

Article

Open Access

Defining honeybee subspecies in an evolutionary context warrants strategized conservation

Lifei Qiu¹, Jiangxing Dong¹, Xingan Li², Sajad H. Parey³, Ken Tan⁴, Michael Orr⁵, Aquib Majeed³, Xue Zhang¹, Shiqi Luo¹, Xuguo Zhou⁶, Chaodong Zhu⁵, Ting Ji⁷, Qingsheng Niu², Shanlin Liu^{1,*}, Xin Zhou^{1,*}

¹ Department of Entomology, College of Plant Protection, China Agricultural University, Beijing 100193, China

² Key Laboratory for Bee Genetics and Breeding, Jilin Provincial Institute of Apicultural Sciences, Jilin, Jilin 132108, China

³ Department of Zoology, School of Biosciences and Biotechnology, Baba Ghulam Shah Badshah University, Rajouri (Jammu and Kashmir) 185234, India

⁴ CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Xishuangbanna, Yunnan 650000, China

⁵ Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

⁶ Department of Entomology, University of Kentucky, Lexington, KY 40546, USA

⁷ Yangzhou University, Yangzhou, Jiangsu 225009, China

ABSTRACT

Despite the urgent need for conservation consideration, strategic action plans for the preservation of the Asian honeybee, *Apis cerana* Fabricius, 1793, remain lacking. Both the convergent and divergent adaptations of this widespread insect have led to confusing phenotypical traits and inconsistent infraspecific taxonomy. Unclear subspecies boundaries pose a significant challenge to honeybee conservation efforts, as it is difficult to effectively prioritize conservation targets without a clear understanding of subspecies identities. Here, we investigated genome variations in 362 worker bees representing almost all populations of mainland *A. cerana* to understand how evolution has shaped its population structure. Whole-genome single nucleotide polymorphisms (SNPs) based on nuclear sequences revealed eight putative subspecies, with all seven peripheral subspecies exhibiting mutually exclusive monophyly and distinct genetic divergence from the widespread central subspecies. Our results demonstrated that most classic morphological traits, including body size, were related to the climatic variables of the local habitats and did not reflect the true evolutionary history of the organism. Thus, such morphological traits were not suitable for subspecific delineation. Conversely, wing vein characters showed relative independence to the environment and supported the subspecies boundaries inferred from nuclear genomes.

This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright ©2023 Editorial Office of Zoological Research, Kunming Institute of Zoology, Chinese Academy of Sciences

Mitochondrial phylogeny further indicated that the present subspecies structure was a result of multiple waves of population divergence from a common ancestor. Based on our findings, we propose that criteria for subspecies delineation should be based on evolutionary independence, trait distinction, and geographic isolation. We formally defined and described eight subspecies of mainland *A. cerana*. Elucidation of the evolutionary history and subspecies boundaries enables a customized conservation strategy for both widespread and endemic honeybee conservation units, guiding colony introduction and breeding.

Keywords: *Apis cerana*; Integrative taxonomy; Species concept; Pollinator insect; Centrifugal diversification; Morphology; Genomics

INTRODUCTION

The Asian honeybee (*Apis cerana*) is an important pollinator

Received: 14 January 2023; Accepted: 13 March 2023; Online: 14 March 2023

Foundation items: The work was supported by the National Natural Science Foundation (NSF) of China (32270475), Program of Ministry of Science and Technology of China (2018FY100403), National Special Support Program for High-level Talents (Ten-Thousand Talents Program), and 2115 Talent Development Program of China Agricultural University through Xin Z. S.L. is supported by Funds for International Cooperation and Exchange of the National Natural Science Foundation of China (3211001043). Sample collection was also supported by the NSF of China (31470123) and Jilin Science and Technology Program (20030561) through X.L. S.H.P. is supported by the National Mission on Himalayan Studies (NMHS) - Almora, Ministry of Environment, Forest and Climate Change, Government of India, through grant GBPN/NMHS-2017-18/MG-12

*Corresponding authors, E-mail: shanlin.liu@cau.edu.cn; xinzhou@cau.edu.cn

native to much of Asia (Figure 1A), providing crucial pollination services for agricultural production, forestry, biodiversity, and natural resource maintenance, supporting more than half of the human population (Abrol, 2013). However, the species faces numerous threats, including the application of pesticides, floral resource shortages, and habitat destruction and fragmentation (Abrol, 2013). In addition to these common challenges shared by most pollinators, the early introduction and extensive importation of the western honeybee (*A. mellifera* Linnaeus 1758) has resulted in direct colony replacements by beekeepers across a large range of the native bees' habitat and introduced new pathogens such as the sacbrood virus (SBV) (Vung et al., 2018) and *A. mellifera* filamentous virus (AmFV) (Hou et al., 2016). Consequently, *A. cerana* populations have experienced a significant reduction in both distribution and size over the past century (Yang, 2009).

Although the need for immediate conservation action for this species is widely acknowledged, strategic plans are rarely discussed. Prioritizing conservation efforts for species with

restricted distribution (including most endangered species) may not present a challenge, but the same cannot be said for taxa with large geographic ranges (Phillimore & Owens, 2006). Notably, regarding native bee populations, cross-population breeding and bee farming practices are obvious issues (Wang et al., 2021). In China, the importation of exotic colonies from remote provinces to local villages to promote regional economies (Wang et al., 2021; Zhao, 2018) presents significant difficulties for assessing the potential impacts of human interference on local honeybees, given the lack of understanding of genealogy and fitness variations in different environments. Conservation actions require a clear definition of target species or subspecies and the differentiation of evolutionary factors from contemporary human disturbance. Although both may influence populations, the latter can be immediately ameliorated by implementing proper actions (Gutiérrez & Helgen, 2013; Morrison III et al., 2009; Requier et al., 2019).

The establishment of proper conservation units is an

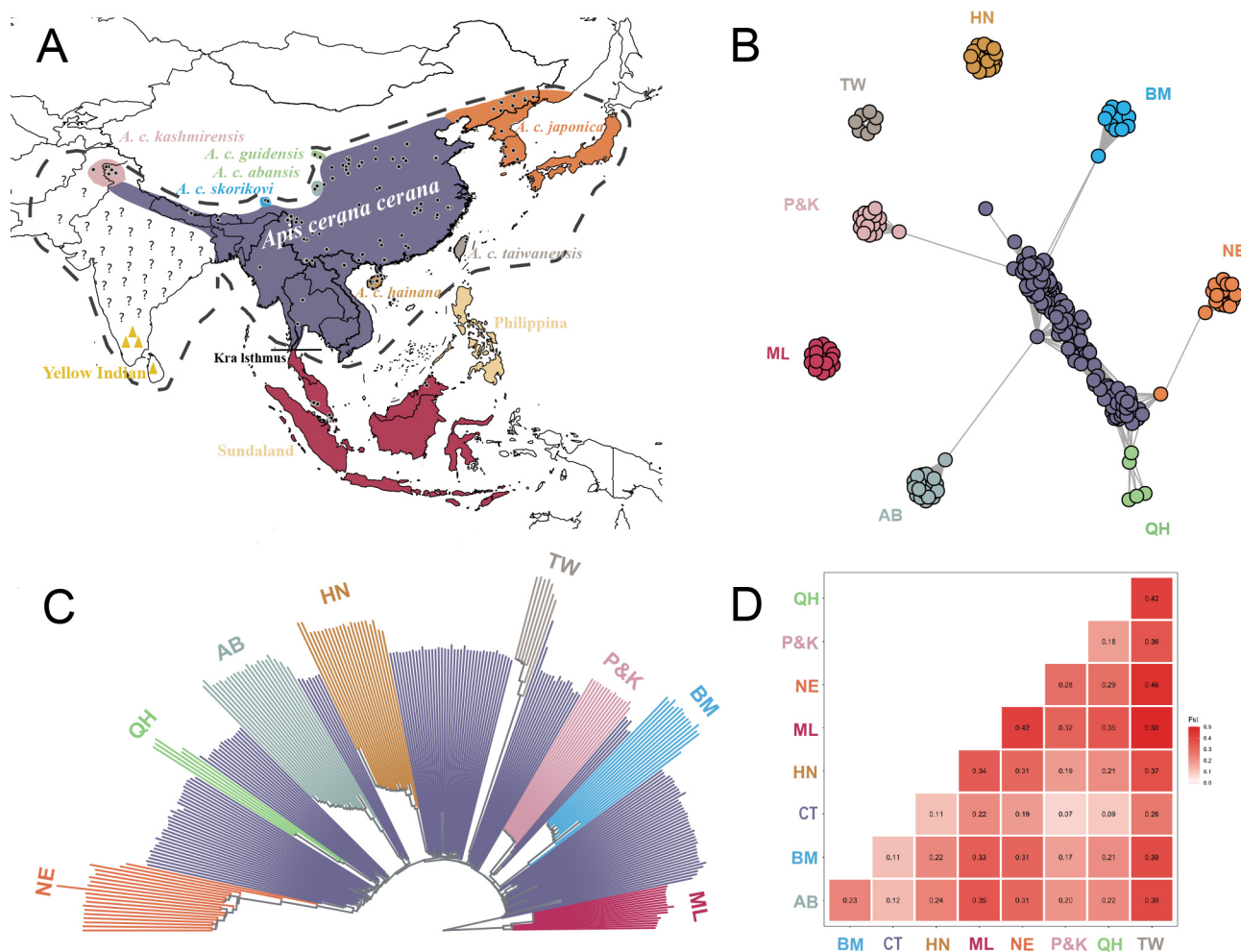


Figure 1 Population structure, phylogenetic relationships, and genetic divergence of mainland Asian honeybee *A. cerana*

A: Native ranges of eight subspecies constituting the mainland lineage of *A. cerana* (indicated by dash lines), and Sundaland, Palawan, and Luzon-Mindanao lineages. Note, cavity-nesting bees of India (yellow) and Philippines (beige) have recently been recognized as outgroups to the common ancestor of Sundaland and mainland lineages (Lo et al., 2010). Dots represent sampling sites. B: Population structure based on nuclear SNPs and NetView. C: Phylogenetic tree based on nuclear SNPs, showing phylogenetic monophyly of seven peripheral descendant subspecies and Malaysian out-group. D: F_{ST} matrix constructed from estimates of pairwise F_{ST} between all eight subspecies of mainland Asian honeybee and Malaysian out-group. Codes for subspecies and geographic populations are: *A. c. cerana* (central group, CT), *A. c. japonica* (Northeast group, NE), *A. c. guidensis* (Qinghai group, QH), *A. c. abansis* (Aba group, AB), *A. c. skorikovi* (Bomi group, BM), *A. c. kashmirensis* (Pakistan and Kashmir group, K&P), *A. c. hainana* (Hainan group, HN), *A. c. taiwanensis* (Taiwan group, TW), and Malaysian out-group (ML).

essential prerequisite for successful conservation. However, the infraspecific taxonomy of *A. cerana* remains enigmatic. Over the past two centuries, approximately 20 equivocal “subspecies” (Chen, 1993; Engel, 1999; Ilyasov et al., 2018, 2019; Lo et al., 2010; Ma & Huang, 1981; Maa, 1953; Radoszkowski, 1887; Smith & Hagen, 1996; Yang et al., 1986; Zhuang, 1989), 31 putative biometric groups (Hepburn et al., 2001b), and various ecotypes (National Animal Genetic Resources Committee, 2011) and morphoclusters (Radloff et al., 2010) have been proposed. In particular, taxonomists hold divergent views at the subspecies level, ranging from the recognition of eight subspecies (Engel, 1999) to the identification of only six morphoclusters across Asia (Radloff et al., 2010). Recently, based on whole-genome single nucleotide polymorphisms (SNPs) of representative populations, we revealed the presence of seven genetically distinctive groups within mainland *A. cerana* (the only group distributed across tropical and temperate regions), showing genomic divergence at the subspecies level (Ji et al., 2020). The reasons for this long-lasting debate are multifaceted (Engel, 1999; Hepburn et al., 2001b; Peng et al., 1989; Radloff et al., 2010), but include discrepancies in sampling strategy, conflicting results produced by varied characteristics (e.g., morphology and genetics), and innate challenges in achieving a uniform subspecies concept (De Queiroz, 2007, 2020).

Geographic coverage of population sampling can impact the representativeness of genealogy and the subsequent interpretation of results. Broad sampling of a wide-ranging species is inevitably challenging, and key lineages many remain unsampled. For instance, in various studies conducted in the Far East, the inferences of *A. cerana* colonization routes between mainland China and Japan varied with the inclusion of new samples (Ilyasov et al., 2018, 2019). Incomplete sampling seems to have at least partially contributed to previous designations of invalid subspecies delineations. For example, three *A. cerana* subspecies and one endemic type identified in local studies in the Far East are now hypothesized to belong to the same subspecies (Ilyasov et al., 2018, 2019; Ji et al., 2020; Liu et al., 2022; National Animal Genetic Resources Committee, 2011).

The continued emergence of new characters and evolution of new methodologies have added yet another level of complexity to subspecies delineation. In particular, the application of novel molecular techniques has cast doubt upon prior studies. For example, recent genomic analyses have largely redefined the distribution of the Aba race of *A. cerana* in montane regions (Chen et al., 2018a; Ji et al., 2020). Once thought to span a large transition zone from the Sichuan basin to the Qinghai-Xizang (Tibet) Plateau (QTP) (central and northern Sichuan, eastern Qinghai, and southern Gansu) (Peng et al., 1989; National Animal Genetic Resources Committee, 2011; Yang et al., 1995), it is now thought to be restricted to the upper reaches of the Dadu River. This substantial difference in range could strongly influence the efficacy of conservation strategies, as an action plan based on the larger hypothesized distribution range may result in no management efforts in the smaller area where these honeybees actually inhabit. However, the lack of criteria for judging which set of characters (e.g., morphology vs. genetics) is more reliable remains a dilemma.

