



OPEN

# Mitochondrial genomes illuminate the evolutionary history of the Western honey bee (*Apis mellifera*)

Erik Tihelka<sup>1</sup>, Chenyang Cai<sup>2,3</sup>, Davide Pisani<sup>3,4</sup> & Philip C. J. Donoghue<sup>3</sup>

Western honey bees (*Apis mellifera*) are one of the most important pollinators of agricultural crops and wild plants. Despite the growth in the availability of sequence data for honey bees, the phylogeny of the species remains a subject of controversy. Most notably, the geographic origin of honey bees is uncertain, as are the relationships among its constituent lineages and subspecies. We aim to infer the evolutionary and biogeographical history of the honey bee from mitochondrial genomes. Here we analyse the full mitochondrial genomes of 18 *A. mellifera* subspecies, belonging to all major lineages, using a range of gene sampling strategies and inference models to identify factors that may have contributed to the recovery of incongruent results in previous studies. Our analyses support a northern African or Middle Eastern origin of *A. mellifera*. We show that the previously suggested European and Afrotropical cradles of honey bees are the result of phylogenetic error. Monophyly of the M, C, and O lineages is strongly supported, but the A lineage appears paraphyletic. *A. mellifera* colonised Europe through at least two pathways, across the Strait of Gibraltar and via Asia Minor.

As probably the single most significant pollinator of agricultural crops and wild plants and as a producer of a variety of foodstuffs with nutritional, medical, and cosmetic uses such as honey and propolis, the importance of understanding the evolutionary history of *Apis mellifera* Linnaeus, 1758 cannot be overstated. While the Western honey bee is native to Europe, Africa, the Middle East, and parts of Asia<sup>1</sup>, the ability of *A. mellifera* to colonise virtually all habitable biomes on Earth and adapt to diverse bioclimatic conditions is a living proof of the species' remarkable morphological and behavioural plasticity<sup>2</sup>. To date, over 30 separate subspecies (or 'geographical races') have been described<sup>3–7</sup>. The distinctiveness of subspecies is in many cases not readily apparent by examining live or pinned individuals and identification has to be carried out based on quantitative morphometric<sup>6,8,9</sup> or molecular analyses<sup>5,10,11</sup>. The classification of honey bee subspecies has important practical implications for apiculture. Beekeepers have long recognised that bee races differ in a number of behavioural traits such as calmness, swarming intensity, honey production, ability to utilize different sources of forage, and resistance to disease<sup>6,12–15</sup>. A pan-continental study of European subspecies has shown that locally-adapted stock enjoys better survival rates and may have lower pathogen levels<sup>16,17</sup>, highlighting the importance of conserving the genetic diversity of native honey bee populations<sup>7,18</sup>.

A reliable backbone phylogeny of the Western honey bee would allow apidologists to study the evolution of economically important traits in this species and the basis of its adaptation to different environmental conditions<sup>19</sup>. Although more gene sequences from across the entire native range of *A. mellifera* are available now than ever before<sup>20</sup>, a number of long-standing questions has still not been answered. Notably, the geographic origin of the Western honey bee has historically been a subject of controversies<sup>1,21,22</sup>. The most recent whole genome analysis has lent support to either a north eastern African or a Middle Eastern origin of *A. mellifera*<sup>23</sup>, but discriminating between these two hypotheses has been proven challenging using any dataset. In stark contrast,

<sup>1</sup>Department of Animal Science, Hartpury College, Hartpury GL19 3BE, UK. <sup>2</sup>State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, and Centre for Excellence in Life and Palaeoenvironment, Chinese Academy of Sciences, Nanjing 210008, China. <sup>3</sup>School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK. <sup>4</sup>School of Biological Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK. ✉email: erik.tihelka@hartpury.ac.uk; cycai@nigpas.ac.cn

analyses of complete mitochondrial genomes have recovered *A. m. mellifera* as the basalmost subspecies, thus placing the origin of honey bees within northern Europe<sup>24–30</sup>. A European origin of honey bees does not seem unreasonable, as the oldest unequivocal fossil representatives of the genus *Apis* are known from the Oligocene of France and Germany, and fossil European honey bees also show high degrees of morphological disparity<sup>4,31–34</sup>. On the other hand, given that all other living species of the genus *Apis* occur in Asia, the idea that *A. mellifera* may have originally dispersed from the east has enjoyed lasting popularity since the 1950s<sup>1,35–38</sup>. Unexpectedly, support for an origin of honey bees in tropical Africa was recovered by an analysis of 1,136 SNPs from 341 bees<sup>2</sup>. Notably, most recent phylogenomic studies investigating honey bee origins<sup>1,2,39</sup> have relied on the neighbour joining (NJ) algorithm which is prone to systematic errors and can yield misleading topologies<sup>40</sup>.

The relationships among the individual subspecies of *A. mellifera* are likewise unclear, with studies yielding ambiguous results. Traditionally, the Western honey bee has been divided into four evolutionary lineages: the A lineage (subspecies native to Africa), the M lineage (western and northern Europe), the C lineage (southern and eastern Europe), and the O lineage (Caucasus, Turkey, Middle East, Cyprus, Crete) based on morphometric and molecular data<sup>2,6,41–43</sup>. In recent years, two new purportedly isolated lineages have been proposed based on molecular data. Honey bees from Ethiopia deviate substantially from the A lineage into which they have been originally placed and were thus referred to a distinct group of their own, the Y lineage<sup>44</sup>. A sixth lineage from Syria and Lebanon has been identified as clearly divergent from the O lineage based on neighbour joining of microsatellite loci and mitochondrial DNA<sup>45,46</sup>. Here we refer to this proposed group as the S lineage. Moreover, some studies do not recognise the distinction between the C and O lineages<sup>21,38,47</sup>.

To shed light on the controversial inter-relationships among honey bee subspecies and to provide a reliable backbone phylogeny of *A. mellifera*, we used a collection of recently sequenced complete mitochondrial genomes representing more than half of the described honey bee subspecies from across its entire native range. We analysed our data with methods that allow for the identification of sources of phylogenetic incongruence, utilizing different gene sampling strategies and inference models.

## Material and methods

**Sequence data.** The mitochondrial genome of *A. mellifera* consists of around 16,463 bp and includes 13 protein coding genes, 22 transfer RNA (tRNA) genes, two ribosomal (rRNA) genes, and one control region<sup>48</sup>. All mitochondrial genomes of *A. mellifera* sequenced to date were obtained from GenBank in January 2020 alongside with the mitochondrial genomes of *A. cerana*, *A. florea*, and *A. dorsata*, which were used as outgroups. In total, 18 *A. mellifera* mitogenomes were analysed, representing more than half of its known subspecific diversity. GenBank accession numbers are provided in Table S1.

