. Vignal, D. Wragg, A. Wallberg, M. T. Webster, M. A. Pinto, Developing reduced SNP assays from whole-genome sequence data to estimate introgression in an organism with complex genetic patterns, the Iberian honeybee (*Apis mellifera iberiensis*). *Evol. Appl.* 11, 1270–1282 (2018).

24. M. Parejo, D. Wragg, L. Gauthier, A. Vignal, P. Neumann, M. Neuditschko, Using whole-genome sequence information to foster conservation efforts for the european dark honey bee, *Apis mellifera mellifera*. *Front. Ecol. Evol.* 4, 140 (2016).

25. I. Muñoz, D. Henriques, L. Jara, J. S. Johnston, J. Chávez-Galarza, P. De La Rúa, M. A. Pinto, SNPs selected by information content outperform randomly selected microsatellite loci for delineating genetic identification and introgression in the endangered dark European honeybee (*Apis mellifera mellifera*). *Mol. Ecol. Resour.* 17, 783–795 (2017).
26. I. Muñoz, D. Henriques, J. S. Johnston, J. Chávez-Galarza, P. Kryger, M. A. Pinto, Reduced SNP panels for genetic identification and introgression analysis in the dark honey bee (*Apis mellifera mellifera*). *PLOS ONE* 10, e0124365 (2015).

27. M. A. Pinto, D. Henriques, J. Chávez-Galarza, P. Kryger, L. Garnery, R. van der Zee, B. Dahle, G. Soland-Reckeweg, P. De la Rúa, R. Dall'Olio, N. L. Carreck, J. S. Johnston, 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 (2014).

28. Y. Ji, The geographical origin, refugia, and diversification of honey bees (*Apis spp.*) based on biogeography and niche modeling. *Apidologie* 52, 367–377 (2021).

29. K. S. Lamm, B. D. Redelings, Reconstructing ancestral ranges in historical biogeography: Properties and prospects. *J. Syst. Evol.* 47, 369–382 (2009).

30. A. M. Harris, M. DeGiorgio, Admixture and ancestry inference from ancient and modern samples through measures of population genetic drift. *Hum. Biol.* 89, 21–46 (2017).

31. E. Corsetti, N. Azpiazu, Functional dissection of the splice variants of the Drosophila gene homothorax (hth). *Dev. Biol.* 384, 72–82 (2013).

32. S. Biswas, R. J. Russell, C. J. Jackson, M. Vidovic, O. Ganeshina, J. G. Oakeshott, C. Claudianos, Bridging the synaptic gap: Neuroligins and neurexin I in *Apis mellifera*. *PLOS ONE* 3, e3542 (2008).

33. B. A. Harpur, S. M. Kadri, R. O. Orsi, C. W. Whitfield, A. Zayed, Defense response in Brazilian honey bees (*Apis mellifera scutellata* × *spp.*) is underpinned by complex patterns of admixture. *Genome Biol. Evol.* 12, 1367–1377 (2020).

34. F. Mondet, A. Beaurepaire, A. McAfee, B. Locke, C. Alaux, S. Blanchard, B. Danka, L. C. Yves, Honey bee survival mechanisms against the parasite Varroa destructor: A systematic review of phenotypic and genomic research efforts. *Int. J. Parasitol.* 50, 433–447 (2020).
35. H. M. G. Lattorff, J. Buchholz, I. Fries, R. F. Moritz, A selective sweep in a Varroa destructor resistant honeybee (Apis mellifera) population. Infect. Genet. Evol. 31, 169–176 (2015).

36. V. Zanni, D. A. Galbraith, D. Annoscia, C. M. Grozinger, F. Nazzi, Transcriptional signatures of parasitization and markers of colony decline in Varroa-infested honey bees (Apis mellifera). Insect Biochem. Mol. Biol. 87, 1–13 (2017).

37. E. Amiri, J. J. Herman, M. K. Strand, D. R. Tarpy, O. Ruepell, Egg transcriptome profile responds to maternal virus infection in honey bees, Apis mellifera. Infect. Genet. Evol. 85, 104558 (2020).