The inconsistent application of subspecies definitions also represents a substantial conceptual challenge. Understanding species boundaries is essential for answering questions

related to evolution, adaptation, conservation, and ecology (Christmas et al., 2021; De Queiroz, 2007; Gutiérrez & Helgen, 2013; Hillis, 2019; Morrison III et al., 2009). However, subspecies are relatively less regulated and are not explicitly covered by the International Code of Zoological Nomenclature. Despite considerable species definitions (Groves et al., 2017), multiple lines of evidences have been applied to infer evolutionary or reproductive isolation and subsequent completion of genetic sorting between sister taxon pairs that share a common ancestor. However, a central challenge is the need to determine isolation, which is difficult to implement for all but the best-known species (Malinsky et al., 2015). Because subspecies reflect an intermediate evolutionary phase of a focal species, they are transient units on the verge of diverging into new incipient species, or alternatively, being replaced by other subspecies, younger descendants, or their hybrids (De Queiroz, 2020). Subsequently, given the lower levels of divergence expected among subspecies, the integration of multiple forms of data (e.g., molecular, morphology) is expected to produce the best results (Guillot et al., 2012).

In accordance with the evolutionary model established in our previous work (Ji et al., 2020), we herein formally describe eight mainland *A. cerana* subspecies. Compared with Ji et al. (2020), we focused on taxonomic delineations and descriptions of the proposed subspecies, with the designation of type specimens for each newly described subspecies and for all previously defined subspecies lacking types. Our streamlined pipeline integrated genetic divergence, phylogenetic relationships, and morphology to test the hypothesis that genetically and morphologically distinct population groups revealed within mainland *A. cerana* are valid subspecies. We demonstrated that multiple independent divergence events may have shaped the current population structure, resulting in a total of eight subspecies, three of which are described for the first time. In addition to clarifying conservation units, we also explored the unique threats faced by these subspecies, highlighting their evolutionary histories and genetic features, and providing important information for strategizing and enacting novel conservation priorities.

MATERIALS AND METHODS

Genome sequencing

Whole-genome sequences of 306 worker bee samples were obtained from our previous study (Ji et al., 2020, NCBI archive PRJNA592293). Additionally, 56 worker bees collected from Pakistan, Kashmir, Myanmar, Malaysia, and China (Supplementary Table S1) were newly sequenced in this study. The combined data covered all eight provisional subspecies of the mainland Asian honeybee (with *A. cerana indica* recently recognized as an outgroup to the common ancestor of the Sundaland and mainland lineages) (Lo et al., 2010). The 56 new samples were collected on flowers, and the head and thorax were then processed for DNA extraction using phenol chloroform. Genomic DNA was sequenced on the BGISEQ-500 platform at the BGI, China (3.5 to 5 Gb per sample). The newly sequenced data were deposited in the Sequence Read Archive database (PRJNA870246).

We applied Fastp v0.13.1 (Chen et al., 2018b) to remove low-quality reads with the parameters “-q 10 -n 10 -u 40.” The clean reads of each sample were mapped to the genome assembly of *A. cerana* (Ref ID: ACSNU-2.0, GCF_001442555.1) using BWA ALN (Li & Durbin, 2009) with

polymerase chain reaction (PCR) duplications being masked using SamBamba (Tarasov et al., 2015). We then applied the Genome Analysis Toolkit v4.0.4 (McKenna et al., 2010) for short variant identification and retained biallelic SNP sites that met the following criteria: (1) average depth $>1/3$ and $<2\times$ mean depth of the whole dataset; (2) quality score >20 ; (3) genotype quality >30 ; (4) minor allele frequency >0.01 ; (5) proportion of missing genotypes $<50\%$. Genome heterozygosity and allelic missing rate were estimated for each individual using VCFtools (Danecek et al., 2011).

Population structure and phylogeny

We inferred kinship relationships between individual pairs using VCFtools (Danecek et al., 2011) following Manichaikul et al. (2010) and only retained unrelated individuals for downstream analyses. Individuals with $>10\%$ missing data were also removed. Finally, a total of 273 worker bees (75.6%) were retained for subsequent analyses (Supplementary Table S1). We applied a balanced minimum evolution method implemented in Fastme (Lefort et al., 2015) for phylogenetic tree construction based on a whole-genome identical-by-state (IBS) matrix calculated using Plink v1.9 (Purcell et al., 2007). Pairwise phylogenetic distances (PD) between different individuals were calculated using an in-house Perl script. The IBS matrix was also used for genetic structure estimation with a mutual k -nearest neighbor graph method implemented in NetView (Neuditschko et al., 2012).

We performed admixture analyses of bee populations from Hainan (HN), Bomi (BM), Northeast (NE), Qinghai (QH), Taiwan (TW), and Pakistan & Kashmir (KP). For each bee population, we created a dataset that respectively included: (1) all individuals from the focal locality; (2) individuals from Malaysia as an outgroup; (3) individuals from adjacent regions that belong to the central population according to Ji et al. (2020) (Supplementary Table S1).

We estimated the admixture level for each group using NGSAdmix (Skotte et al., 2013) based on the genotype likelihood calculated using ANGSD (Korneliussen et al., 2014) with the following parameters: base quality ≥ 20 , mapping quality ≥ 20 , minor allele frequency (MAF) ≥ 0.05 , SNP P -value $\geq 1e-6$, and informative individuals ≥ 10 . For admixture analysis, we applied a predefined K value of 3 and obtained a convergent estimation when the three highest likelihood values fell within the range of five likelihood units.

Mitochondrial phylogeny

From the genome sequences of each individual honeybee, we obtained a consensus sequence for the mitochondrial genome using the doFasta function imbedded in ANGSD (Korneliussen et al., 2014), requiring a depth $\geq 5X$. We then obtained all 13 protein coding genes (PCGs) for each sample based on the reference annotation information (NC_014295.1). In addition, the mitochondrial genome of *A. nigrocincta* (NC_038114.1) was applied to root the tree. After sequence alignment of the 13 PSGs using MAFFT (Nakamura et al., 2018), we concatenated the alignments to generate a “supergene” and inferred a phylogenetic tree using IQTREE (Minh et al., 2020) with the best-fit substitution models estimated by ModelFinder (Kalyaanamoorthy et al., 2017). To present a simplified view for better illustration, we included only a single representative for each unique haplotype from the same sampling location.

Morphological analyses

The honeybee worker specimens used for morphological

analyses were obtained from two sources: (1) A total of 119 specimens collected from seven peripheral provisional subspecies population throughout China, India, and Japan from 2010 to 2021. (2) A total of 207 colonies from 101 localities, covering much of the distribution of the central population, recorded in the Kunming Honeybee Database (Tan et al., 2006). The geographical origins of all samples representing 234 colonies from 115 localities are listed in Supplementary Table S2.

Morphological measurements

The morphological traits described in Ruttner (1988), as well as two additional effective characters (cubital index and hamuli number), were employed to characterize *A. cerana*. Overall, the 35 morphometric characters included 12 relevant to size, 16 relevant to forewing, five relevant to color, one relevant to hair, and one relevant to hamuli (Tan et al., 2008).

Morphometric analysis

We first assessed differences among different datasets, i.e., legacy data obtained from Tan et al. (2006) and the Morphological Bank and new data obtained from the current study. We then performed independent principal component analysis (PCA) for size-, wing-, and color-related traits to remove traits showing apparent divergences across data sources or outlier samples from each data source. Subsequently, we used the filtered trait set to conduct cluster analysis using samples from Hainan Island and Northeast China/Korea/Japan populations, for which overlapping sampling is found between legacy and new data. We divided the focal populations into eight provisional subspecies, with the central subspecies (i.e., *A. cerana cerana*) further divided into three sub-clusters (Northern cerana, Indo Chinese cerana and Himalayan cerana), based on combined genomic (Ji et al., 2020) and morphological evidence (Radloff et al., 2010). Mean values and standard deviations were computed for each character for all colonies using the morphometric Image Analysis System 11 program. The “princomp” function in R was used for PCA to identify possible morphoclusters. Morphometric similarities were investigated using hierarchical cluster analyses based on location means of Chinese populations and group means of adjacent country populations.

Bioclimatic analysis

We extracted bioclimatic variables for all sampling sites using DIVA-GIS (www.diva-gis.org). We performed PCA using the “princomp” function in R to obtain the first two principal components. Correlations between Dim 1–5 scores of PCA for morphological characteristics and bioclimatic variables, including altitude, latitude, longitude, and Bio1–19, were calculated using the R package “corrplot”. *: $P < 0.05$; **: $P < 0.01$; ***: $P < 0.001$.

Geometric morphometrics

A total of 221 forewings from *A. cerana* worker bees collected from 24 sites (see Supplementary Table S3) were used for geometric morphometric analysis. All wing specimens were photographed using a Nikon D700 camera (Japan). These images were then used in subsequent morphological analyses. A total of 20 forewing landmarks were identified following our previous work (Ji et al., 2020). All landmarks were digitized using tpsDig v2.0 and tpsUtil v1.40 (http://life.bio.sunysb.edu/morph/). We divided the 24 sampling sites into eight groups following our subspecies definitions. Size-corrected shape was used to conduct canonical variable

analysis (CVA) in Morphoj v1.06 to determine the degree of morphological discrimination among subspecies. Mahalanobis distances (MDs) among groups and *P*-values from permutation tests (10 000 iterations) were calculated during CVA. The thin plate spline (TPS) method in Morphoj v1.06 (Klingenberg, 2011) was used to identify morphological differences in wing shape between each pair of subspecies. We estimated morphological divergences between peripheral and central subspecies using MD. We calculated the correlation efficiency between fixation index (F_{ST}) and MD among groups to evaluate the level of coherence between genetic and morphological divergences. All calculations were performed using SPSS v20.0.

RESULTS

Provisional subspecies suggested by nuclear genetic divergence, phylogenetic relationships, and evolutionary history

Phylogenetic relationships based on nuclear genomes (nrDNA) showed that, in congruence with our previous findings (Ji et al., 2020), the six peripheral population groups (i.e., Hainan, Taiwan, Northeast, Bomi, Aba, and Qinghai) and the out-group (Malaysia) (Figure 1A) were genetically distinct from the central group, each forming a monophyletic lineage (Figure 1B). In addition, the newly sampled population from Pakistan and Kashmir (P&K) formed a new monophyletic group, albeit at a relatively lower divergence level from the central group compared to other peripheral groups (Figure 1A–C; Supplementary Figure S1). Thus, all genetic divergence values between the central and peripheral groups ($F_{ST}=0.135\pm0.06$, Figure 1D) were comparable to subspecies variations found in *A. mellifera* ($F_{ST}=0.163\pm0.073$ (Dogantzis et al., 2021)). With reference to historical taxonomic studies on relevant geographic populations of mainland *A. cerana*, we defined these genetically distinct groups as provisional subspecies: *A. cerana cerana* (central group), *A. c. japonica* (Northeast group), *A. c. guidensis* (Qinghai group), *A. c.*

abansis (Aba group), *A. c. skorikovi* (Bomi group), *A. c. kashmirensis* (Pakistan and Kashmir group), *A. c. hainana* (Hainan group), and *A. c. taiwanensis* (Taiwan group). The taxonomic considerations will be discussed in the subsequent subspecies section.

Although our results suggested clear groupings, genetic introgression has occurred frequently between the proposed peripheral and central subspecies. For example, individuals from the Liaoning population showed a much closer phylogenetic relationship to the central subspecies compared to the remaining Northeast samples (Supplementary Figure S1), with admixture analysis suggesting it may be a hybridized population between the central and Northeast subspecies (Supplementary Figure S2). Similarly, all but the Kashmir population contained hybridized individuals, suggesting recent independent genetic introgressions between peripheral and central (ancestral) subspecies (Supplementary Figures S1–S3). Although no obvious hybridization was detected in the Kashmir population, this could be attributed to insufficient sampling in intervening areas where interbreeding is more likely.

Mitochondrial phylogeny indicates hidden history of multiple radiation waves

Unlike nuclear phylogeny, where the aforementioned genetic subspecies all split from the common central subspecies (Figure 1C), mitochondrial genomes revealed a temporal order among subspecies divisions (Figure 2; Supplementary Figure S4). These differences in phylogenetic structure are likely due to variations in evolutionary rates between nuclear and mitochondrial genomes, as mitochondrial genes are known to complete lineage sorting more rapidly (Barr et al., 2005; Li et al., 2017). Alternatively, genetic drift associated with recent population bottlenecks may have led to a similar pattern (Larson et al., 2012). Within the mainland lineage, *A. c. taiwanensis* (TW) and *A. c. hainana* (HN) diverged consecutively, following the split of the mainland lineage from the ML outgroup (Figure 2). Each of these early divergent subspecies formed a separate monophyletic group in the

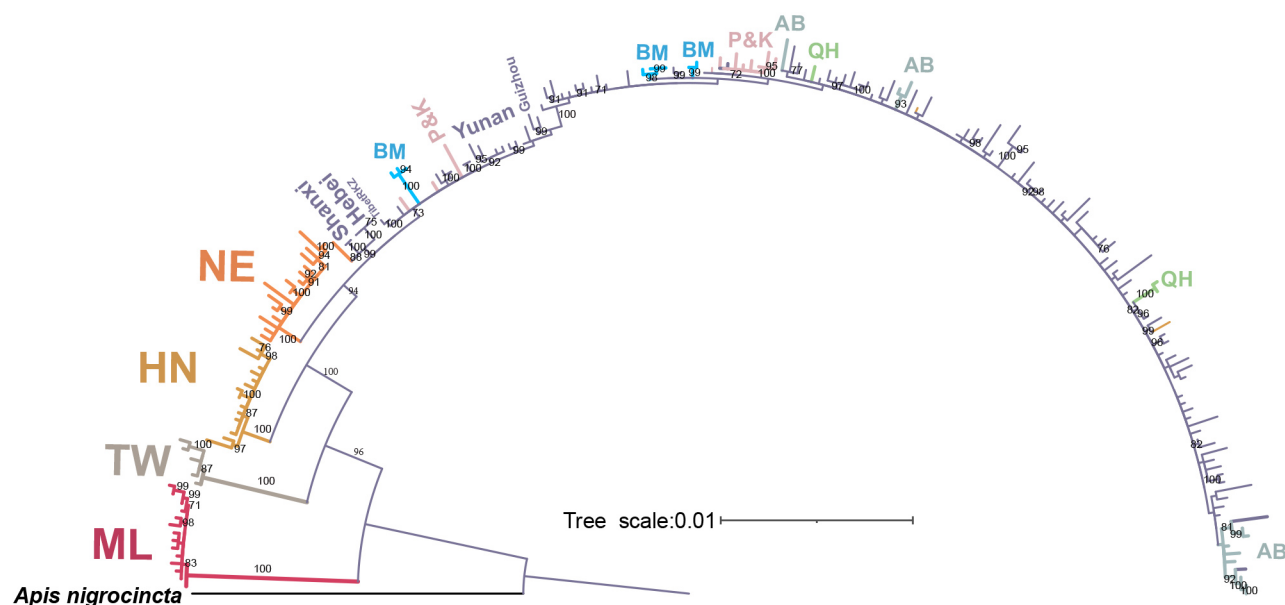


Figure 2 Phylogenetic relationships of subspecies of mainland *A. cerana* and Malaysian lineage based on mitochondrial genomes

Maximum-likelihood phylogeny was rooted using *A. nigrocincta*. Internodes with support values >70 are marked with numbers. Color codes follow those of Figure 1.

mitochondrial phylogeny, mirroring the nuclear genome results (Supplementary Figure S5). Following these early divisions, three clades formed a trichotomy in the mitochondrial phylogeny – one representing the *A. c. japonica* (NE) subspecies; one including populations from Hebei, Shanxi, and Liaoning; and one including the four remaining peripheral subspecies, *A. c. kashmirensis* (KP), *A. c. skorikovi* (BM), *A. c. abansis* (AB), and *A. c. guidensis* (QH), and geographic populations of *A. c. cerana* (CT). The Liaoning samples from the second clade showed mitochondrial introgression between *A. c. japonica* and the northern China populations (*A. c. cerana*) (Figure 2), in congruence with findings based on nuclear genomes, thus confirming hybridization. The third clade contained multiple peripheral subspecies and geographic populations of *A. c. cerana*, none of which showed mutual monophyly, reflecting recent divergence history and ongoing hybridization (Supplementary Figure S6). A more detailed tree structure is shown in Supplementary Figure S4. Overall, the mitochondrial phylogeny indicated that the current population structure of mainland *A. cerana* resulted from multiple waves of divergence across evolutionary history.