Data for the 13 protein coding genes and two RNA molecules were downloaded, aligned, and concatenated using PHYLOSUITE v1.2.1<sup>49</sup>. Protein-encoding genes were unambiguously aligned owing to few gaps and their codon-based structure using the G-INS-i algorithm implemented in the MAFFT v 7.313 plugin<sup>50</sup>. The ribosomal RNAs 16S and 18S were aligned in MAFFT using the E-INS-I algorithm.

**Phylogenomic analyses.** To test the effects of gene sampling on topology recovery, we prepared three datasets: the first and second codon positions only (P12), P12 and the two rRNA genes (P12RNA), and all codon positions together (P123). The third codon position of the protein-coding genes has been shown to suffer high degrees of saturation, which can potentially lead to biased phylogenies, and as such its exclusion is recommended to reduce data heterogeneity<sup>51,52</sup>.

To examine potential sources of phylogenomic conflict resulting from inappropriate model selection, we analysed the sequences using both site-heterogeneous Bayesian inference (BI) and site-homogeneous maximum likelihood (ML) methods. The inappropriate selection of inference models represents one of the key sources of phylogenomic error<sup>53</sup>. In particular, high compositional and rate heterogeneity of sequences can result into simpler site-homogeneous models yielding results affected by systematic error such as long branch attraction<sup>54,55</sup>. Such types of error are often difficult to notice, because they may be strongly supported and recovered consistently<sup>56</sup>. On the other hand, more complex site-heterogeneous models that account for compositional and rate heterogeneity generally fit data better and have consequently been applied to resolving difficult phylogenomic problems such as the origin of eukaryotes or the basal branching order in Metazoa<sup>54,55,57–63</sup>.

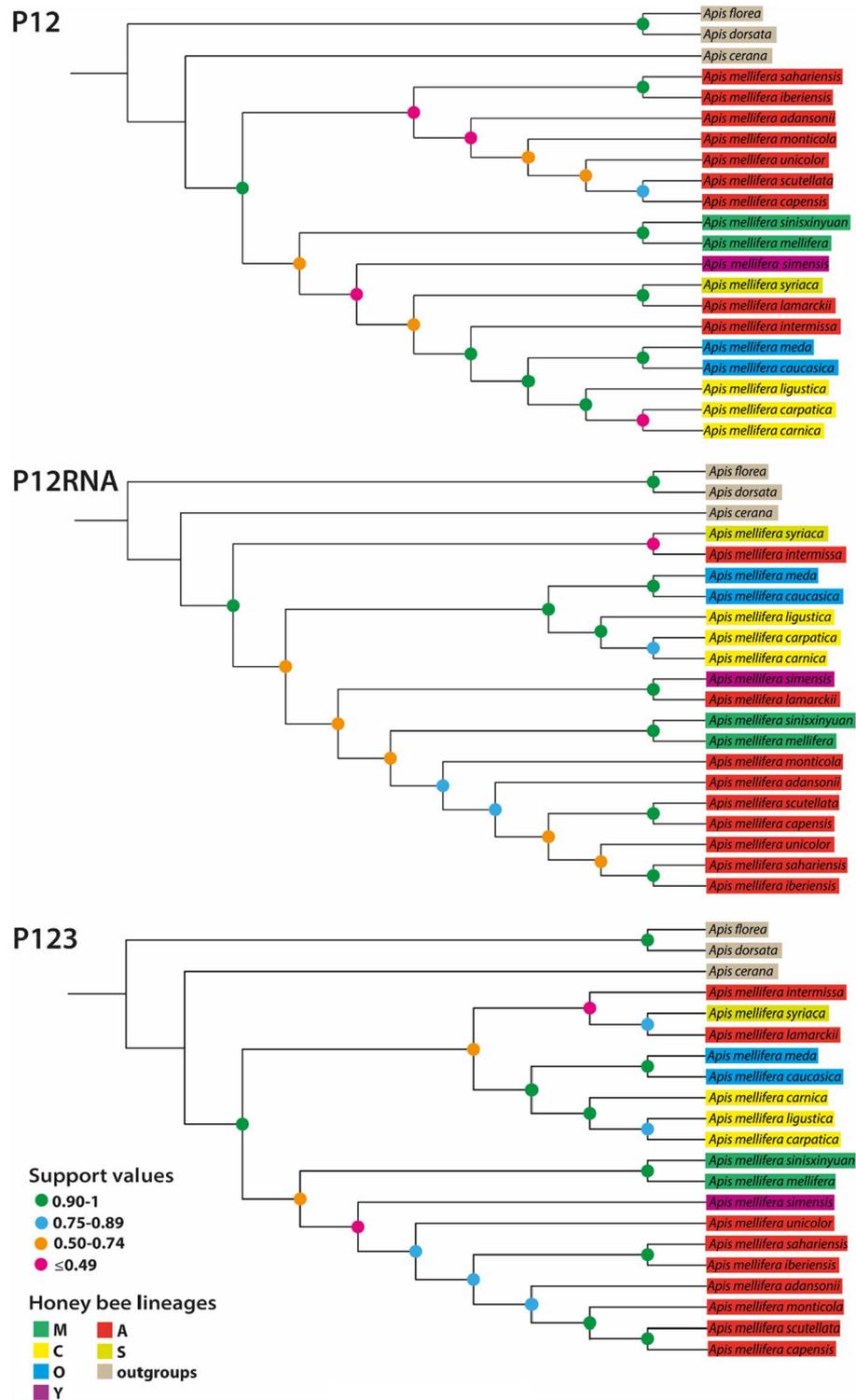
The BI site-heterogeneous mixture model CAT-GTR + G was run in PHYLOBAYES MPI 1.764. Two independent Markov chain Monte Carlo (MCMC) chains were run until convergence (maxdiff < 0.3). For each PHYLOBAYES run, we used the bpcomp program to generate output of the largest (maxdiff) and mean (meandiff) discrepancy observed across all bipartitions.

For the ML analyses, the concatenated datasets partitioned by codon were analysed with PARTITIONFINDER 2.1.1 to select the most appropriate models<sup>65</sup>. The ‘greedy’ algorithm was used to search ‘all’ models under the Bayesian information criterion, with branch lengths unlinked. The partitioning scheme is reported in Table S2. ML analyses using the selected models were performed in IQ-TREE v 1.6.12. Analyses were run using 1,000 ultra-fast bootstraps<sup>66</sup>.

All illustrations were prepared with iTOL v. 5.5.1<sup>67</sup>, Adobe Photoshop v. 21.2, and Adobe Illustrator v. 24.2.