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

39. R. Ashby, S. Forêt, I. Searle, R. Malesza, MicroRNAs in honey bee caste determination. Sci. Rep. 6, 18794 (2016).

40. Q. W. Chan, M. Y. Chan, M. Logan, Y. Fang, H. Higo, L. J. Foster, Honey bee protein atlas at organ-level resolution. Genome Res. 23, 1951–1960 (2013).

41. J. Chávez-Galarza, D. Henriques, J. S. Johnston, M. Carneiro, J. Rufino, J. C. Patton, M. A. Pinto, 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, 2973–2992 (2015).

42. J. Chávez-Galarza, L. Garnery, D. Henriques, C. J. Neves, W. Loucif-Ayad, J. S. Johnston, M. A. Pinto, Mitochondrial DNA variation of Apis mellifera iberiensis: Further insights from a large-scale study using sequence data of the tRNA^leu^-cox2 intergenic region. Apidologie 48, 533–544 (2017).
43. F. Cánovas, P. De la Rúa, J. Serrano, J. Galián, Geographical patterns of mitochondrial DNA variation in *Apis mellifera iberiensis* (Hymenoptera: Apidae). *J. Zool. Syst. Evol. Res.* **46**, 24–30 (2008).

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

45. L. Boardman, A. Eimanifar, R. Kimball, E. Braun, S. Fuchs, B. Grünwald, J. D. Ellis, The mitochondrial genome of the Spanish honey bee, *Apis mellifera iberiensis* (Insecta: Hymenoptera: Apidae), from Portugal. *Mitochondrial DNA Part B* **5**, 17–18 (2020).

46. M. A. El-Niweiri, R. F. Moritz, Mitochondrial discrimination of honeybees (*Apis mellifera*) of Sudan. *Apidologie* **39**, 566–573 (2008).

47. T. G. Hailu, P. D’Alvise, A. Tofilski, S. Fuchs, J. Greiling, P. Rosenkranz, M. Hasselmann, Insights into Ethiopian honey bee diversity based on wing geomorphometric and mitochondrial DNA analyses. *Apidologie* **51**, 1182–1198 (2020).

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

49. G. L. Conte, M. E. Arnegard, C. L. Peichel, D. Schluter, The probability of genetic parallelism and convergence in natural populations. *Proc. Biol. Sci.* **279**, 5039–5047 (2012).

50. Y. Ji, X. Li, T. Ji, J. Tang, L. Qiu, J. Hu, J. Dong, S. Luo, S. Liu, P. B. Frandsen, X. Zhou, S. H. Parey, L. Li, Q. Niu, X. Zhou, Gene reuse facilitates rapid radiation and independent adaptation to diverse habitats in the Asian honeybee. *Sci. Adv.* **6**, eabd3590 (2020).

51. M. L. Winston, O. R. Taylor, G. W. Otis, Some differences between temperate European and tropical African and South American honeybees. *Bee World* **64**, 12–21 (1983).
52. B. A. Harpur, A. Dey, J. R. Albert, S. Patel, H. M. Hines, M. Hasselman, L. Packer, A. Zayed, Queens and workers contribute differently to adaptive evolution in bumble bees and honey bees. Genome Biol. Evol. 9, 2395–2402 (2017).

53. K. A. Dogantzis, B. A. Harpur, A. Rodrigues, L. Beani, A. L. Toth, A. Zayed, Insects with similar social complexity show convergent patterns of adaptive molecular evolution. Sci. Rep. 8, 10388 (2018).

54. C. Chen, H. Wang, Z. Liu, X. Chen, J. Tang, F. Meng, W. Shi, Population genomics provide insights into the evolution and adaptation of the eastern honey bee (Apis cerana). Mol. Biol. Evol. 35, 2260–2271 (2018).

55. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014).