Subspecies delineation is supported by morphological characters

We compared morphological traits among the three different datasets (Supplementary Figure S7A–D) and excluded all five color-related traits from subsequent analyses due to obvious divergences across datasets (Supplementary Figure S7C). In addition, 23 outlier samples were identified in the size- and wing-related traits and were also excluded from subsequent analyses (Supplementary Figure S7E, F). The 30 remaining traits were employed to cluster samples from the HN and NE populations, where new samples nested readily within subspecies boundaries delineated by legacy samples, showing consistency in subspecies delineation in the different morphological datasets (Supplementary Figure S7G). Thus, we employed the 30 morphological characters from 211 colonies for downstream analyses. Morphological differentiations among *A. cerana* lineages and provisional subspecies of the mainland lineage were correlated with climatic characters in the local habitats. Based on hierarchical cluster analysis and PCA, the mainland Asian honeybee could be divided into two large morphological clusters, corresponding to tropical and temperate groups. The tropical bees (Indo Chinese *cerana*, *A. c. hainana*, and *A. c. taiwanensis*) formed a compact morphological cluster distinctively separate from those of mainland China, Korea, Japan, and Pakistan & Kashmir (Figures 3A, B; Supplementary Figure S8A). Notably, the Malaysian out-group, the sister lineage to all mainland *A. cerana*, was also clustered with the tropical mainland populations based on hierarchical cluster analysis. PCA showed that PC1 was mostly influenced by body size characters (Figure 3B; Supplementary Figure S8B), explaining 39.6% of the differentiation between the two morphological clusters. The size division between tropical and temperate bees was further supported by analysis using wing length characters only (Supplementary Figure S8C).

Similarly, within both tropical and temperate groups, morphological affinity among provisional subspecies and geographic populations was also correlated with local climatic parameters, where peripheral subspecies inhabiting similar habitats showed apparent morphological convergence

(Supplementary Figure S8D). Further analyses of the correlations between Dim 1–5 scores and environmental factors indicated that Dim 1 (related characters, e.g., hair length and appendage size; Supplementary Figure S8B) was highly correlated with temperature-related environmental factors (Figure 3C; Supplementary Figure S8B). These results suggest that conventional morphological characteristics were largely influenced by geography and microclimate.

Our results also showed that wing vein angles exhibited relatively weak associations with environmental factors (Dim 2 and Dim 3 in Figure 3C, Supplementary Figure S8B). Thus, wing morphological geometrics were applied to validate morphological differences among the eight provisional subspecies defined by nuclear genomic divergences and evolutionary history. All eight provisional subspecies and the Malaysian out-group showed consistent morphological clustering based on wing geometric morphology (Figure 3D), and *P*-values for MD among groups were less than 0.001 (Supplementary Table S4), supporting the subspecies delineations. We discovered a significant correlation between F_{ST} (Figure 1D) and MD (Supplementary Table S4) ($R=0.44$, $P=0.008$), indicating a consistent relationship between genetic and morphological differentiations among groups (Supplementary Figure S8E). In addition, pairwise MDs between peripheral and central subspecies (4.21 ± 0.75) were significantly lower than those between peripheral subspecies (5.49 ± 0.73) ($P<0.001$, two-tailed Mann-Whitney test). The same pattern was observed for F_{ST} estimations (0.135 ± 0.06 vs. 0.28 ± 0.08 , $P<0.001$, two-tailed Mann-Whitney test). These findings indicate that the central subspecies was the closest lineage to each of the seven peripheral subspecies, as supported by both molecular- and morphology-based evidence. The wing shape groups with positive and negative loadings on CV1 and CV2 are shown on Supplementary Figure S8F.

Subspecies descriptions and taxonomic annotations on mainland *A. cerana*

Based on convergent evidence from morphology, nuclear genetic divergence, evolutionary history, and phylogenetic relationships, we herein formally define eight subspecies for the mainland *A. cerana* lineage: i.e., central subspecies *A. c. cerana* and seven peripheral subspecies, *A. c. taiwanensis*, *A. c. hainana*, *A. c. kashmirensis*, *A. c. skorikovi*, *A. c. abansis*, *A. c. guidensis*, and *A. c. japonica*. A revision of the subspecies taxonomy of mainland *A. cerana*, including descriptions of three newly defined subspecies, is provided in Supplementary Text. The holotypes of *A. c. taiwanensis*, *A. c. guidensis*, and *A. c. kashmirensis* and neotypes of *A. c. hainana* and *A. c. abansis*, as well as locations of four highland subspecies, are shown in Supplementary Figures S9–11. Morphometric character measurements of worker bees of the eight *A. cerana* subspecies are provided in Supplementary Table S5.

DISCUSSION

Subspecies of mainland *A. cerana*

In the present study, we analyzed 362 *A. cerana* specimens, representing almost all distinct groups of mainland *A. cerana* known to date. Our results demonstrated that all previously identified genetically distinct groups (i.e., six peripheral groups – NE, QH, AB, BM, HN, and TW) and the newly added KP



group each formed a monophyletic group based on nuclear SNP phylogeny. In addition, all peripheral groups showed notable genetic divergences from the central group, while the latter formed a paraphyly intermingled with all peripheral groups (Figure 1C). In this new analysis, we only retained non-kin individuals to better characterize intra- and inter-population genetic features. With this stringent sampling scheme, genetic divergence between populations was expected to decrease, explaining the slight reduction in F_{ST} between the QH and CT groups compared with our previous work (0.09 and 0.10 in current study and Ji et al. (2020), respectively). Even so, the genetic divergence values detected between pairs of mainland *A. cerana* subspecies (F_{ST} range 0.07–0.26, between each peripheral subspecies and central subspecies, Figure 1D) were on par with those of *A. mellifera* subspecies (0.038–0.263) (Dogantzis et al., 2021). In addition, both phylogenetic distances from the current study (Supplementary Figure S1) and F_{ST} estimates from our previous work (Ji et al., 2020, Supplementary Figure S6B) revealed limited differentiation within the central group, even among distantly located geographic populations. Therefore, we regarded all central group populations as a single subspecies. Based on phylogenetic monophyly and divergence values, we propose that these genetically distinct

Mitochondrial phylogeny further indicated that the current population structure of mainland *A. cerana* was the result of multiple waves of divergence. This newly discovered temporal sequence is an important addition to the centrifugal diversification model (Ji et al., 2020). Island subspecies represented by *A. c. taiwanensis* and *A. c. hainana* were among the first diverging sublineages (Figure 2). The relatively long divergence time (inferred by basal phylogenetic positions) and ocean isolation may explain the genetic distinctiveness of the island subspecies. Alternatively, the observed genetic distinctiveness and basal phylogenetic placement may have resulted from recent genetic drift following population bottlenecks (Larson et al., 2012). Distinguishing between these two evolutionary models requires further investigation, with direct evidence from historical honeybee samples. The *A. c. japonica* subspecies was likely the first continentally distributed group to separate from the common ancestor (Figure 2). It remains unclear how this subspecies maintained genetic isolation from current *A. c. cerana* in the absence of

obvious geographic barriers. Despite the presence of hybrid individuals identified in the Liaoning region, genetic introgression from the central subspecies was not prevalent in *A. c. japonica*. Genomic incompatibility (e.g., structural variations) and differing fitness (e.g., cold hardiness (Liu et al., 2022)) are among the possible mechanisms remaining to be tested. Finally, the most recent range expansion(s) led to parallel invasions of the central ancestor into the peripheral valleys of the QTP (Figure 2). Formed by continuous mountain formation events in the Himalayans, these deep valleys are characterized by abrupt vertical transitions in vegetation, with the tree line confined to mid-hill elevations (Wang et al., 2022). As a cavity-nesting honeybee, *A. cerana* relies on tree hollows to build nests. Therefore, the barren ridges defining these valleys may have effectively served as terrestrial barriers, preventing local honeybee populations from cross-valley migration and direct interbreeding. Compared to the three species that diverged earlier, the four QTP valley subspecies (*A. c. guidensis*, *A. c. abansis*, *A. c. skorikovi*, and *A. c. kashmirensis*) showed elevated levels of admixture with *A. c. cerana*. Both the relatively young divergence time and lack of absolute geographic isolation (lasting secondary contact with *A. c. cerana* at the valley opening) are likely reasons for their relatively low genetic divergences from the central subspecies (Figure 1D), whereas ongoing hybridization and retention of ancestral polymorphisms likely explain the paraphyly based on mitochondrial phylogeny (Figure 2). Individuals from QH (*A. c. guidensis*), AB (*A. c. abansis*), and BM (*A. c. skorikovi*) contained mitochondrial haplotypes derived from *A. c. cerana* (Supplementary Figure S6) and the QH and AB populations exhibited elevated genome-wide heterozygosity rates (Supplementary Figure S3), indicating hybridization with *A. c. cerana*. Conversely, individuals from BM (*A. c. skorikovi*) demonstrated no clear signs of mitochondrial introgression (Supplementary Figure S6, this study), but exhibited a lower genome-wide heterozygosity rate (Supplementary Figure S3), suggesting that retention of ancestral polymorphism is the most likely evolutionary cause of mitonuclear discordance, although additional sampling is needed for confirmation.

Although the divergence order of these young subspecies remains unknown, their evolutionary independence not only helps clarify their genetic boundaries but also the constraining factors that determine their distributional ranges. For the QTP valley subspecies, understanding how valley isolation has shaped their formation provides critical clues for drawing fine distributional borders. Additionally, a global understanding of the phylogenetic relationship based on comprehensive sampling is essential for revealing the genealogy of local populations. For example, in a recent study, the close phylogenetic relationship between *A. cerana* samples collected in the plains region of northern Pakistan was interpreted as the result of human introduction (Tan et al., 2021). However, our new sampling scheme covering major geographic populations of mainland *A. cerana* indicated an alternative hypothesis, namely, that the Pakistani populations were part of the widely distributed *A. c. cerana*. Further sampling of *A. cerana* populations in South Asian countries will better clarify the boundaries among *A. c. kashmirensis*, *A. c. cerana*, *A. c. skorikovi*, and *A. c. indica*. In addition, given the wide distribution range of *A. cerana* in Asia, we anticipate that new subspecies may be discovered in future research, especially in regions not included in the current study, such as deep QTP valleys, central and southern India, and Sri Lanka

(Oldroyd et al., 2006; Shanas et al., 2022).

Integrative taxonomy of honeybee subspecies

Subspecies are taxonomic assignments used to classify evolutionary units showing evolutionary independence and genetic/morphological distinctiveness from other such units (Braby et al., 2012). Our study showed that the *A. cerana* subspecies resulted from multiple divergence events, representing transient evolutionary units along the temporal spectrum of speciation. Thus, all characters used to delineate subspecies are dependent on the divergence time of the focal population from its ancestor and subsequent interactions (e.g., hybridization) with other evolutionary units. However, it remains challenging, and in many cases impossible, to identify a fixed level of variation for any given character, including morphology and genetics.

This issue became evident when quantifying genetic divergence and completeness of lineage sorting among the *A. cerana* subspecies. Notably, the early diverging subspecies (*A. c. taiwanensis*, *A. c. hainana*, and *A. c. japonica*) exhibited larger genetic divergences and exclusive monophyly in both the nuclear and mitochondrial trees. In contrast, the younger QTP valley subspecies experienced more intensive and recent introgression events from *A. c. cerana*, resulting in lower genetic divergences and paraphyly in the mitochondrial tree (Figures 1, 2). Therefore, for these later diverging subspecies, subspecies delineation based on genetic divergence or monophyly alone will require arbitrary thresholds. Moreover, the level of genomic variation observed in the honeybee populations resulted from multiple factors, including genomic divergence due to historical division and isolation, natural selection, ongoing introgression, and anthropogenic influences. Clarification of the relative contributions of these factors is necessary to understand the extent to which observed genetic features are a consequence of evolutionary isolation.

Morphological subspecies delimitation faces similar challenges. Our study demonstrated that honeybee morphology, including characters typically employed in differentiating subspecies (e.g., body size), was strongly related to environmental features, in congruence with findings from regional studies (Tan et al., 2003). Such a pattern has created major challenges in the application of morphological characters in defining honeybee subspecies. This issue is particularly common for species with large geographic ranges, such as *A. cerana*, where geographic populations inhabit and have adapted to highly heterogeneous environments (Hepburn et al., 2001a; Tan et al., 2006). In some cases, a wide-ranging lineage may retain relatively distinctive characteristics despite experiencing continuous within-group genetic exchange, such as in *A. c. cerana*. On the other hand, distantly related (and genetically distinct) populations living under similar climatic conditions may show convergent traits. For example, several independent montane subspecies of *A. cerana* share a darkened body coloration and enlarged body size, whereas tropical populations are almost all lighter in color and smaller in size. In these cases, morphological characters alone would incorrectly group distant lineages as the same or close related subspecies. Thus, morphological traits influenced by the environment should not be used in subspecies delineation, as these traits cannot correctly trace underlying evolutionary history.