## Results

The largest discrepancies (maxdiff) in all PHYLOBAYES runs were < 0.3, indicating they all represent phylogenetically informative runs<sup>68</sup>. Trees recovered with the CAT-GTR + G model are displayed in Fig. 1, trees with computed branch lengths and topologies recovered with the site-homogeneous ML models are presented in Figs. S1–S12. A phylogenetic hypothesis based on the CAT-GTR + G analysis of P12RNA is presented as the



**Figure 1.** Three competing topologies recovered based on different gene sampling approaches with the site-heterogeneous CAT-GTR + G model ignoring branch length for easier legibility. Subspecies are grouped into proposed molecular lineages primarily after Ruttner<sup>6</sup> and Meixner et al.<sup>5</sup>. Support values represent Bayesian posterior probabilities. Abbreviations: P12, first and second codon positions of protein coding mitochondrial genes; P12RNA, first and second codon positions of protein coding mitochondrial genes and two ribosomal (rRNA) genes; P123, protein coding mitochondrial genes without the third codon position excluded.

graphical abstract. Shallow relationships among subspecies typically received high support, although the same was not always the case for deeper nodes, as has been typical of most honey bee phylogenies conducted to date<sup>7,20</sup>.

Our analyses using the site-heterogeneous CAT-GTR + G model always recovered subspecies from northern Africa and the Middle East as sister to the remaining *A. mellifera* subspecies. Full support (Bayesian Posterior Probabilities [BPP] = 1) was recovered for the P12RNA dataset where *A. m. syriaca* and *A. m. intermissa* formed the basalmost clade. The former is native to the eastern Mediterranean region while the latter occurs in Tunisia, Algeria, and in Morocco between the Atlas Mountains and the Mediterranean and Atlantic coasts<sup>69,70</sup>. Analyses of the P12 and P123 datasets have recovered *A. m. sahariensis* (northwest Africa.) + *A. m. iberiensis* (Iberian Peninsula), and *A. m. syriaca* (Near East and Israel) + *A. m. intermissa* (northern Africa) as the basalmost subspecies, respectively, albeit with low support.

ML analyses of P12 and P12RNA have recovered the northern African *A. m. intermissa* as the basalmost subspecies with full support (Maximum Likelihood Bootstrap [MLB] = 100, Figs. S2, S4). A clade comprising *A. m. mellifera* and *A. m. sinisxinyuan* was recovered as sister to the remaining honey bees in the ML analysis of P123 (MLB = 100, Fig. S6).

Regardless of the dataset analysed or the model used, the monophyly of the M lineage was always recovered with maximal support (BPP = 1, MLB = 100). Likewise, lineages C and O were always monophyletic and together formed a monophyletic clade. Notably, the A lineage was never monophyletic in the PHYLOBAYES analyses, with a separation between the basal northern African subspecies and the more derived Afrotropical ones.

## Discussion

**The cradle of Western honey bees: Asia, Africa, Middle East, or Europe?.** Disentangling the phylogeny of *A. mellifera* has long been hampered by the limited availability of molecular data for the majority of honey bee subspecies. Molecular analyses have traditionally recovered conflicting relationships among honey bee races, resulting into uncertainties over the geographic origin of the species<sup>1</sup>.

An Asian origin of honey bees seems rather intuitive and it is probably for this reason that it has long enjoyed support in apidological circles<sup>1,35–37</sup>. Of the ten or so recognised species of the genus *Apis*, all except for *A. mellifera* occur in Asia<sup>71</sup>, so it would be reasonable to expect that this region is also the historical centre of the Western honey bee. The native range of *A. mellifera* in Asia includes Kazakhstan, Mongolia, the south of Russia, and a distinct *A. mellifera* subspecies further occurs in China<sup>3,72</sup>. Even though most recent genomic analyses of honey bees have failed to recover an Asian cradle of the Western honey bee, it has still been regarded as the most parsimonious explanation<sup>1,39</sup>; after all, the basal Asian populations may have simply gone extinct or were not sequenced yet. However, we caution that the evolutionary history of the genus *Apis* may be considerably more complex. Although the oldest members of the genus *Apis* are known from the Oligocene of Western Europe, these species are not necessarily basal and are morphologically similar to the extant giant honey bee *A. dorsata* native to south and southeast Asia<sup>32–34</sup>. This indicates that the genus *Apis* may have been present in Europe long before the origin of *A. mellifera*, and so there is no reason to expect an Asian origin of the species.

The hypothesis of honey bee genesis in tropical or subtropical Africa is relatively recent and has been proposed by the first SNP analysis of honey bee relationships by Whitfield et al.<sup>2</sup>. However, extensive reanalyses of Whitfield and colleagues' original dataset excluding samples with potentially hybrid origins<sup>4</sup> as well as analyses of new SNPs<sup>39</sup> have failed to find unequivocal support for the 'out of Africa' hypothesis. It is notable that all three studies<sup>1,2,39</sup> used the neighbour joining (NJ) algorithm to infer evolutionary relationships. While NJ is computationally superfast and can be used to give a quick reference tree, it is also prone to systematic errors<sup>40</sup> and as such is not generally recommended for phylogenomic studies<sup>56</sup>. Analyses that have used ML inference methods have rejected a deep African origin of *A. mellifera*<sup>23</sup>. Our analyses with the CAT-GTR + G model, that has been designed to specifically counter the effects of systematic error, also never recovered an Afrotropical origin of Western honey bees. It therefore seems that the 'out of Africa' result is a phylogenetic artefact probably caused by a systematic error.

Our analyses have recovered northern African subspecies, and occasionally the Middle Eastern *A. m. syriaca*, as the basalmost Western honey bees. This result is congruent with the largest molecular phylogeny based on whole honey bee genomes from across the species' native range<sup>23</sup> and with Ruttner's et al.<sup>21</sup> classical morphometric analysis of 33 characters on 404 bees. However, this finding is in contrast with ML analyses of honey bee mitochondrial genomes which have consistently recovered the dark European honey bee (*A. m. mellifera*, M lineage) as the basalmost subspecies. Analyses that have recovered a European cradle of *A. mellifera* have invariably used site-homogeneous models<sup>24–30</sup>. In our analyses, we have only recovered the M lineage as the basalmost group when the P123 dataset was analysed under ML. The same result was not recovered with ML analyses of the P12 and P12RNA datasets excluding the heterogeneous third codon position, nor when the data were analysed with the CAT-GTR + G model. Since the European origin of honey bees is only recovered in analyses that do not account for the biasing effects of data compositional heterogeneity, we conclude that it is most likely artefactual.