56. C. G. Elsik, K. C. Worley, A. K. Bennett, M. Beye, F. Camara, C. P. Childers, D. C. de Graaf, G. Debyser, J. Deng, B. Devreese, E. Elhaik, J. D. Evans, L. J. Foster, D. Graur, R. Guigo; HGSC production teams, K. J. Hoff, M. E. Holder, M. E. Hudson, G. J. Hunt, H. Jiang, V. Joshi, R. S. Khetani, P. Kosarev, C. L. Kovar, J. Ma, R. Maleszka, R. F. A. Moritz, M. C. Munoz-Torres, T. D. Murphy, D. M. Muzny, I. F. Newsham, J. T. Reese, H. M. Robertson, G. E. Robinson, O. Rueppell, V. Solovyev, M. Stanke, E. Stolle, J. M. Tsuruda, M. Van Vaerenbergh, R. M. Waterhouse, D. B. Weaver, C. W. Whitfield, Y. Wu, E. M. Zdobnov, L. Zhang, D. Zhu, R. A. Gibbs, Finding the missing honey bee genes: Lessons learned from a genome upgrade. BMC Genomics 15, 86 (2014).

57. F. J. Sedlazeck, P. Rescheneder, A. Von Haeseler, NextGenMap: Fast and accurate read mapping in highly polymorphic genomes. Bioinformatics 29, 2790–2791 (2013).

58. R. Poplin, V. Ruano-Rubio, M. A. De Pristo, T. J. Fennell, M. O. Carneiro, G. A. Van der Auwera, D. E. Kling, L. D. Gauthier, A. Levy-Moonshine, D. Roazen, K. Shakir, J. Thibault, S. Chandran, C. Whelan, M. Lek, S. Gabriel, M. J. Daly, B. Neale, D. G. Mac Arthur, E. Banks, Scaling accurate genetic variant discovery to tens of thousands of samples. BioRxiv, 201178 (2017).
59. G. A. Van der Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.1–11.10.33 (2013).

60. D. H. Alexander, K. Lange, Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinformatics* **12**, 246 (2011).

61. E. Linck, C. Battey, Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Mol. Ecol. Resour.* **19**, 639–647 (2019).

62. X. Zheng, D. Levine, J. Shen, S. M. Gogarten, C. Laurie, B. S. Weir, A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328 (2012).

63. R Core Team, R: A language and environment for statistical computing. **201**, (2013).

64. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

65. T. Jombart, adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).

66. E. Paradis, J. Claude, K. Strimmer, APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).

67. J. K. Pickrell, J. K. Pritchard, Inference of population splits and mixtures from genome-wide allele frequency data. *PLOS Genet.* **8**, e1002967 (2012).

68. S. R. Browning, B. L. Browning, Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
69. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).

70. N. J. Matzke, BioGeoBEARS: BioGeography with Bayesian (and likelihood) evolutionary analysis in R Scripts. *R package, version 0.2* **1**, 2013 (2013).

71. N. J. Matzke, Probabilistic historical biogeography: New models for founder-event speciation, imperfect detection, and fossils allow improved accuracy and model-testing. *Front. Biogeogr.* **5**, (2013).

72. R. H. Ree, S. A. Smith, Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Syst. Biol.* **57**, 4–14 (2008).

73. R. H. Ree, Detecting the historical signature of key innovations using stochastic models of character evolution and cladogenesis. *Evolution* **59**, 257–265 (2005).

74. F. Ronquist, Dispersal-vicariance analysis: A new approach to the quantification of historical biogeography. *Syst. Biol.* **46**, 195–203 (1997).

75. M. J. Landis, N. J. Matzke, B. R. Moore, J. P. Huelsenbeck, Bayesian analysis of biogeography when the number of areas is large. *Syst. Biol.* **62**, 789–804 (2013).

76. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin; 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

77. G. Watterson, On the number of segregating sites in genetical models without recombination. *Theor. Popul. Biol.* **7**, 256–276 (1975).

78. H. Liu, Y. Jia, X. Sun, D. Tian, L. D. Hurst, S. Yang, Direct determination of the mutation rate in the bumblebee reveals evidence for weak recombination-associated mutation and an approximate rate constancy in insects. *Mol. Biol. Evol.* **34**, 119–130 (2017).
79. N. Patterson, P. Moorjani, Y. Luo, S. Mallick, N. Rohland, Y. Zhan, T. Genschoreck, T. Webster, D. Reich, Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).