The delineation of species and subspecies is integral to

conservation because they are the units most used for conservation management. However, taxonomic research is always evolving, based on continuous effort in formulating working hypotheses and ever-improving species/subspecies definitions (Orr et al., 2021). This process is built upon many pioneering studies, with the goal of understanding the taxonomic boundaries and evolutionary mechanisms that have led to the biodiversity we see today. How we do this has immense ramifications for conservation and science. Therefore, we propose that multiple lines of evidence, including, but not limited to, morphology, genetics, biology, and ecology, should be employed to determine subspecies boundaries. This practice should abide by the general rules of taxonomy (e.g., establishment and curation of type specimens) and data sharing in public domains to better enable subsequent studies to leverage and synthesize data.

To ensure best practices and facilitate replicability and comparability, we propose that honeybee subspecies must meet the following criteria: (1) Different subspecies should have undergone independent evolutionary histories supported by integrative evidence of monophyly, preferably including molecular data. Given the difference in inheritance patterns between nuclear and mitochondrial genes, discordance in phylogenetic history is frequently revealed by nuclear and mitochondrial data (Toews & Brelsford, 2012), especially in young and radiating species (Shaw, 2002). Here, we used nuclear genome sequences as the primary data to better reconstruct the evolutionary history of *A. cerana* populations. Additionally, we employed mitochondrial data to specify evolutionary mechanisms causing genetic admixture among populations, such as hybridization and retention of ancestral polymorphism (Garner et al., 1992; Hausdorf et al., 2011; Rodríguez et al., 2010); (2) Each subspecies should be confined to a distribution that is largely disjunct from other such lineages; and (3) Subspecies should be diagnosable based on a suite of combined characteristics, as well as molecular, morphological, and other types of evidence.

Conservation of mainland *A. cerana*

From an evolutionary perspective, delineating honeybee subspecies necessitates determining the relative contributions of evolutionary and contemporary events to the current status of a focal population. Such an endeavor perfectly aligns with conservation initiatives, where the challenges posed to the studied organism lie in both historical and ongoing processes.

We identified seven peripheral subspecies within the mainland *A. cerana* lineage, each with an independent history and constrained to their current ranges by various geographical and biological features. Compared with the widely distributed *A. c. cerana* subspecies, these isolated subspecies exhibited distinct genetic differentiation and restricted habitat ranges, indicating that each should be treated as an independent conservation unit. Clarification of subspecies boundaries is crucial in prioritizing conservation efforts. Once these target units are defined, further dissection of potential historical and contemporary threats can help to identify conservation priorities, in line with the concept of evolutionarily significant units (ESUs) (Ryder, 1986), which have been extensively discussed in the context of conservation biology (Casacci et al., 2014; De Guia & Saitoh, 2007; Dogantzis & Zayed, 2019; Requier et al., 2019).

Conservation strategies for peripheral subspecies should aim to maintain their integrity, in line with their parallel

evolutionary history and isolated distributions. For example, *A. c. taiwanensis* and *A. c. hainana* likely formed via geographical isolation. On concern, elevated hybridization with *A. c. cerana* has been found in the Hainan subspecies, in concordance with extensive colony introductions from the Guangxi and Guangdong provinces in the 1990s (Zhao et al., 2017). This has had a significant influence on the current genetic structure of the Hainan subspecies, highlighting the need for immediate intervention and prohibition of colony introductions from the mainland to prevent further genetic degradation.

Conversely, geographic populations belonging to the central subspecies *A. c. cerana* showed much more homogenized genetic structure due to long-term introgression. Major geographic barriers, including the Qinling Mountains, which define the western boundary between the Palearctic and Oriental Realms, showed limited effects in preventing introgression among geographic populations within *A. c. cerana*. Therefore, a primary conservation goal for this subspecies should include prevention of population compartmentalization caused by anthropogenic activities, such as urbanization, habitat fragmentation, uniform crop plantation, and replacement with introduced honeybee species such as *A. mellifera* (Abrol, 2013; Yang, 2009). Furthermore, the knowledge that honeybee subspecies have independently evolved and adapted to their current habitats (Ji et al., 2020) warrants increased caution in honeybee cross-breeding. For example, queens of montane (e.g., *A. c. guidensis* and *A. c. abansis*) and northern subspecies (*A. c. japonica*) and races are often introduced to queenless tropical colonies to improve honey production (Wang et al., 2021). Hybridized progeny often show rapid trait regression, leading to the persistent importation of even more exogenous queens. Hybridization regression is unsurprising considering that local subspecies are adapted to specific tropical or montane environments. However, to what extent these anthropogenic activities are changing the genetic structure of the target population remains largely untested. Unlike in the western honeybee *A. mellifera*, most *A. cerana* breeding practices have been conducted within the native range of the species. Understanding how these activities may alter the genetics and fitness of endemic honeybees is crucial for their future survival. The clear delineation of conservation targets provides a pathway to understand ongoing impacts and track long-term influences, enabling more effective management and the preservation of these unique honeybees.

DATA AVAILABILITY

The newly sequenced data have been archived at the NCBI SRA under accession No. PRJNA870246, as well as the Genome Sequence Archive (<https://ngdc.cncb.ac.cn/gsz>) and Science Data Bank (<https://www.scidb.cn/en>) under accessions PRJCA015358 and 10.57760/sciencedb.07636, respectively.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Xin Z. and S.L. designed the study. L.Q. performed morphometrics analysis. S.L. conducted population genomics and mitochondrial analyses. J.D. assisted with sequencing data analysis. S.H.P. and A.M. acquired Indian

specimens. Q.N., T.J., X.L., and Xin Z. organized sampling. X.L. provided samples from Northeast China and Japan. M.O., C.Z., X Zhang., Xuguo Z., and S. Luo assisted in manuscript writing. K.T. provided morphological data for a series of geographic populations. Xin Z., S.L., and L.Q. wrote the first drafts, and all authors contributed to and proof-read the manuscript. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We gratefully acknowledge sample contributions from many collaborators, including Hu Li, Syed I.A. Shah, Qingtao Wu, Zeqing Niu, Dan Zhang, Jinglin Gao, Yihai Zhong, Jiahui Hu, Zhenhua Chen, and Jialong Huang, among many others. We thank Guanhuang Yang, Chao Chen, and Wei Shi from the Institute of Apicultural Research of the Chinese Academy of Agricultural Sciences for help in tracing type specimens of *A. c. hainana* and *A. c. abansis* subspecies. S.H.P. acknowledges the National Biodiversity Authority (NBA) of the Government of India for providing approval to send honeybee samples to Xin Z. for non-commercial research purposes.

REFERENCES

- Abrol DP. 2013. Asiatic Honeybee *Apis cerana*: Biodiversity Conservation and Agricultural Production. Dordrecht: Springer.
- Barr CM, Neiman M, Taylor DR. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. *New Phytologist*, **168**(1): 39–50.
- Braby MF, Eastwood R, Murray N. 2012. The subspecies concept in butterflies: has its application in taxonomy and conservation biology outlived its usefulness?. *Biological Journal of the Linnean Society*, **106**(4): 699–716.
- Casacci LP, Barbero F, Balletto E. 2014. The “Evolutionarily Significant Unit” concept and its applicability in biological conservation. *Italian Journal of Zoology*, **81**(2): 182–193.
- Chen C, Wang HH, Liu ZG, et al. 2018a. Population genomics provide insights into the evolution and adaptation of the Eastern Honey Bee (*Apis cerana*). *Molecular Biology and Evolution*, **35**(9): 2260–2271.
- Chen SF, Zhou YQ, Chen YR, et al. 2018b. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**(17): i884–i890.
- Chen YC. 1993. Apiculture of China. Beijing: China Agricultural Publication. (in Chinese)
- Christmas MJ, Jones JC, Olsson A, et al. 2021. Genetic barriers to historical gene flow between cryptic species of Alpine Bumblebees revealed by comparative population genomics. *Molecular Biology and Evolution*, **38**(8): 3126–3143.
- Danecek P, Auton A, Abecasis G, et al. 2011. The variant call format and VCFtools. *Bioinformatics*, **27**(15): 2156–2158.
- De Guia APO, Saitoh T. 2007. The gap between the concept and definitions in the Evolutionarily Significant Unit: the need to integrate neutral genetic variation and adaptive variation. *Ecological Research*, **22**(4): 604–612.
- De Queiroz K. 2007. Species concepts and species delimitation. *Systematic Biology*, **56**(6): 879–886.
- De Queiroz K. 2020. An updated concept of subspecies resolves a dispute about the taxonomy of incompletely separated lineages. *Herpetological Review*, **51**(3): 459–461.
- Dogantzis KA, Tiwari T, Conflitti IM, et al. 2021. Thrive out of Asia and the adaptive radiation of the western honey bee. *Science Advances*, **7**(49): eabj2151.
- Dogantzis KA, Zayed A. 2019. Recent advances in population and quantitative genomics of honey bees. *Current Opinion in Insect Science*, **31**: 93–98.
- Engel MS. 1999. The taxonomy of recent and fossil honey bees (Hymenoptera: Apidae; *Apis*). *Journal of Hymenoptera Research*, **8**(2): 165–196.
- Garnery L, Cornuet JM, Solignac M. 1992. Evolutionary history of the honey bee *Apis mellifera* inferred from mitochondrial DNA analysis. *Molecular Ecology*, **1**(3): 145–154.
- Groves CP, Cotterill FPD, Gippoliti S, et al. 2017. Species definitions and conservation: a review and case studies from African mammals. *Conservation Genetics*, **18**(6): 1247–1256.
- Guillot G, Renaud S, Ledevin R, et al. 2012. A unifying model for the analysis of phenotypic, genetic, and geographic data. *Systematic Biology*, **61**(6): 897–911.
- Gutiérrez EE, Helgen KM. 2013. Outdated taxonomy blocks conservation. *Nature*, **495**(7441): 314.
- Hausdorf B, Wilkens H, Strecker U. 2011. Population genetic patterns revealed by microsatellite data challenge the mitochondrial DNA based taxonomy of *Astyanax* in Mexico (Characidae, Teleostei). *Molecular Phylogenetics and Evolution*, **60**(1): 89–97.
- Hepburn HR, Radloff SE, Verma S, et al. 2001a. Morphometric analysis of *Apis cerana* populations in the southern Himalayan region. *Apidologie*, **32**(5): 435–447.
- Hepburn HR, Smith DR, Radloff SE, et al. 2001b. Intraspecific categories of *Apis cerana*: morphometric, allozymal and mtDNA diversity. *Apidologie*, **32**(1): 3–23.
- Hillis DM. 2019. Species delimitation in herpetology. *Journal of Herpetology*, **53**(1): 3–12.
- Hou CS, Li BB, Luo YX, et al. 2016. First detection of *Apis mellifera* filamentous virus in *Apis cerana cerana* in China. *Journal of Invertebrate Pathology*, **138**: 112–115.
- Ilyasov RA, Park J, Takahashi J, et al. 2018. Phylogenetic uniqueness of Honeybee *Apis cerana* from the Korean peninsula inferred from the mitochondrial, nuclear, and morphological Data. *Journal of Apicultural Science*, **62**(2): 189–214.
- Ilyasov RA, Youn HG, Lee ML, et al. 2019. Phylogenetic relationships of Russian Far-East *Apis cerana* with other north Asian populations. *Journal of Apicultural Science*, **63**(2): 289–314.
- Ji YK, Li XG, Ji T, et al. 2020. Gene reuse facilitates rapid radiation and independent adaptation to diverse habitats in the Asian honeybee. *Science Advances*, **6**(51): eabd3590.
- Kalyaanamoorthy S, Minh BQ, Wong TKF, et al. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, **14**(6): 587–589.
- Klingenberg CP. 2011. MORPHOJ: an integrated software package for geometric morphometrics. *Molecular Ecology Resources*, **11**(2): 353–357.
- Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, **15**(1): 356.
- Larson G, Karlsson EK, Perri A, et al. 2012. Rethinking dog domestication by integrating genetics, archeology, and biogeography. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(23): 8878–8883.
- Lefort V, Desper R, Gascuel O. 2015. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Molecular Biology and Evolution*, **32**(10): 2798–2800.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, **25**(14): 1754–1760.
- Li YY, Zhang R, Liu SL, et al. 2017. The molecular evolutionary dynamics of oxidative phosphorylation (OXPHOS) genes in Hymenoptera. *BMC Evolutionary Biology*, **17**(1): 269.
- Liu NN, Liu HM, Ju Y, et al. 2022. Geometric morphology and population genomics provide insights into the adaptive evolution of *Apis cerana* in Changbai Mountain. *BMC Genomics*, **23**(1): 64.
- Lo N, Gloag RS, Anderson DL, et al. 2010. A molecular phylogeny of the genus *Apis* suggests that the giant honey bee of the Philippines, *A. brevilingula* Maa, and the plains honey bee of southern India, *A. indica*