Overall, the origins of *A. mellifera* appear to lie in northern Africa or the Middle East. Distinguishing between these two regions is difficult at the present stage since only a few mitochondrial genomes of local bee races have been sequenced. The sister relationship between the African *A. m. intermissa* and the Middle Eastern *A. m. syriaca*, which have been recovered as the basalmost clade in our CAT-GTR + G analysis of P12RNA, suggests that the ancestral range of *A. mellifera* possibly encompassed both regions. Therefore, to narrow down the geographical origin of the Western honey bee, data for more subspecies occurring in the region will be required.

**Relationships among *Apis mellifera* subspecies.** Given the meagre amount of morphological variation among honey bee subspecies, they have historically been defined primarily by quantitative measurements of around 40 principal characters such colour, pilosity, wing venation characters, and the sizes of various body

parts<sup>6</sup>. Unfortunately, quantitative morphological characters are difficult to analyse phylogenetically<sup>73</sup> and suffer from widespread homoplasy<sup>74</sup>. Nevertheless, our analysis recovered strong support for three out of the four morphologically defined honey bee lineages: O, C, and M. The lineages O and C formed a strongly supported monophyletic group, an affinity already suggested by morphometry<sup>6</sup> and earlier analyses of mitochondrial markers<sup>38,74</sup>. The two M lineage subspecies, the European *A. m. mellifera* and the Chinese *A. m. sinisxinyuan*, were also recovered in all analyses. The apparent paraphyly of the A lineage is difficult to interpret, as it was not recovered with high support. Nonetheless, the separation of northern and south African honey bees into two distinct clades has previously been detected in analyses of a limited number of mitochondrial markers<sup>44,75</sup>, and so the problem requires further investigation.

Aside from the four traditional morphological groups, molecular studies have suggested the existence of at least two additional lineages. The Y lineage has been recognised as a distinctive grouping of honey bees from Ethiopia based on parsimony analyses of microsatellite markers and mitochondrial DNA<sup>44</sup>. The distinctiveness of Ethiopian bees, traditionally placed within the A lineage, has further been demonstrated by pheromone<sup>76</sup> and morphometric analyses<sup>77,78</sup>, although the latter also found a high degree of introgression between Ethiopian populations and the neighbouring well-defined African subspecies. We found a close proximity between the Ethiopian *A. m. simensis* and part of the paraphyletic A lineage, clearly separating it from the geographically close O lineage. These results are in line with the ML analysis of Cridland et al.<sup>23</sup>. The phylogenetic position of *A. m. simensis* was poorly supported or not supported at all in our CAT-GTR+G analyses. Therefore, the validity of *A. m. simensis* as a separate lineage cannot be rejected at present but is contingent upon future validation.

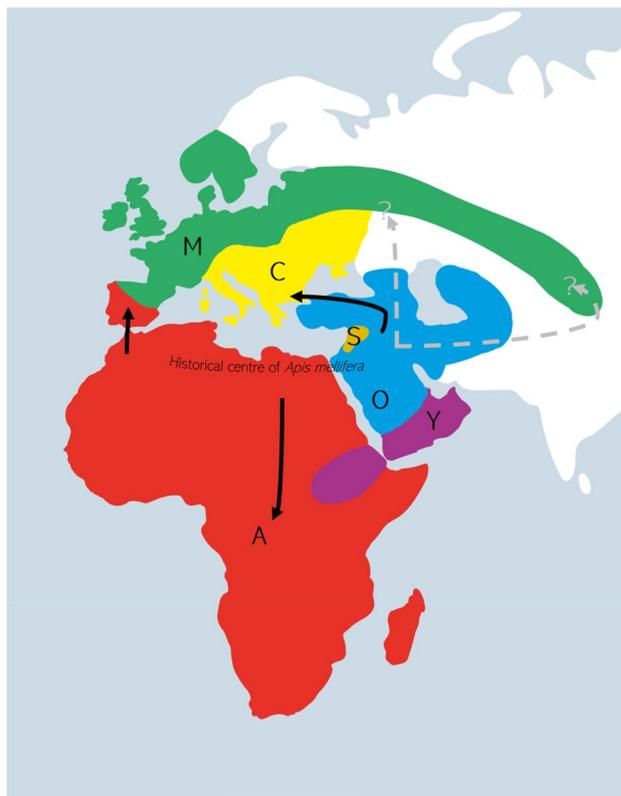
The S lineage, consisting of *A. m. syriaca*, has been considered as a separate bee lineage by Alburaki et al.<sup>46</sup>. This subspecies occurs throughout Syria, Lebanon and north Jordan and has been assigned to the A lineage based on morphological characters<sup>6,21</sup>, although a mitochondrial analysis placed it within the O lineage<sup>42</sup>. The position of the Syrian honey bee has been recovered with no or little support, although a sister relationship with *A. m. lamarkii* was found in the CAT-GTR+G analysis of the P12 dataset (BPP = 0.93, Fig. 1). As such the phylogenetic position of the Syrian bee remains an open question.

**Biogeographic history of Western honey bees.** Regardless of the dataset analysed, our Bayesian analyses consistently recovered a strongly supported sister group relationship between the closely co-occurring *A. m. sahariensis* native to the oasis regions of the Sahara Desert and *A. m. iberiensis* from Spain and Portugal (Fig. 1). The Iberian honey bee was nested within the A lineage in our analyses regardless of the dataset analysed. While it has been placed into the M lineage based on morphological characters<sup>6,79,80</sup>, molecular analyses indicate that the group has a significant degree of admixture with the A lineage, especially in the south of the Iberian peninsula, and consequently it has been assigned to the A lineage by some authors<sup>5,38,81,82</sup>. This indicates that a part of the ancestral northern African honey bee population probably colonised Europe across the Strait of Gibraltar. This hypothesis has been proposed by morphometric analyses<sup>21</sup> but was questioned by early analyses based on mitochondrial markers<sup>38,74</sup>. It should be noted that these analyses used parsimony and distance methods, which are outperformed by model-based BI and ML methods<sup>83,84</sup>. We therefore consider a cross-Gibraltar dispersal from northern Africa to Europe as highly likely (Fig. 2).

A second migration route from northern Africa and the Middle East to Europe likely occurred via Asia Minor or the Caucasus. This is suggested by the strongly supported clade comprising the lineages O and C (Fig. 1). While the basal O lineage bees occur in the Near East and Caucasus, the more derived C lineage bees are native to southern Europe. Whether honey bees dispersed into Europe via Turkey or also crossed the Caucasus and North Asia remains to be determined, as whole mitochondrial sequences are not available yet for the phylogenetically important subspecies *A. m. anatolica*, *A. m. macedonica*, and *A. m. cecropia*. We tentatively consider the former as more likely, since a cross-Turkish dispersal is supported by morphometric data<sup>21</sup>. Moreover, a distance analysis of ND2 sequences recovered the Turkish *A. m. anatolica* as sister to the Near Eastern *A. m. meda*<sup>85</sup>.