80. M. Raghavan, M. Steinrücken, K. Harris, S. Schiffels, S. Rasmussen, M. De Giorgio, A. Albrechtsen, C. Valdiosera, M. C. Ávila-Arcos, A.-S. Malaspinas, A. Eriksson, I. Moltke, M. Metspalu, J. R. Homburger, J. Wall, O. E. Cornejo, J. V. Moreno-Mayar, T. S. Korneliussen, T. Pierre, M. Rasmussen, P. F. Campos, P. de Barros Damgaard, M. E. Allentoft, J. Lindo, E. Metspalu, R. Rodríguez-Varela, J. Mansilla, C. Henrickson, A. Seguin-Orlando, H. Malmström, T. Stafford Jr., S. S. Shringarpure, A. Moreno-Estrada, M. Karmin, K. Tambets, A. Bergström, Y. Xue, V. Warmuth, A. D. Friend, J. Singarayer, P. Valdes, F. Balloux, I. Leboreiro, J. L. Vera, H. Rangel-Villalobos, D. Pettener, D. Luiselli, L. G. Davis, E. Heyer, C. P. E. Zollikofer, M. S. Ponce de León, C. I. Smith, V. Grimes, K.-A. Pike, M. Deal, B. T. Fuller, B. Arriaza, V. Standen, M. F. Luz, F. Ricaut, N. Guidon, L. Osipova, M. I. Voevoda, O. L. Posukh, O. Balanovsky, M. Lavryashina, Y. Bogunov, E. Khusnutdinova, M. Gubina, E. Balanovska, S. Fedorova, S. Litvinov, B. Malyarchuk, M. Derenko, M. J. Mosher, D. Archer, J. Cybulski, B. Petzelt, J. Mitchell, R. Worl, P. J. Norman, P. Parham, B. M. Kemp, T. Kivisild, C. Tyler-Smith, M. S. Sandhu, M. Crawford, R. Villems, D. G. Smith, M. R. Waters, T. Goebel, J. R. Johnson, R. S. Malhi, M. Jakobsson, D. J. Meltzer, A. Manica, R. Durbi, Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science* **349**, aab3884 (2015).

81. P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Flying* **6**, 80–92 (2012).

82. D. Molodtsova, B. A. Harpur, C. F. Kent, K. Seevananthan, A. Zayed, Pleiotropy constrains the evolution of protein but not regulatory sequences in a transcription regulatory network influencing complex social behaviors. *Front. Genet.* **5**, 431 (2014).

83. D. W. Huang, B. T. Sherman, R. A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
84. C. G. Elsik, A. Tayal, C. M. Diesh, D. R. Unni, M. L. Emery, H. N. Nguyen, D. E. Hagen, Hymenoptera Genome Database: Integrating genome annotations in HymenopteraMine. *Nucleic Acids Res.* **44**, D793-D800 (2016).

85. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

86. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

87. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

88. A. Wallberg, I. Bunikis, O. V. Pettersson, M.-B. Mosbech, A. K. Childers, J. D. Evans, A. S. Mikheyev, H. M. Robertson, G. E. Robinson, M. T. Webster, A hybrid de novo genome assembly of the honeybee, *Apis mellifera*, with chromosome-length scaffolds. *BMC Genomics* **20**, 275 (2019).

89. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and samtools. *Bioinformatics* **25**, 2078–2079 (2009).

90. B. A. Harpur, M. M. Guarna, E. Huxter, H. Higo, K.-M. Moon, S. E. Hoover, A. Ibrahim, A. P. Melathopoulos, S. Desai, R. W. Currie, S. F. Pernal, L. J. Foster, A. Zayed, Integrative genomics reveals the genetics and evolution of the honey bee’s social immune system. *Genome Biol. Evol.* **11**, 937–948 (2019).

91. P. G. Meirmans, P. W. Hedrick, Assessing population structure: FST and related measures. *Mol. Ecol. Resour.* **11**, 5–18 (2011).