- Fabricius, are valid species. *Systematic Entomology*, **35**(2): 226–233.
- Ma DF, Huang WC. 1981. Apiculture in the New China. *Bee World*, **62**(4): 163–166.
- Maa TC. 1953. An inquiry into the systematics of the tribus Apidini or honeybees (Hym.). *Treubia*, **21**(3): 525–640.
- Malinsky M, Challis RJ, Tyers AM, et al. 2015. Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science*, **350**(6267): 1493–1498.
- Manichaikul A, Mychaleckyj JC, Rich SS, et al. 2010. Robust relationship inference in genome-wide association studies. *Bioinformatics*, **26**(22): 2867–2873.
- McKenna A, Hanna M, Banks E, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, **20**(9): 1297–1303.
- Minh BQ, Schmidt HA, Chernomor O, et al. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, **37**(5): 1530–1534.
- Morrison III WR, Lohr JL, Duchon P, et al. 2009. The impact of taxonomic change on conservation: Does it kill, can it save, or is it just irrelevant?. *Biological Conservation*, **142**(12): 3201–3206.
- Nakamura T, Yamada KD, Tomii K, et al. 2018. Parallelization of MAFFT for large-scale multiple sequence alignments. *Bioinformatics*, **34**(14): 2490–2492.
- National Animal Genetic Resources Committee. 2011. Animal Genetic Resources in China: Bees. Beijing: China Agriculture Press. (in Chinese)
- Neuditschko M, Khatkar MS, Raadsma HW. 2012. NetView: a high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. *PLoS One*, **7**(10): e48375.
- Oldroyd BP, Reddy MS, Chapman NC, et al. 2006. Evidence for reproductive isolation between two colour morphs of cavity nesting honey bees (*Apis*) in south India. *Insectes Sociaux*, **53**(4): 428–434.
- Orr MC, Ferrari RR, Hughes AC, et al. 2021. Taxonomy must engage with new technologies and evolve to face future challenges. *Nature Ecology & Evolution*, **5**(1): 3–4.
- Peng YS, Nasr ME, Locke SJ. 1989. Geographical races of *Apis cerana* Fabricius in China and their distribution. Review of recent Chinese publications and a preliminary statistical analysis. *Apidologie*, **20**(1): 9–20.
- Phillimore AB, Owens IPF. 2006. Are subspecies useful in evolutionary and conservation biology?. *Proceedings of the Royal Society B: Biological Sciences*, **273**(1590): 1049–1053.
- Purcell S, Neale B, Todd-Brown K, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**(3): 559–575.
- Radloff SE, Hepburn C, Hepburn HR, et al. 2010. Population structure and classification of *Apis cerana*. *Apidologie*, **41**(6): 589–601.
- Radoszkowski O. 1887. Hyménoptères de Korée. *Horae Societatis Entomologicae Rossicae*, **21**: 428–436.
- Requier F, Garnery L, Kohl PL, et al. 2019. The conservation of native Honey Bees is crucial. *Trends in Ecology & Evolution*, **34**(9): 789–798.
- Rodríguez F, Pérez T, Hammer SE, et al. 2010. Integrating phylogeographic patterns of microsatellite and mtDNA divergence to infer the evolutionary history of chamois (genus *Rupicapra*). *BMC Evolutionary Biology*, **10**(1): 222.
- Ruttner F. 1988. Biogeography and Taxonomy of Honeybees. Berlin: Springer.
- Ryder OA. 1986. Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology & Evolution*, **1**(1): 9–10.
- Shanas S, Anju KG, Mashhoor K. 2022. Identity of cavity nesting honey bees of the Indian subcontinent with description of a new species (Hymenoptera: Apidae: Apinae: Apini: *Apis*). *Entomon*, **47**(3): 197–220.
- Shaw KL. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation: What mtDNA reveals and conceals about modes of speciation in Hawaiian crickets. *Proceedings of the National Academy of Sciences of the United States of America*, **99**(25): 16122–16127.
- Skotte L, Korneliussen TS, Albrechtsen A. 2013. Estimating individual admixture proportions from next generation sequencing data. *Genetics*, **195**(3): 693–702.
- Smith DR, Hagen RH. 1996. The biogeography of *Apis cerana* as revealed by mitochondrial DNA sequence data. *Journal of the Kansas Entomological Society*, **69**(4): 294–310.
- Tan HW, Naeem M, Ali H, et al. 2021. Genome sequence of the Asian honeybee in Pakistan sheds light on its phylogenetic relationship with other honeybees. *Insects*, **12**(7): 652.
- Tan K, Fuchs S, Koeniger N, et al. 2003. Morphological characterization of *Apis cerana* in the Yunnan province of China. *Apidologie*, **34**(6): 553–561.
- Tan K, Hepburn HR, Radloff SE, et al. 2008. Multivariate morphometric analysis of the *Apis cerana* of China. *Apidologie*, **39**(3): 343–353.
- Tan K, Meixner MD, Fuchs S, et al. 2006. Geographic distribution of the eastern honeybee, *Apis cerana* (Hymenoptera: Apidae), across ecological zones in China: Morphological and molecular analyses. *Systematics and Biodiversity*, **4**(4): 473–482.
- Tarasov A, Vilella AJ, Cuppen E, et al. 2015. Sambamba: fast processing of NGS alignment formats. *Bioinformatics*, **31**(12): 2032–2034.
- Toews DPL, Brelsford A. 2012. The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**(16): 3907–3930.
- Vung NN, Kim I, Lee MY, et al. 2018. Controlling sacbrood virus disease in *Apis cerana* colonies with biological methods in Korea. *Journal of Apiculture*, **33**(4): 283–295.
- Wang XY, Wang T, Xu JF, et al. 2022. Enhanced habitat loss of the Himalayan endemic flora driven by warming-forced upslope tree expansion. *Nature Ecology & Evolution*, **6**(7): 890–899.
- Wang Z, Chen DH, Wang Q, et al. 2021. Research on the development status and countermeasures of the bee seed industry in China. *Journal of Bee*, **41**(11): 12–18. (in Chinese)
- Yang GH. 2009. The effect of *Apis cerana cerana* on forest ecosystem. *Apiculture of China*, **60**(4): 5–7, 10. (in Chinese)
- Yang GH, Lin GL, Sun QH, et al. 1995. The distribution, morphological variation and behavior of Chinese honeybee in northwestern plateau of Sichuan province. *Scientia Agricultura Sinica*, **28**(S1): 202–206. (in Chinese)
- Yang GH, Xu SY, Kuang BY, et al. 1986. The distribution of the genus *Apis cerana* Fab. in China and some of its subspecies. *Journal of Yunnan Agricultural University*, **1**(1): 89–92. (in Chinese)
- Zhao HL. 2018. Problems and countermeasures of bee cultivation in poor mountainous areas of Hainan. *Anhui Agricultural Science Bulletin*, **24**(12): 10–12. (in Chinese)
- Zhao WZ, Wang M, Liu YQ, et al. 2017. Phylogeography of *Apis cerana* populations on Hainan island and southern mainland China revealed by microsatellite polymorphism and mitochondrial DNA. *Apidologie*, **48**(1): 63–74.
- Zhuang DA. 1989. New subspecies of *Apis cerana*. *Southwest China Journal of Agricultural Sciences*, **2**(4): 61–65. (in Chinese)

Supplementary Materials

Defining honeybee subspecies in an evolutionary context warrants strategized conservation

Lifei Qiu¹, Jiangxing Dong¹, Xingan Li², Sajad H. Parey³, Ken Tan⁴, Michael Orr⁵, Aquib Majeed³,
Xue Zhang¹, Shiqi Luo¹, Xuguo Zhou⁶, Chaodong Zhu⁵, Ting Ji⁷, Qingsheng Niu², Shanlin Liu^{1,*}, and
Xin Zhou^{1,*}

1. Department of Entomology, College of Plant Protection, China Agricultural University, Beijing 100193, China
2. Key Laboratory for Bee Genetics and Breeding, Jilin Provincial Institute of Apicultural Sciences, Jilin, Jilin 132108, China
3. Department of Zoology, School of Biosciences and Biotechnology, Baba Ghulam Shah Badshah University, Rajouri (Jammu and Kashmir) 185234, India
4. CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Xishuangbanna, Yunnan 650000, China
5. Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China
6. Department of Entomology, University of Kentucky, Lexington, KY 40546, USA
7. Yangzhou University, Yangzhou, Jiangsu 225009, China

*Corresponding authors, E-mail: shanlin.liu@cau.edu.cn; xinzhou@cau.edu.cn

Supplementary Text

A revision of subspecies taxonomy of Mainland Asian honeybee *Apis cerana*

Apis cerana cerana Fabricius, 1793 (Fabricius, 1792-1793)

Apis cerana pekinga Zhuang, 1989 (Zhuang, 1989). Nomen nudum.

Apis cerana bijieca Zhuang, 1989 (Zhuang, 1989). Nomen nudum.

Apis cerana kweiyanga Zhuang, 1989 (Zhuang, 1989). Nomen nudum.

Apis cerana twolakeca Zhuang, 1989 (Zhuang, 1989). Nomen nudum.

Apis cerana shankianga Zhuang, 1989 (Zhuang, 1989). Nomen nudum.

Apis ceran fantsun mountaina Zhuang, 1989 (Zhuang, 1989). Nomen nudum.

Apis ceran shenp countica Zhuang, 1989 (Zhuang, 1989). Nomen nudum.

Apis ceran cathayca Zhuang, 1989 (Zhuang, 1989). Nomen nudum.

Apis cerana twolareca Chen, 1993 (Chen, 1993). Nomen nudum.

Type locality. The type samples were reported as being collected from the eastern mountains of China, without a specific locality (Fabricius, 1792-1793). The type locality was recently concluded to be Jiangxi Province (Kuang et al., 2007) within the current range of the central lineage in our study.

Distribution. This subspecies occurs as far west as Afghanistan and expands southward along the central mountain ranges. Along the Himalayan uplift in southeastern Tibet, it is replaced by *Apis cerana skorikovi* and further south in India, it is replaced by *A. c. indica* (with overlapping regions likely in northern or central India). The subspecies then ranges across most of central and southern China, including most mountain valleys along the east and southeast edges of the Qinghai-Xizang (Tibet) plateau, with exceptions where two montane subspecies (*A. c. abansis* and *A. c. guidensis*) occur. *A. c. cerana* then ranges along eastern China up to the Huabei plain and south to the Kra Isthmus of Thailand, including most of China, Afghanistan, Pakistan, Nepal, Bhutan, Northern India, Bangladesh, Myanmar, Laos, Cambodia, Vietnam and Northern Thailand.

Diagnosis. The morphological characters of this subspecies are highly variable due to strong climatic influences associated with local habitats. The most reliable morphological diagnosis is the wing characters described in the results. Additional morphometric values are provided in Supplementary Table S5.

Evolutionary synopsis. This subspecies represents the descendant of the central ancestral population, from which the current peripheral subspecies were derived (Ji et al., 2020). The historical range of *Apis cerana cerana* is hypothesized to have undergone multiple waves of expansion and contraction following climatic oscillation and glaciations. During the last of such cycles, the subspecies has likely expanded northward from its glacial resort around the Sundaland region till it reaches the current range limits, forming secondary contacts with all peripheral subspecies, except for *A. c. taiwanensis* and *A. c. hainana*. Despite its large range and diverse climates associated with local habitats, genetic structure was weak within the subspecies, where intensive hybridization and introgression are detected among regional populations.

Apis cerana taiwanensis Qiu, Liu & Zhou, ssp. nov.

Supplementary Figure S9: A~C

ZooBank taxon LSID: E8BB2998-9CBD-4E8F-AA04-28876B5E3EAD

Type locality. Holotype: worker; China: Taiwan Province, Jiayi Co., October 6th, 2020 (deposited in the Department of Entomology, China Agricultural University, Beijing, China). Paratypes: Three workers, same collection data as the holotype (deposited in the Department of Entomology, China Agricultural University, Beijing, China).

Distribution. This subspecies occurs on the island of Taiwan.

Diagnosis. *Apis cerana taiwanensis* is of small body size, with a forewing length around $816.93 \pm 9.12 \mu\text{m}$. Workers are yellowish, with a yellow metasoma and scutellum. Additional morphometric values are provided in Supplementary Table S5.

Evolutionary synopsis. After the split of the Mainland *Apis cerana* lineage from the out-group, a frontier population of the common central ancestor expanded its range into the current Taiwan Island during the peak of glacial impact and then became isolated by the straight formed during the interglacial period. This population eventually evolved into *A. c. taiwanensis*. Gene introgression is rare in this subspecies, as a consequence of its early divergence and significant barriers to inbreeding with mainland populations. Alternatively, recent population bottleneck and genetic drift may have also contributed to the current population structure.

Apis cerana hainana Yang & Kuang, 1986 (Yang et al., 1986)

Supplementary Figure S9: D~F

Apis cerana hainanensis Hepburn, Smith, Radloff & Otis, 2001 (Hepburn et al., 2001). Nomen nudum.

Type locality. Holotype: worker; China: Hainan Province, Qiongzong Co., September 1976 (deposited in the Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing, China). Paratypes: worker; China: Hainan Province, Wenchang Co., September 1976 (deposited in the Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing, China). The holotype and paratype specimens are confirmed missing. We thus establish neotypes in the present work.

Neotype: worker; China: Hainan Province, Wenchang Co., June 16th 2020 (deposited in the Department of Entomology, China Agricultural University, Beijing, China). Paratypes: three workers; China: Hainan Province, Danzhou Co., March 26th 2013 (deposited in the Department of Entomology, China Agricultural University, Beijing, China)

Distribution. This subspecies occurs primarily on the main island of Hainan, with a remote (>20 Km from any adjacent coast) and isolated population also found on the Yongxing Island of the South China Sea.

Diagnosis. This subspecies is of smaller size, with a forewing length of $791.56 \pm 10.12 \mu\text{m}$. The body color is yellowish, with a yellow metasoma, scutellum and legs. Additional morphometric values are provided in Supplementary Table S5.

Evolutionary synopsis. *Apis cerana hainana* probably underwent a similar history as *A. c. taiwanensis* during a later glacial cycle. We observed an elevated level of gene introgression between the Hainan subspecies and the mainland population. This result

is consistent with the extensive introduction of mainland colonies in recent decades (Zhao et al., 2017). The history of the Yongxing Island population remains as a mystery.

***Apis cerana japonica* Radoszkowski, 1877 (Radoszkowski, 1887)**

Apis cerana Changbai Mountain race (National Animal Genetic Resources Committee, 2011)

Apis cerana ussuriensis Ilyasov, 2019 (Ilyasov et al., 2021).

Apis cerana koreana Ilyasov et al., 2019 (Ilyasov et al., 2018).

Type locality. Japan.

Distribution. This subspecies occurs in North Korea, South Korea, Primorsky Krai and Khabarovsk Krai of Russia, Northeast China and Japan.

Diagnosis. This subspecies is slightly larger, with two body color types: yellow-gray and black-gray, and a high Cubital index. About 1/3 of the workers in northeast China have a small protruding vein on the radiomedial crossvein of the forewing. Additional morphometric values are provided in Supplementary Table S5.

Evolutionary synopsis. The present study confirms that populations found in Northeastern China, South and North Korea, and Japan all belong to the same subspecies *Apis cerana japonica*, which diverged from the common ancestor after *A. c. hainana*. The current distributional range of *A. c. japonica* suggests that one of the historical range expansions of Mainland *A. cerana* had reached the far north, including at least the current northern range limit of the species. The descendant of this frontier population became adapted to the temperate climate, which is characterized by prolonged winter and extreme annual temperature variations. The mechanism by which it maintains its separation from the central subspecies, *A. c. cerana*, however, remains unknown. Hybrid individuals are detected in the region where they meet (Liaoning), but *A. c. japonica* have not shown major introgression signals derived from the invasion of the central population. Given that there are no obvious major geographic barriers in between the ranges of the two subspecies, it might be possible that other mechanisms have played a role in fortifying their genetic distinctiveness (e.g., genomic structural variation, differences in fitness).