The dispersal of M lineage honey bees into Europe and Asia is hard to explain, since the position of the M lineage was recovered as close to A lineage bees in CAT-GTR+G analyses, but without strong support. Chen et al.<sup>3</sup> suggested that *A. m. mellifera* and *A. m. sinisxinyuan* are close to O and C lineage bees based on a NJ analysis. This result would imply that the origin of the M lineage lies in Asia Minor or the Near East. Clearly, the origin and position of the M lineage warrants further study.

Understanding what historical drivers controlled the dispersal of Western honey bees is difficult. The only unequivocal evidence about the timing of evolutionary events can be obtained from the fossil record. Ideally, it would be possible to use subfossil remains of honey bees identified to subspecies level to calibrate the *A. mellifera* phylogeny<sup>86</sup>, which would allow us to correlate splitting events within the tree with known bioclimatic events. Unfortunately, the subfossil record of *A. mellifera* is scant. Inclusions in copal are often cited as the oldest subfossils of Western honey bees, but the copal itself is of uncertain provenance<sup>87–89</sup>. Copal is moreover notoriously difficult to date and may be anywhere from 5 million to several years old<sup>90</sup>. Other honey bee remains are only known from archaeological sites, at most several thousand years old<sup>91,92</sup>. As such, studies have so far yielded highly divergent results; the split between *A. cerana* and *A. mellifera* has been variously estimated as having occurred between 6 and 25 million years ago<sup>38,72,93,94</sup> and the rapid divergence between *A. mellifera* subspecies between 0.3 and 1.3 million years ago<sup>6,38,75</sup>. Wallberg et al.<sup>39</sup> used the genealogical concordance method to estimate that honey bee subspecies diverged between 13,000–38,000 years ago, which would correspond to the last glacial maximum, implying that the expansion of *A. mellifera* from its region of origin into Europe began after the retreat of the ice sheets. The expansion of the honey bee into the Afrotropics may have been controlled by climate-induced desertification and vegetation shifts in the Pleistocene<sup>75</sup>. However, it is clear that the timescale of honey bee diversification will have to be revisited with more subfossil at hand to verify these hypotheses.



**Figure 2.** Hypothesis of *Apis mellifera* origin and dispersal routes. Solid black arrows are based on well-supported relationships inferred from Bayesian analysis of mitochondrial genomes while dashed grey arrows represent hypothetical dispersal routes. Lineage ranges are based largely on Ruttner<sup>6</sup> and Dogantzis and Zayed<sup>20</sup>.

### Future directions

While mitochondrial genomes provide a useful insight into the evolutionary history of the Western honey bee, several caveats must be pointed out. Only slightly more than half of the currently recognised honey bee subspecies have their mitochondrial genomes sequenced and some genomes from the phylogeographically important Middle East and Asia Minor are still lacking. This means that our understanding of honey bee origin and dispersal is still correspondingly incomplete.

The low support values recovered at some nodes could possibly result from the fact that the genomes of closely related honey bee subspecies are usually very similar<sup>1,23</sup>. Bootstrap support for groups is calculated as the proportion of times a given clade is recovered when a subset of the data is resampled<sup>95</sup>. As such, bootstrap support does not necessarily reflect phylogenetic signal, but assesses data redundancy, which is expected to be low in sequences with low variability<sup>96</sup>. Similarly, Bayesian posterior probabilities appear low when analysing sequences of potentially hybrid origin<sup>97</sup>. We expect that better supported topologies can be obtained by increasing gene sampling in future analyses.

Natural and human-induced hybridisation of honey bee populations has resulted into a considerable degree of genetic admixture among the subspecies<sup>98–102</sup>. This is the case for feral and managed bee populations within the Old World as well as for populations imported into the Americas and Australia<sup>103,104</sup>. As a result, most honey bee subspecies have not experienced long periods of isolation<sup>1</sup> and the evolutionary trees produced by us represent only pragmatic approximations of honey bee evolutionary history. Ultimately, it may be more appropriate to think of honey bee evolution as a web of intertwining populations rather than a strictly dichotomous branching tree. However, disentangling the honey bee ‘evolutionary web’ is likewise contingent upon obtaining more whole genome sequences from across its native range.

### Conclusions

1. The Western honey bee originated in northern Africa or the Near East. This conclusion is congruent with morphological<sup>21</sup> and the most extensive molecular<sup>23</sup> datasets.
2. Earlier hypotheses on an Afrotropic<sup>2</sup> or European<sup>27</sup> origin of *A. mellifera* appear to be erroneous. They are never recovered when methods for countering the effects of rate heterogeneity are applied to datasets.
3. *A. mellifera* colonised Europe from northern Africa or the Near East via at least two routes: across the Strait of Gibraltar and via Turkey, although a possible route via the Caucasus or North Asia is subject to validation once more genomes of subspecies native to these areas are sequenced.

4. The A lineage has been recovered as paraphyletic, split between north and south African subspecies, albeit with low support.
5. The O and C lineages form a strongly supported monophyletic group in all analyses. The basal O lineage bees are native to the Near East and Caucasus, while the more recently diverged C lineage bees occur in southern Europe.
6. *A. m. mellifera* is grouped together with the Chinese *A. m. sinixinyuan*. The M lineage was suggested to have split from the O and C lineages in Asia Minor<sup>3</sup>, but our analyses recovered a poorly supported affinity with African subspecies. Thus, the origin of the M lineage is open to further investigation.
7. Future studies of honey bee origins should prioritise obtaining samples from northern Africa and the Middle East, as these regions are home to a high diversity of genetically diverse subspecies<sup>46</sup>. These samples will be especially important for testing the monophyly of the A lineage and the positions of the recently proposed S and Y lineages.

## Data availability

The NCBI accession numbers of the sequences analysed are listed in the Supplementary Information. The analysed datasets and all output files are available online at Mendeley Data <https://data.mendeley.com/datasets/fk9whcw3pp/1> (<https://doi.org/10.17632/fk9whcw3pp.1>).