***Apis cerana skorikovi* Engel, 1999 (Engel, 1999)**

Apis cerana skorikovi Maa, 1944 (Maa, 1944). Nomen nudum. Maa proposed this subspecies in 1944 but did not provide sufficient description. Engel (1999) treated this proposal as Nomen nudum but retained the original subspecies name in recognition of Maa's contribution (Engel, 1999).

Apis cerana Himalayan race, Ruttner, 1988 (Ruttner, 1988).

Apis cerana Xizang race, Peng et al., 1989 (Peng et al., 1989).

Apis cerana himalaya Smith, 1991 (Smith, 1992). Nomen nudum.

Type locality. worker; Nepal: KoPasi near Panauti, March 19th, 1982, B. A. Underwood (deposited in the Cornell University Insect Collection, Ithaca)

Distribution. This subspecies likely has a constrained distribution in Tibet at the elevation range of 2000~4000 m, along the Palongzangbu river valley, including Bomi County (Supplementary Figure S10).

Diagnosis. This subspecies is fairly gray in overall body coloration, with some areas being slightly more yellowish while others are closer to black (abdominal T4 is entirely black). The abdominal T3 is characterized by a pair of yellow spots. Distinct white hairs are evenly distributed on the thoracic and abdominal terga, contributing to the overall grayish appearance. *Apis cerana skorikovi*, like *A. c. abansis*, also differs from *A. c. cerana* in the presence of a dark scutellum and slightly larger body size (forewing length of $846.60 \pm 28.61 \mu\text{m}$). Additional morphometric values are provided in Supplementary Table S5.

Evolutionary synopsis. The historical literature on the Tibet subspecies has been confusing. The subspecies defined here has a very strict range, confined only to the Palongzangbu River Valley. Introgression is not common among the samples examined so far, likely a result of the relatively complete geographic barrier formed by surrounding high mountain ridges.

***Apis cerana guidensis* Qiu, Liu & Zhou, ssp. nov.**

Supplementary Figure S11: A~C

ZooBank taxon LSID: D0FE86E8-00F2-46EA-9157-6CD201117921

Type locality. Holotype: worker; China: eastern Qinghai Province, Guide Co., July 16th, 2020 (deposited in the Department of Entomology, China Agricultural University, Beijing, China). Paratypes: Three workers, same collection data as the holotype (deposited in the Department of Entomology, China Agricultural University, Beijing, China).

Distribution. This subspecies is confined to the upper reaches of the Yellow River Valley in Qinghai Province, China, including Guide and Xunhua Counties (Supplementary Figure S10).

Diagnosis. *Apis cerana guidensis* is of a large body size, with a forewing length of $864.07 \pm 17.98 \mu\text{m}$. The body color is relatively dark, with a dark metasoma and legs. The Cubital index is high. Additional morphometric values are provided in Supplementary Table S5.

Evolutionary synopsis. This newly defined subspecies is confined to the upper reaches of the Yellow River valley, representing the most northern ranging Qinghai-Xizang (Tibet) valley subspecies. But the order of divergence among the four valley subspecies remains unclear. *Apis cerana guidensis* and *A. c. abansis* share several morphological traits, such as large body size and darkened coloration, reflecting likely convergent evolution of the mountain bees.

***Apis cerana abansis* Yang & Kuang, 1986 (Yang et al., 1986)**

Supplementary Figure S11: D~F

Apis cerana heimifera Engel, 1999 (Engel, 1999)

Apis cerana abaensis Hepburn, Smith, Radloff & Otis, 2001 (Hepburn et al., 2001).
Nomen nudum.

Type locality. Holotype: worker; China: northern Sichuan Province, Jinchuan Co., July 1976 (deposited in the Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing, China). Paratypes: worker; China: northern Sichuan

Province, Maerkang Co., September 1982 (deposited in the Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing, China). The holotype and paratype specimens are confirmed missing. We thus establish neotypes in the present work.

Neotype: worker; China: northern Sichuan Province, Maerkang Co., April 8th, 2021 (deposited in the Department of Entomology, China Agricultural University, Beijing, China). Paratypes: Three workers, same collection data as holotype (deposited in the Department of Entomology, China Agricultural University, Beijing, China).

Distribution. This subspecies occurs along the Dadu River valley in Sichuan Province, including Barkam, Xiaojin and Jinchuan Counties (Supplementary Figure S10).

Diagnosis. This is the largest *Apis cerana* race in China, with a forewing length of 865.00 ± 8.79 μm . The workers have an overall dark appearance, with a dark metasoma, scutellum, and legs. The orange bands on abdominal terga III-V are very narrow or completely absent. Additional morphometric values are provided in Supplementary Table S5.

Evolutionary synopsis. The subspecies status of *Apis cerana abansis* has been proposed earlier. However, its specific range was considered to cover a large transition zone of the Sichuan basin to the Qinghai-Xizang (Tibet) Plateau (Peng et al., 1989), expanding several valleys of the eastern and northeastern Qinghai-Xizang (Tibet) Plateau. Based on genetic structure and evolutionary independence, we now understand that each of these valley populations are a result of parallel expansion and adaptation events without direct introgression in between, therefore they do not belong to the same subspecies. The previous understanding of the cross-valley range was probably caused by convergent morphology of mountain populations.

***Apis cerana kashmirensis* Qiu, Liu & Zhou, ssp. nov.**

Supplementary Figure S11: G~I

ZooBank taxon LSID: 0E27DCB7-D0C2-41CF-BCC6-6929BFED04C3

Type locality. Holotype: worker; India: Harwan Srinagar Jammu & Kashmir, October 25th, 2018 (deposited in the Department of Entomology, China Agricultural University, Beijing, China). Paratypes: Three workers, same collection data as holotype (deposited in the Department of Entomology, China Agricultural University, Beijing, China); five workers, same collection data as the holotype (deposited in the Department of Zoology, BGSB University Rajouri, J&K, India); three workers, same collection data as holotype (deposited in the Zoological Survey of India, Kolkata, India).

Distribution. This subspecies occurs mainly in the Kashmir Valley (Supplementary Figure S10).

Diagnosis. *Apis cerana kashmirensis* is of large body size, with a forewing length of 846.93 ± 13.25 μm . The body color is dark, with a dark metasoma and legs. Additional morphometric values are provided in Supplementary Table S5.

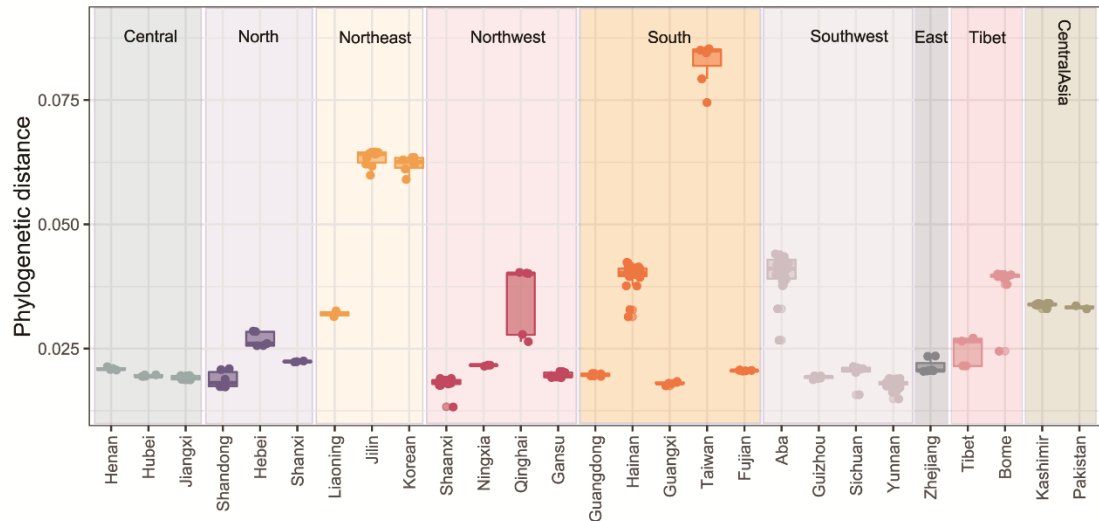
Evolutionary synopsis. This newly defined subspecies is largely confined in the Vale of Kashmir, a deep asymmetrical basin lying between the Pir Panjal Range and the western end of the Great Himalayas, with an average elevation of 1,620 m, covering the Kashmir region and northern Pakistan. Gene introgression with *Apis cerana cerana*

has not been detected, but additional sampling from the overlapping range might better reveal the genetic structure of the subspecies. Note that populations from both Gyirong and Xigazê Counties, China, the two valleys carved into the southern slope of the Himalaya, showed admixtures between *A. c. kashmirensis* and *A. c. cerana*. These results suggested that the ancestral population of *A. c. kashmirensis* might have a larger historical range covering the western to at least the mid-range of the Himalayas.

References

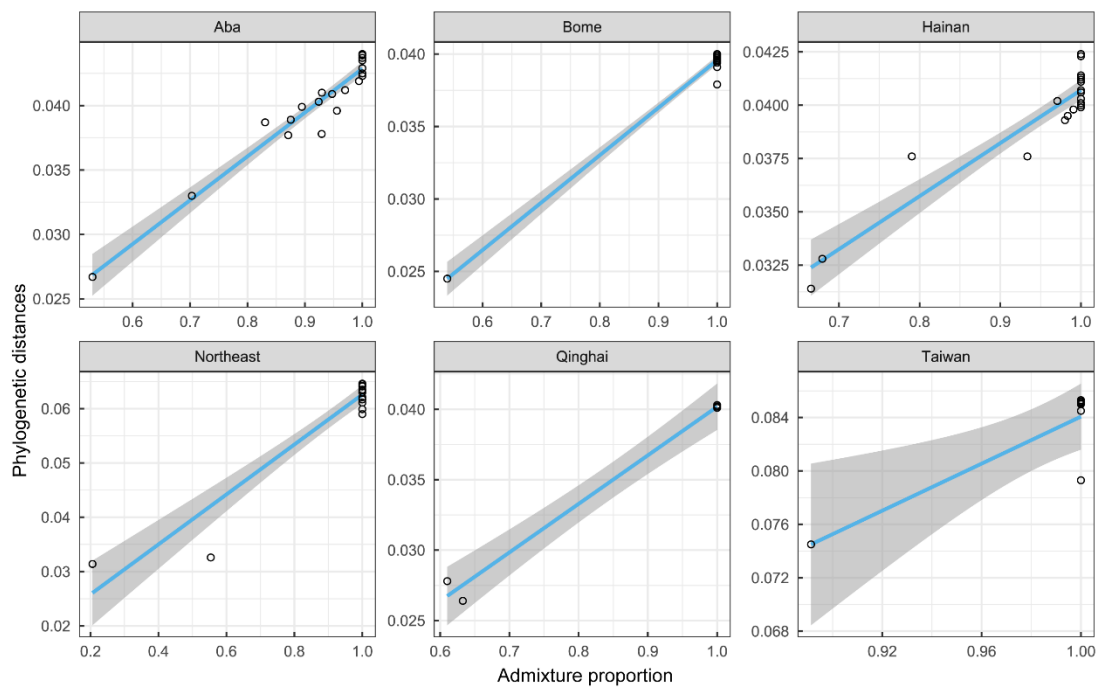
- Chen YC. 1993. Apiculture of China. Beijing: China Agricultural Publication. (in Chinese)
- Engel MS. 1999. The taxonomy of recent and fossil honey bees (Hymenoptera: Apidae; *Apis*). *Journal of Hymenoptera Research*, **8**(2): 165–196.
- Fabricius. 1792–1793. Eutomological Systematica.
- Hepburn HR, Smith DR, Radloff SE, et al. 2001. Intraspecific categories of *Apis cerana*: morphometric, allozymal and mtDNA diversity. *Apidologie*, **32**(1): 3–23.
- Ilyasov RA, Han GY, Lee ML, et al. 2021. Genetic properties and evolution of Asian honey bee *Apis cerana ussuriensis* from primorsky Krai, Russia. *Russian Journal of Genetics*, **57**(5): 568–581.
- Ilyasov RA, Park J, Takahashi J, et al. 2018. Phylogenetic uniqueness of Honeybee *Apis cerana* from the Korean peninsula inferred from the mitochondrial, nuclear, and morphological Data. *Journal of Apicultural Science*, **62**(2): 189–214.
- Ji YK, Li XG, Ji T, et al. 2020. Gene reuse facilitates rapid radiation and independent adaptation to diverse habitats in the Asian honeybee. *Science Advances*, **6**(51): eabd3590.
- Kuang HO, Kuang BY, Wang JP, et al. 2007. The type locality investigation of *Apis cerana cerana* Fabricius. In: Proceedings of the 3rd China Animal Husbandry Science and Technology Forum. 477–479. (in Chinese)
- Maa TC. 1944. On the classification and phylogeny of the Chinese honeybees. *Entomol. Shaowuana*, **1**: 4–5. (in Chinese)
- National Animal Genetic Resources Committee. 2011. Animal Genetic Resources in China: Bees. Beijing: China Agriculture Press. (in Chinese)
- Peng YS, Nasr ME, Locke SJ. 1989. Geographical races of *Apis cerana* Fabricius in China and their distribution. Review of recent Chinese publications and a preliminary statistical analysis. *Apidologie*, **20**(1): 9–20.
- Radoszkowski O. 1887. Hyménoptères de Korée. *Horae Societatis Entomologicae Rossicae*, **21**: 428–436.
- Ruttner F. 1988. Biogeography and Taxonomy of Honeybees. Berlin: Springer.
- Smith DR. 1992. Mitochondrial DNA and honey bee biogeography. In: Smith DR. Diversity in the Genus *Apis*. Boulder: Westview Press, 131–176.
- Yang GH, Xu SY, Kuang BY, et al. 1986. The distribution of the genus *Apis cerana* Fab. in China and some of its subspecies. *Journal of Yunnan Agricultural University*, **1**(1): 89–92. (in Chinese)

- Zhao WZ, Wang M, Liu YQ, et al. 2017. Phylogeography of *Apis cerana* populations on Hainan island and southern mainland China revealed by microsatellite polymorphism and mitochondrial DNA. *Apidologie*, **48**(1): 63–74.
- Zhuang DA. 1989. New subspecies of *Apis cerana*. *Southwest China Journal of Agricultural Sciences*, **2**(4): 61–65. (in Chinese)



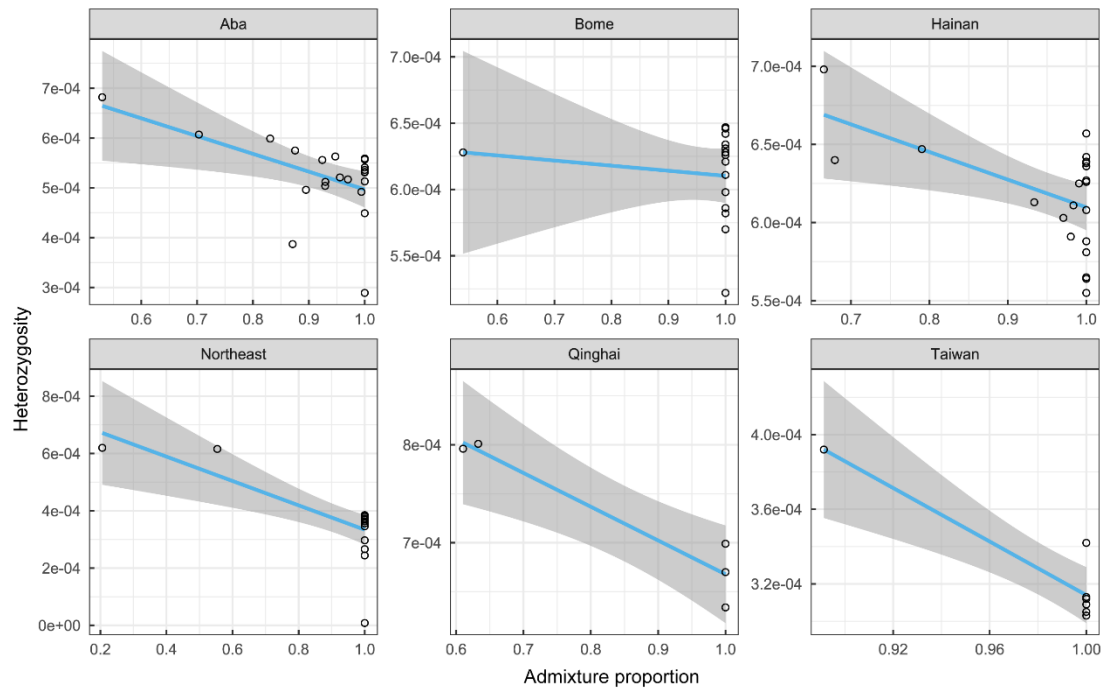
Supplementary Figure S1 Phylogenetic distance (PD) between geographic populations of mainland *A. cerana* and root lineage (Malaysia)

Local populations were ordered and grouped by geographic affinity. For each local population, PD was calculated according to a formula: $PD = \text{avg}(d_{xy}) - \text{avg}(d_{xz})$, where avg is the average, d_{xy} is the distance of an individual from the root lineage, and d_{xz} is the distance between individuals of the same local population.



Supplementary Figure S2 Correlation between PD and admixture proportion (AP)

PD is defined in Supplementary Figure S1 and AP represents proportion of an individual's alleles with ancestry in corresponding local population. Note, no individual in the Kashmir and Pakistan population showed an obvious low AP, and thus was not included in the current figure.

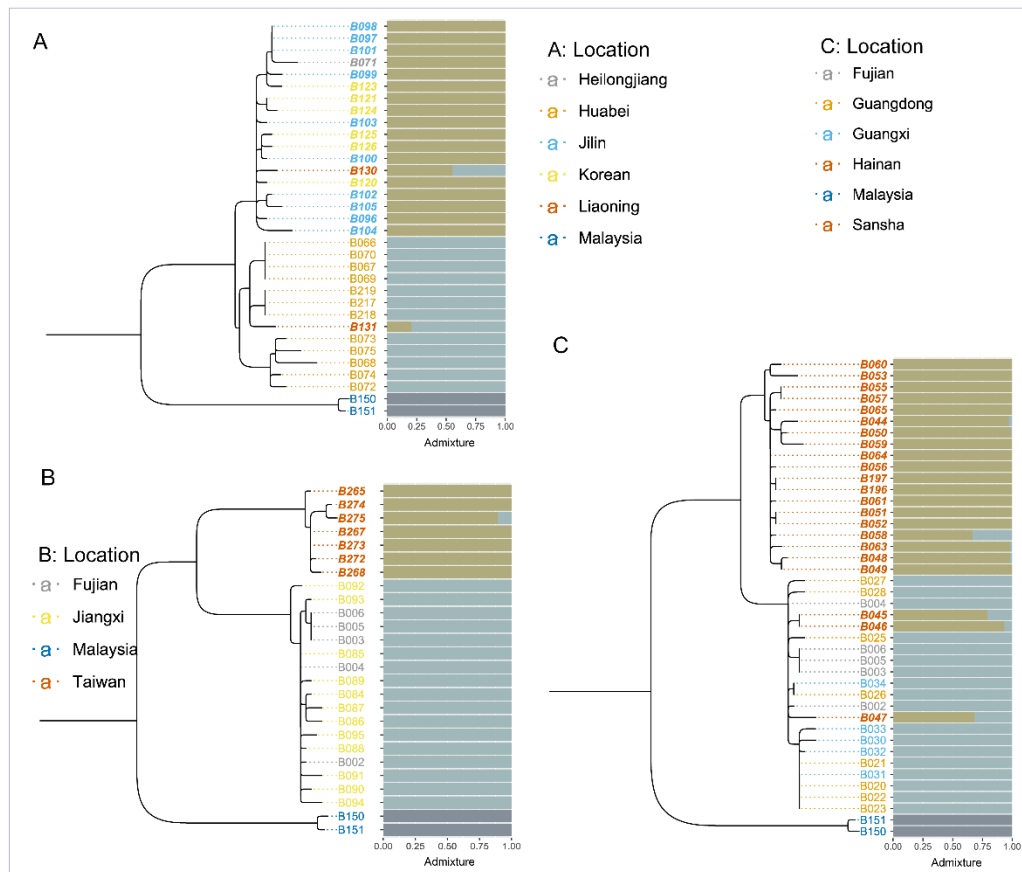


Supplementary Figure S3 Correlation between whole-genome heterozygosity and AP

AP is defined in Supplementary Figure S2.

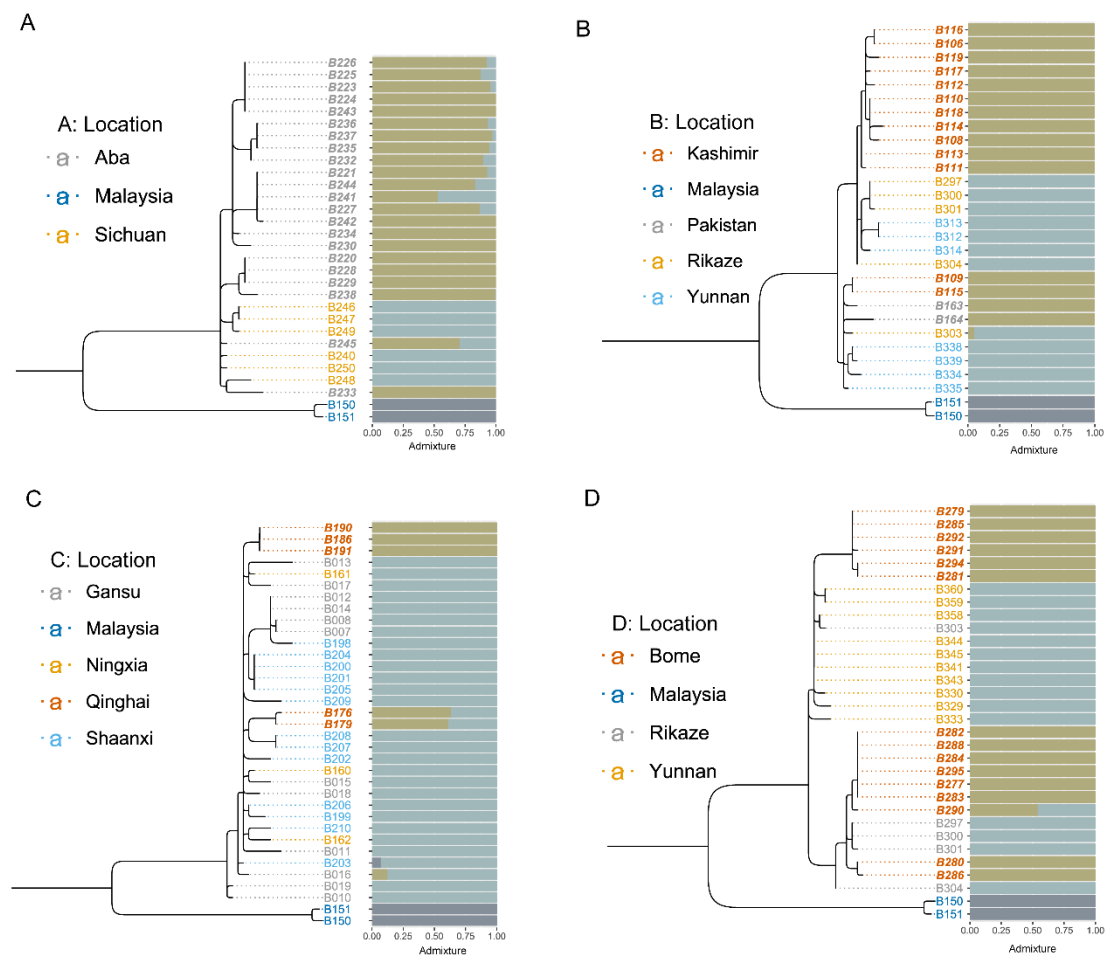
Supplementary Figure S4 Phylogenetic relationships of subspecies of mainland *A. cerana* and Malaysian lineage based on mitochondrial genomes

Supplementary Figure S4 is listed as a separate PDF file due to its large size.



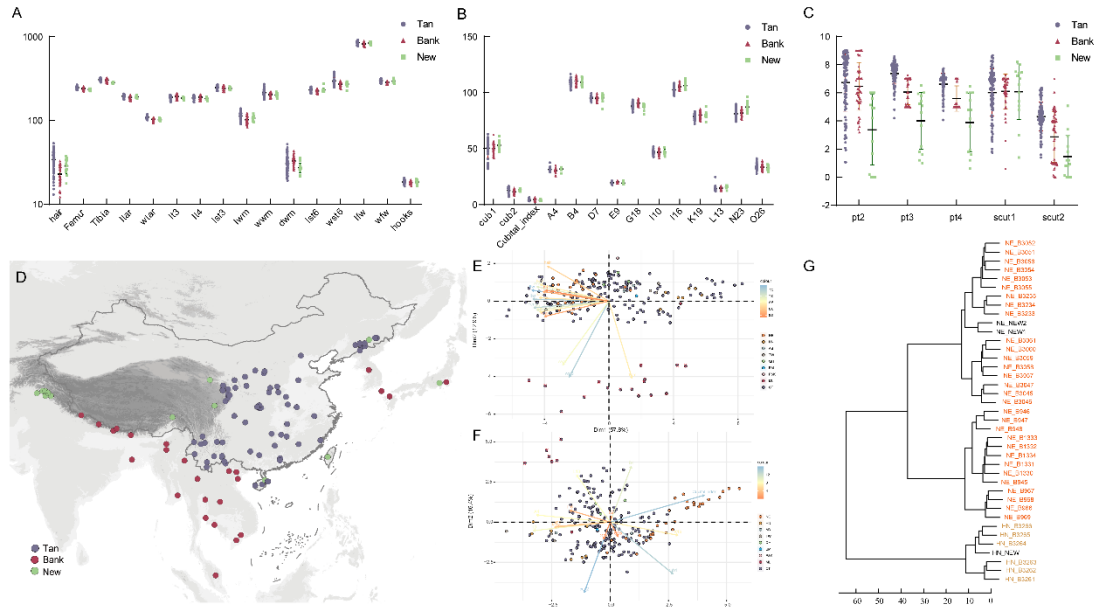
Supplementary Figure S5 Phylogenetic tree based on mitochondrial genomes and admixture estimations for Northeast (A), Taiwan (B), and Hainan (C) populations

Tip labels in bold and italics indicate individuals from focal population, whereas individuals from peripheral and Malaysia populations are labeled in plain text. Text colors correspond to locations indicated by legends of each panel. To obtain a simplified view for better illustration, we included only two representative individuals from the Malaysia population.



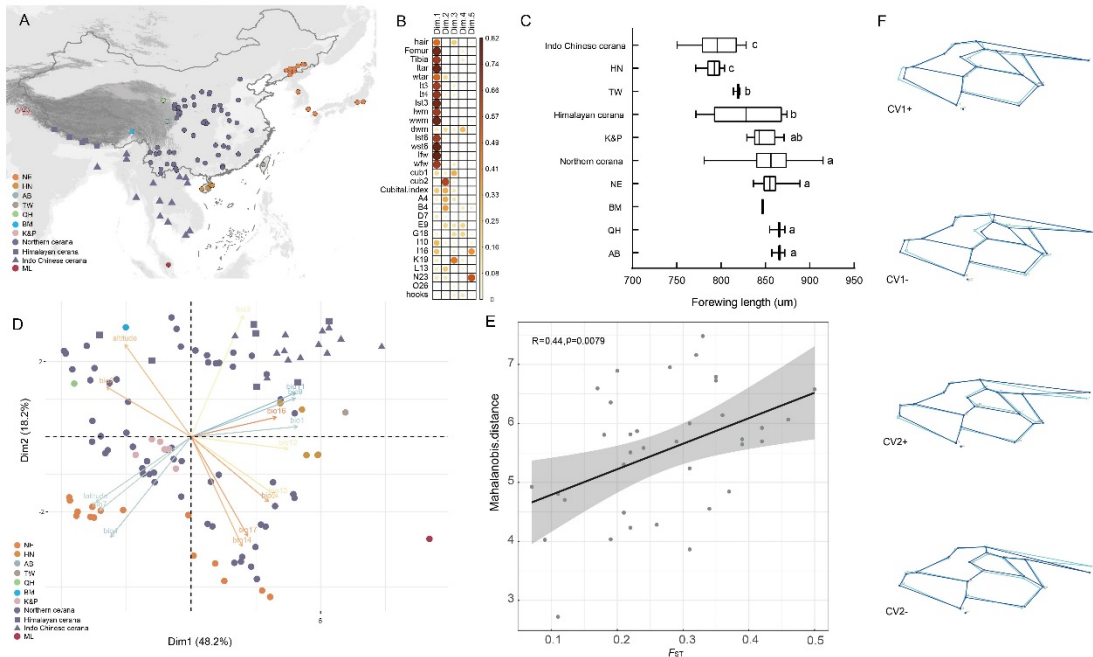
Supplementary Figure S6 Phylogenetic tree based on mitochondrial genomes and admixture estimations for Aba (A), Pakistan & Kashmir (B), Qinghai (C), and Bomi (D) populations

Figure legends follow those in Supplementary Figure S9.