Received: 25 March 2020; Accepted: 11 August 2020

Published online: 03 September 2020

## References

1. Han, F., Wallberg, A. & Webster, M. T. From where did the Western honeybee (*Apis mellifera*) originate?. *Ecol. Evol.* **2**, 1949–1957 (2012).
2. Whitfield, C. W. *et al.* Thrice out of Africa: Ancient and recent expansions of the honey bee, *Apis mellifera*. *Science* **314**, 642–645 (2006).
3. Chen, C. *et al.* Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinixinyuan* n. ssp. *Mol. Biol. Evol.* **33**, 1337–1348 (2016).
4. Engel, M. S. The taxonomy of recent and fossil honey bees (Hymenoptera: Apidae; *Apis*). *J. Hymenoptera Res.* **8**, 165–196 (1999).
5. Meixner, M. D. *et al.* Standard methods for characterising subspecies and ecotypes of *Apis mellifera*. *J. Apic. Res.* **52**, 1–28 (2013).
6. Ruttner, F. *Biogeography and Taxonomy of Honeybees* (Springer, Berlin, 1988).
7. Yunusbaev, U. B. *et al.* The role of whole-genome studies in the investigation of honey bee biology. *Russ. J. Genet.* **55**, 815–824 (2019).
8. Bustamante, T., Baiser, B. & Ellis, J. D. Comparing classical and geometric morphometric methods to discriminate between the South African honey bee subspecies *Apis mellifera scutellata* and *Apis mellifera capensis* (Hymenoptera: Apidae). *Apidologie* <https://doi.org/10.1007/s13592-019-00651-6> (2019).
9. Čermák, K. & Kašpar, F. A method of classifying honey bee races by their body characters. *Pszczelnicze Zeszyty Naukowe* **44**, 81–86 (2000).
10. Bouga, M. *et al.* A review of methods for discrimination of honey bee populations as applied to European beekeeping. *J. Apic. Res.* **50**, 51–84 (2011).
11. Oleksa, A. & Tofilski, A. Wing geometric morphometrics and microsatellite analysis provide similar discrimination of honey bee subspecies. *Apidologie* **46**, 49–60 (2015).
12. Adam, B. *In Search of the Best Strains of Bees*. (Walmar Verlag Zell, 1968).
13. De la Rúa, P., Fuchs, S. & Serrano, J. Biogeography of European honey bees. in *Beekeeping and Conserving Biodiversity of Honey Bees* 15–52 (Northern Bee Books, 2005).
14. Hepburn, H. R. & Crewe, R. M. Portrait of the Cape honeybee, *Apis mellifera capensis*. *Apidologie* **22**, 567–580 (1991).
15. Uzunov, A. *et al.* Swarming, defensive and hygienic behaviour in honey bee colonies of different genetic origin in a pan-European experiment. *J. Apic. Res.* **53**, 248–260 (2014).
16. Büchler, R. *et al.* The influence of genetic origin and its interaction with environmental effects on the survival of *Apis mellifera* L. colonies in Europe. *J. Apic. Res.* **53**, 205–214 (2014).
17. Francis, R. M. *et al.* Effect of genotype and environment on parasite and pathogen levels in one apiary—a case study. *J. Apic. Res.* **53**, 230–232 (2014).
18. Requier, F. *et al.* The conservation of native honey bees is crucial. *Trends Ecol. Evol.* **34**, 789–798 (2019).
19. Altaye, S. Z., Meng, L. & Li, J. Molecular insights into the enhanced performance of royal jelly secretion by a stock of honeybee (*Apis mellifera ligustica*) selected for increasing royal jelly production. *Apidologie* **50**, 436–453 (2019).
20. Dogantzis, K. A. & Zayed, A. Recent advances in population and quantitative genomics of honey bees. *Curr. Opinion Insect Sci.* **31**, 93–98 (2019).
21. Ruttner, F., Tassencourt, L. & Louveaux, J. Biometrical-statistical analysis of the geographical variability of *Apis mellifera*. *Apidologie* **9**, 363–381 (1978).
22. Wilson, E. O. *The Insect Societies* (Belknap Press, London, 1971).
23. Cridland, J. M., Tsutsui, N. D. & Ramírez, S. R. The complex demographic history and evolutionary origin of the Western honey bee, *Apis mellifera*. *Genome Biol. Evol.* **9**, 457–472 (2017).
24. Boardman, L. *et al.* The complete mitochondrial genome of *Apis mellifera unicolor* (Insecta: Hymenoptera: Apidae), the Malagasy honey bee. *Mitochondrial DNA Part B* **4**, 3286–3287 (2019).
25. Boardman, L. *et al.* The mitochondrial genome of the Carniolan honey bee, *Apis mellifera carnica* (Insecta: Hymenoptera: Apidae). *Mitochondrial DNA Part B* **4**, 3288–3290 (2019).
26. Boardman, L. *et al.* The mitochondrial genome of *Apis mellifera simensis* (Hymenoptera: Apidae), an Ethiopian honey bee. *Mitochondrial DNA Part B* **5**, 9–10 (2020).
27. Boardman, L. *et al.* The complete mitochondrial genome of *Apis mellifera jemenitica* (Insecta: Hymenoptera: Apidae), the Arabian honey bee. *Mitochondrial DNA Part B* **5**, 875–876 (2020).
28. Boardman, L. *et al.* The mitochondrial genome of the Maltese honey bee, *Apis mellifera ruttneri* (Insecta: Hymenoptera: Apidae). *Mitochondrial DNA Part B* **5**, 877–878 (2020).
29. Boardman, L. *et al.* The mitochondrial genome of the Spanish honey bee, *Apis mellifera iberiensis* (Insecta: Hymenoptera: Apidae), from Portugal. *Mitochondrial DNA Part B* **5**, 17–18 (2020).

30. Boardman, L. *et al.* The complete mitochondrial genome of the West African honey bee *Apis mellifera adansonii* (Insecta: Hymenoptera: Apidae). *Mitochondrial DNA Part B* **5**, 11–12 (2020).
31. Engel, M. S. Fossil honey bees and evolution in the genus *Apis* (Hymenoptera: Apidae). *Apidologie* **29**, 265–281 (1998).
32. Kotthoff, U., Wappler, T. & Engel, M. S. Miocene honey bees from the Randeck Maar of southwestern Germany (Hymenoptera, Apidae). *Zookeys* <https://doi.org/10.3897/zookeys.96.752> (2011).
33. Kotthoff, U., Wappler, T. & Engel, M. S. Greater past disparity and diversity hints at ancient migrations of European honey bee lineages into Africa and Asia. *J. Biogeogr.* **40**, 1832–1838 (2013).
34. Nel, A., Martínez-Delclòs, X., Arillo, A. & Peñalver, E. A review of the Eurasian fossil species of the bee *Apis*. *Palaeontol.* **42**, 243–285 (1999).
35. Culliney, T. W. Origin and evolutionary history of the honeybees *Apis*. *Bee World* **64**, 29–38 (1983).
36. Deodikar, G. B., Thakar, C. V. & Shah, P. N. Cyto-genetic studies in Indian honey-bees. *Proc. Indian Acad. Sci. Sect. B* **49**, 194–206 (1959).
37. Deodikar, G. B., Thakar, C. V. & Tonapi, K. V. Evolution in the genus *Apis*. *Indian Bee J.* **23**, 86–91 (1961).
38. Cornuet, J. M. & Garnery, L. Mitochondrial DNA variability in honeybees and its phylogeographic implications. *Apidologie* **22**, 627–642 (1991).
39. Wallberg, A. *et al.* A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nat. Genet.* **46**, 1081–1088 (2014).
40. Brinkmann, H., van der Giezen, M., Zhou, Y., de Raucourt, G. P. & Philippe, H. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. *Syst. Biol.* **54**, 743–757 (2005).
41. Franck, P., Garnery, L., Celebrano, G., Solignac, M. & Cornuet, J.-M. Hybrid origins of honeybees from Italy (*Apis mellifera ligustica*) and Sicily (*A. m. sicula*). *Mol. Ecol.* **9**, 907–921 (2000).
42. Franck, P., Garnery, L., Solignac, M. & Cornuet, J.-M. Molecular confirmation of a fourth lineage in honeybees from the Near East. *Apidologie* **31**, 167–180 (2000).
43. Ruttner, F. *Naturgeschichte der Honigbienen* (Ehrenwirth Verlag, Ehrenwirth, 1992).
44. Franck, P. *et al.* Genetic diversity of the honeybee in Africa: microsatellite and mitochondrial data. *Heredity* **86**, 420–430 (2001).
45. Alburaki, M., Moulin, S., Legout, H., Alburaki, A. & Garnery, L. Mitochondrial structure of Eastern honeybee populations from Syria, Lebanon and Iraq. *Apidologie* **42**, 628 (2011).
46. Alburaki, M. *et al.* A fifth major genetic group among honeybees revealed in Syria. *BMC Genet.* **14**, 117 (2013).
47. Franck, P., Garnery, L., Solignac, M. & Cornuet, J.-M. The origin of West European subspecies of honeybees (*Apis mellifera*): New insights from microsatellite and mitochondrial data. *Evolution* **52**, 1119–1134 (1998).
48. Crozier, R. H. & Crozier, Y. C. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* **133**, 97–117 (1993).
49. Zhang, D. *et al.* PhyloSuite: An integrated and scalable desktop platform for streamlined molecular sequence data management and evolutionary phylogenetics studies. *Mol. Ecol. Res.* **20**, 348–355 (2020).
50. Katoh, K. & Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
51. Rota-Stabelli, O., Lartillot, N., Philippe, H. & Pisani, D. Serine codon-usage bias in deep phylogenomics: Pancrustacean relationships as a case study. *Syst. Biol.* **62**, 121–133 (2013).
52. Breinholt, J. W. & Kawahara, A. Y. Phylotranscriptomics: Saturated third codon positions radically influence the estimation of trees based on next-gen data. *Genome Biol. Evol.* **5**, 2082–2092 (2013).
53. Philippe, H. *et al.* Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biol.* **9**, e1000602 (2011).
54. Lartillot, N., Brinkmann, H. & Philippe, H. Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. *BMC Evol. Biol.* **7**, S4 (2007).
55. Williams, T. A., Cox, C. J., Foster, P. G., Szöllösi, G. J. & Embley, T. M. Phylogenomics provides robust support for a two-domains tree of life. *Nat. Ecol. Evol.* **4**, 138–147 (2020).
56. Bleidorn, C. *Phylogenomics: An Introduction* (Springer International Publishing, Berlin, 2017). <https://doi.org/10.1007/978-3-319-54064-1>.
57. Simion, P. *et al.* A large and consistent phylogenomic dataset supports sponges as the sister Group to all other animals. *Curr. Biol.* **27**, 958–967 (2017).
58. Feuda, R. *et al.* Improved modeling of compositional heterogeneity supports sponges as sister to all other animals. *Curr. Biol.* **27**, 3864–3870.e4 (2017).
59. Pisani, D. *et al.* Genomic data do not support comb jellies as the sister group to all other animals. *Proc. Natl. Acad. Sci.* **112**, 15402–15407 (2015).
60. Betts, H. C. *et al.* Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. *Nat. Ecol. Evol.* **2**, 1556–1562 (2018).
61. Lozano-Fernandez, J. *et al.* Pancrustacean evolution illuminated by taxon-rich genomic-scale data sets with an expanded remi-pede sampling. *Genome Biol. Evol.* **11**, 2055–2070 (2019).
62. Lozano-Fernandez, J. *et al.* Increasing species sampling in chelicerate genomic-scale datasets provides support for monophyly of Acari and Arachnida. *Nat. Commun.* **10**, 1–8 (2019).
63. Cai, C., Tihelka, E., Pisani, D. & Donoghue, P. C. J. Data curation and modeling of compositional heterogeneity in insect phylogenomics: a case study of the phylogeny of Dytiscoidea (Coleoptera: Adepaga). *Mol. Phylogenet. Evol.* **147**, 106782 (2020).
64. Lartillot, N., Lepage, T. & Blanquart, S. PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinform.* **25**, 2286–2288 (2009).
65. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. & Calcott, B. PartitionFinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **34**, 772–773 (2017).
66. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274 (2015).
67. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v4: Recent updates and new developments. *Nucl. Acids Res.* **47**, W256–W259 (2019).
68. Lartillot, N., Rodrigue, N., Stubbs, D. & Richer, J. PhyloBayes MPI: Phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst. Biol.* **62**, 611–615 (2013).
69. Zaitoun, S. T., Al-Ghzawi, A.-M. & Shannag, H. K. Population dynamics of the Syrian Honeybee, *Apis mellifera syriaca*, under semi-arid Mediterranean conditions. *Zool. Middle East* **21**, 129–132 (2000).
70. Haddad, N. J. *et al.* Draft genome sequence of the Algerian bee *Apis mellifera intermissa*. *Genom. Data* **4**, 24–25 (2015).
71. Arias, M. C. & Sheppard, W. S. Phylogenetic relationships of honey bees (Hymenoptera: Apinae: Apini) inferred from nuclear and mitochondrial DNA sequence data. *Mol. Phylogenet. Evol.* **37**, 25–35 (2005).
72. Sheppard, W. S. & Meixner, M. D. *Apis mellifera pomonella*, a new honey bee subspecies from Central Asia. *Apidologie* **34**, 367–375 (2003).
73. Macholán, M. *Základy fylogenetické analýzy* (Masaryk University, Masaryk, 2014).
74. Garnery, L., Cornuet, J.-M. & Solignac, M. Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis. *Mol. Ecol.* **1**, 145–154 (1992).

75. Arias, M. C. & Sheppard, W. S. Molecular phylogenetics of honey bee subspecies (*Apis mellifera* L.) inferred from mitochondrial DNA sequence. *Mol. Phylogenet. Evol.* **5**, 557–566 (1996).
76. Radloff, S. E. & Hepburn, H. R. Multivariate analysis of honeybee populations, *Apis mellifera* L. (Hymenoptera: Apidae), from western central Africa: morphometrics and pheromones. *Afr. Entomol.* **5**, 195–204 (1997).
77. Amssalu, B., Nuru, A., Radloff, S. E. & Hepburn, H. R. Multivariate morphometric analysis of honeybees (*Apis mellifera*) in the Ethiopian region. *Apidologie* **35**, 71–81 (2004).
78. Meixner, M. D., Leta, M. A., Koeniger, N. & Fuchs, S. The honey bees of Ethiopia represent a new subspecies of *Apis mellifera*—*Apis mellifera simensis* n. ssp. *Apidologie* **42**, 425–437 (2011).
79. Cornuet, J. M. & Fresnaye, J. Etude biométrique de colonies d'abeilles d'Espagne et du Portugal. *Apidologie* **20**, 93–101 (1989).
80. Arias, M. C., Rinderer, T. E. & Sheppard, W. S. Further characterization of honey bees from the Iberian Peninsula by allozyme, morphometric and mtDNA haplotype analyses. *J. Apic. Res.* **45**, 188–196 (2006).
81. Cánovas, F., de la Rúa, P., Serrano, J. & Galián, J. Geographical patterns of mitochondrial DNA variation in *Apis mellifera iberiensis* (Hymenoptera: Apidae). *J. Zool. Syst. Evol. Res.* **46**, 24–30 (2008).
82. Chávez-Galarza, J. *et al.* Mitochondrial DNA variation of *Apis mellifera iberiensis*: further insights from a large-scale study using sequence data of the tRNA<sub>Leu-cox2</sub> intergenic region. *Apidologie* **48**, 533–544 (2017).
83. Huelsenbeck, J. P. Performance of phylogenetic methods in simulation. *Syst. Biol.* **44**, 17–48 (1995).
84. Felsenstein, J. *Inferring Phylogenies* (Sinauer Associates, Sunderland, 2004).
85. Kandemir, I., Kence, M., Sheppard, W. S. & Kence, A. Mitochondrial DNA variation in honey bee (*Apis mellifera* L.) populations from Turkey. *J. Apic. Res.* **45**, 33–38 (2006).
86. Donoghue, P. C. & Benton, M. J. Rocks and clocks: calibrating the Tree of Life using fossils and molecules. *Trends Ecol. Evol.* **22**, 424–431 (2007).
87. Foord, A. S. Notes on a collection of East Coast amber belonging to Mrs Burwood of Yarmouth. *Trans. Norfolk Norwich Nat. Soc.* **5**, 92–95 (1890).
88. Cockerell, T. D. A. Some European fossil bees. *Entomol.* **42**, 313–317 (1909).
89. Zeuner, F. E. & Manning, F. J. A monograph on fossil bees (Hymenoptera: Apoidea). *Bull. Brit. Mus. Nat. Hist. Geol.* **27**, 151–268 (1976).
90. Stroiński, A. & Szwedo, J. *Yuripopoverus africanus* gen. et sp. n from East African copal (Hemiptera: Fulgoromorpha: Ricaniidae). *Pol. J. Entomol.* **80**, 679–688 (2011).
91. Ruttner, F., Milner, E. & Dews, J. E. *The Dark European Honey Bee*, *Apis mellifera mellifera* Linnaeus 1758. (British Isles Bee Breeders Association, 1990).
92. Bloch, G. *et al.* Industrial apiculture in the Jordan valley during Biblical times with Anatolian honeybees. *Proc. Natl. Acad. Sci.* **107**, 11240–11244 (2010).
93. Sheppard, W. S. & Berlocher, S. H. Allozyme variation and differentiation among four *Apis* species. *Apidologie* **20**, 419–431 (1989).
94. Ramírez, S. R. *et al.* A molecular phylogeny of the stingless bee genus *Melipona* (Hymenoptera: Apidae). *Mol. Phylogenet. Evol.* **56**, 519–525 (2010).
95. Felsenstein, J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**, 783–791 (1985).
96. Young, A. D. & Gillung, J. P. Phylogenomics—principles, opportunities and pitfalls of big-data phylogenetics. *Syst. Entomol.* **n/a**, (2019).
97. Eme, L., Spang, A., Lombard, J., Stairs, C. W. & Ettema, T. J. G. Archaea and the origin of eukaryotes. *Nat. Rev. Microbiol.* **15**, 711–723 (2017).
98. Crewe, R. M., Hepburn, H. R. & Moritz, R. F. A. Morphometric analysis of two southern African races of honeybee. *Apidologie* **25**, 61–70 (1994).
99. Harpur, B. A., Minaei, S., Kent, C. F. & Zayed, A. Management increases genetic diversity of honey bees via admixture. *Mol. Ecol.* **21**, 4414–4421 (2012).
100. Harpur, B. A., Minaei, S., Kent, C. F. & Zayed, A. Admixture increases diversity in managed honey bees: Reply to De la Rúa *et al.* *Mol. Ecol.* **22**, 3211–3215 (2013).
101. De la Rúa, P., Jaffé, R., Dall'olio, R., Muñoz, I. & Serrano, J. Biodiversity, conservation and current threats to European honeybees. *Apidologie* **40**, 263–284 (2009).
102. de la Rúa, P. *et al.* Conserving genetic diversity in the honeybee: Comments on Harpur *et al.* (2012). *Mol. Ecol.* **22**, 3208–3210 (2013).
103. Harpur, B. A. *et al.* Assessing patterns of admixture and ancestry in Canadian honey bees. *Insect. Soc.* **62**, 479–489 (2015).
104. Chapman, N. C. *et al.* Hybrid origins of Australian honeybees (*Apis mellifera*). *Apidologie* **47**, 26–34 (2016).

## Acknowledgements

E.T. is grateful to Mattia Giacomelli (University of Bristol, UK) for advice regarding the bioinformatic aspects of the study. This work was supported by the Strategic Priority Research Program (B), Grant number XDB26000000; Natural Science Foundation of China, Grant numbers 41688103, 31402008, 41602009; Youth Innovation Promotion Association of the CAS Grant number 2018347; and the Newton International Fellowship from the Royal Society awarded to C.C.

## Author contributions

E.T. and C.C. designed the study, E.T. conducted the analyses and drafted the manuscript, all authors contributed to result interpretation and writing of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-71393-0>.

**Correspondence** and requests for materials should be addressed to E.T. or C.C.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020