Supplementary Figure S7 Analysis of morphometric differences in legacy and new samples

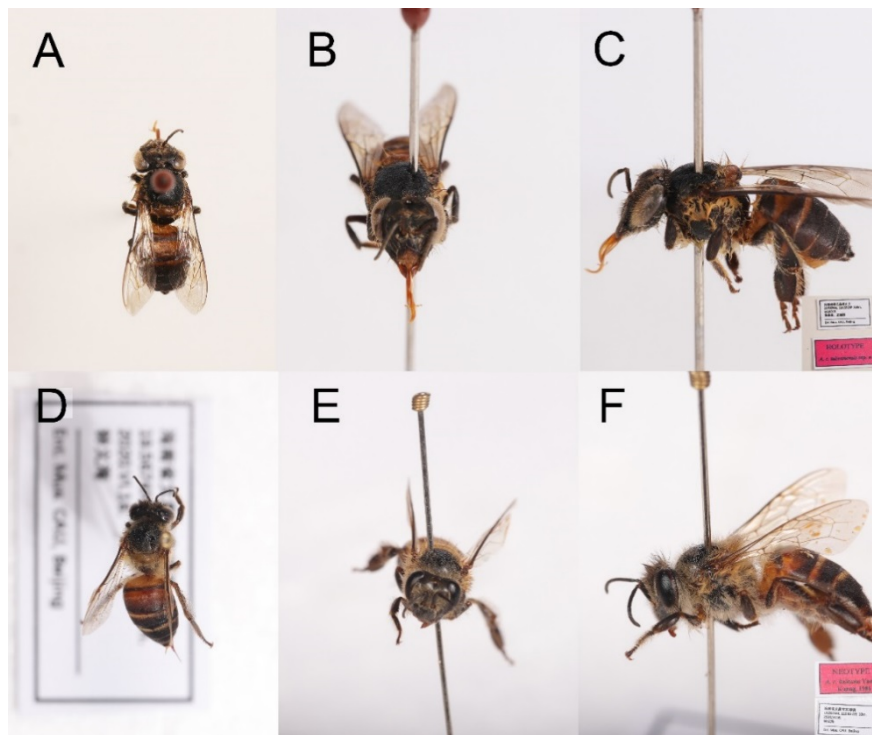
A: Scatterplots of size-related traits and hook numbers obtained from legacy (Tan et al., 2006 and Morphometric Bank) and new samples from present study. B: Scatterplots of wing-related trait values from legacy and new samples. C: Scatterplots of color-related trait values from legacy and new samples. D: Sample distribution in different datasets. E: Principal component plot of all samples based on 15 size-related traits and hook numbers. “X” represents outliers removed from subsequent analyses. F: Principal component plot of all samples based on 14 wing-related traits. “X” represents outliers removed from subsequent analyses. G: Cluster plots of legacy and new samples from Hainan Island and NE populations. New samples measured in present study (“HN_NEW” and “NE_NEW”) are in black.



Supplementary Figure S8 Morphological and bioclimatic variables of different subspecies

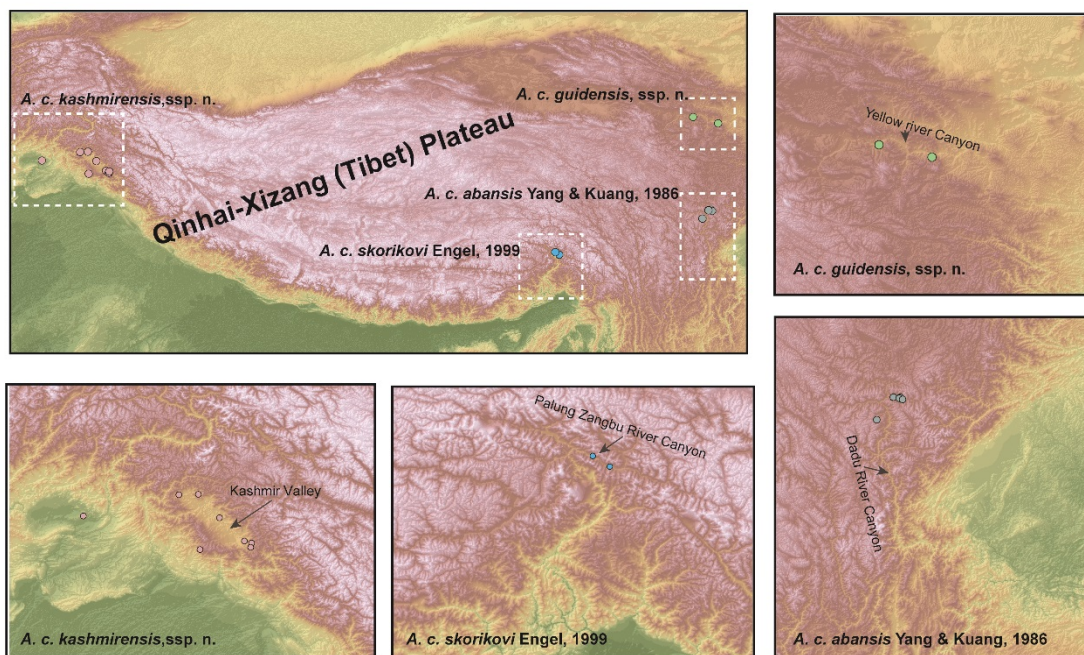
A: Collection sites for samples used in morphological analysis. B: Contribution of each morphological variable to

first five principal components revealed in PCA. Size and color of each circle represent quality of representation of each variable, with darker or larger circles indicating larger contribution from morphological trait to respective principal component. C: Boxplot of *A. cerana* in forewing length across geographical populations. D: PCA of 21 bioclimatic variables for all sampling sites. E: Correlation between F_{ST} and Mahalanobis distance between each pair of groups. F: Wireframes represent *A. cerana* wing configuration from positive and negative scores on first and second canonical variate axes (CV1 and CV2). Lines in light blue represent average configuration of wing. Lines in navy blue represent first and second canonical variables.

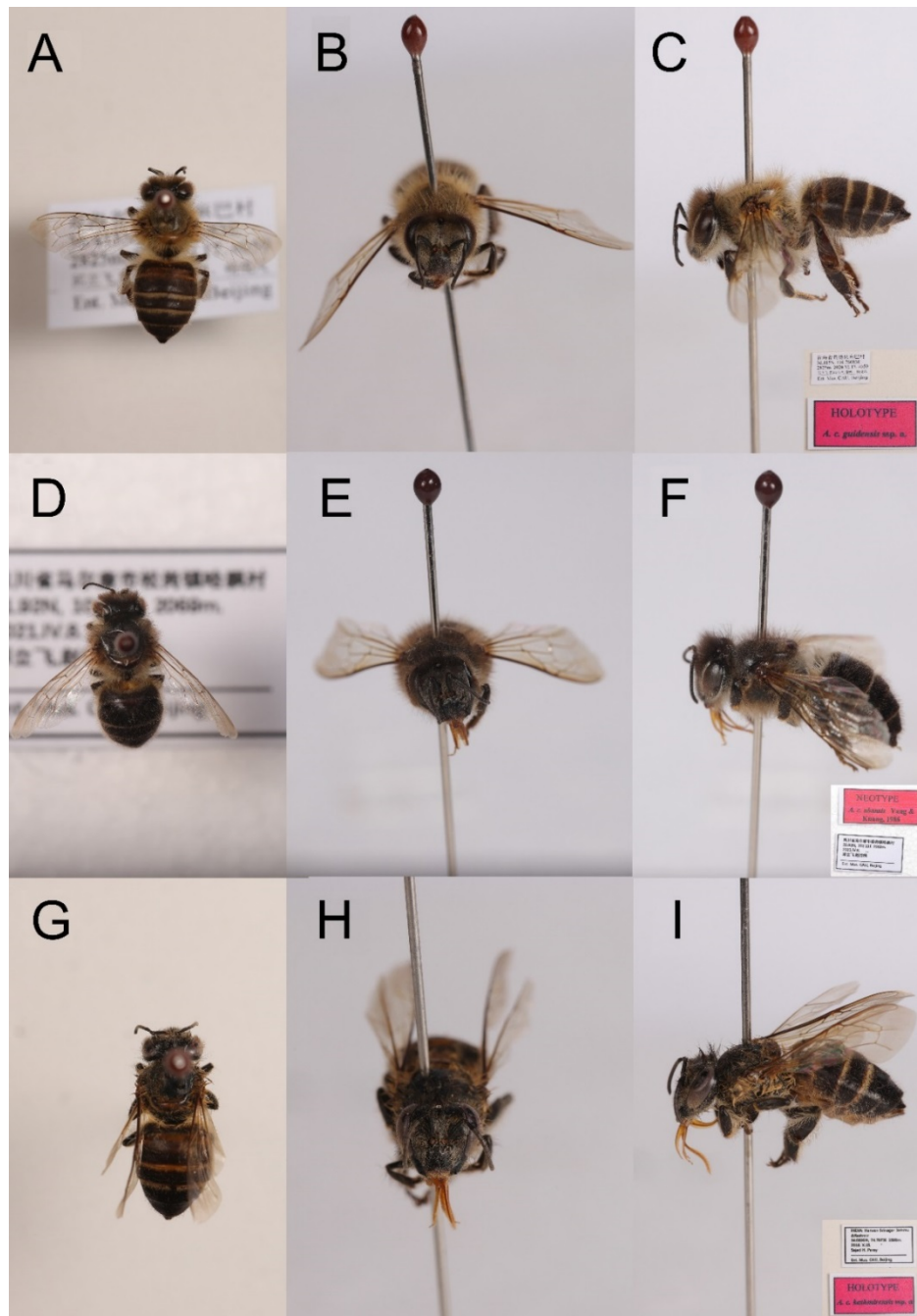


Supplementary Figure S9 Holotypes of *A. c. taiwanensis* and neotypes of *A. c. hainana*

A-C: *A. taiwanensis* **ssp. n.**; D-F: *A. c. hainana* Yang & Kuang, 1986.



Supplementary Figure S10 Locations of four high-land subspecies distributed in canyons and valleys around the edge of the Qinghai-Xizang (Tibet) Plateau



Supplementary Figure S11 Holotypes of *A. c. guidensis* and *A. c. kashmirensis* and neotypes of *A. c. abansis* Yang & Kuang, 1986

A-C: *A. c. guidensis* ssp. n.; D-F: *A. c. abansis* Yang & Kuang, 1986; G-I: *A. c. kashmirensis* ssp. n.

Supplementary Table S1. Information of specimens used for genome re-sequences.

Supplementary Table S2. Information of specimens used for standard morphometrics.

Supplementary Table S3. Information of specimens used for geometric morphometrics.

Supplementary Tables S1-S3 are listed as a separate excel file due to their large size.

Supplementary Table S4 Mahalanobis distances among groups and *P*-values from permutation tests (10 000 iterations)

	CT	NE	HN	K&P	AB	ML	QH	BM	TW
CT	~	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
NE	4.0369	~	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
HN	2.7218	3.8646	~	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001
K&P	4.9254	6.9528	6.3556	~	<.0001	<.0001	<.0001	<.0001	<.0001
AB	4.7048	5.9997	5.5844	6.8927	~	<.0001	<.0001	<.0001	<.0001
ML	4.2311	5.9241	4.5523	7.1594	6.7909	~	<.0001	<.0001	<.0001
QH	4.0262	5.6921	5.3021	5.8081	5.5095	6.7269	~	<.0001	<.0001
BM	4.8059	5.238	5.8155	6.5938	5.8678	7.4823	4.487	~	<.0001
TW	4.2826	6.0654	4.8452	6.1402	5.7312	6.5799	5.696	5.6455	~

Supplementary Table S5 Means and standard deviations (SD) of 30 morphometric character measurements (μm or angles in degrees) of eight *A. cerana* worker honeybee subspecies

	<i>A.c. cerana</i>	<i>A.c. japonica</i>	<i>A.c. hainana</i>	<i>A.c. abaensis</i>	<i>A.c. taiwanensis</i>	<i>A.c. guidensis</i>	<i>A.c. skorikovi</i>	<i>A.c. kashmirensis</i>
hair	32.86 $\pm$ 9.59	33.09 $\pm$ 9.76	21.1 $\pm$ 2.55	31.33 $\pm$ 6.59	24.6 $\pm$ 4.31	35.53 $\pm$ 4.07	31.07 $\pm$ 4.88	29.79 $\pm$ 5.67
Femur	244.75 $\pm$ 8.85	244.31 $\pm$ 8.92	236.32 $\pm$ 5.17	227.53 $\pm$ 9.44	235.73 $\pm$ 7.68	237.53 $\pm$ 8.32	236.07 $\pm$ 10.89	228.79 $\pm$ 7.71
Tibia	306.56 $\pm$ 11.51	305.78 $\pm$ 11.64	292.77 $\pm$ 6.56	286.07 $\pm$ 5.7	279.47 $\pm$ 8.39	287.93 $\pm$ 8.11	286.13 $\pm$ 13.26	280.5 $\pm$ 9.83
ltar	192.87 $\pm$ 8.72	192.15 $\pm$ 8.5	184.48 $\pm$ 3.97	191.8 $\pm$ 7.78	191.8 $\pm$ 10.02	196.13 $\pm$ 8.29	188.67 $\pm$ 9.2	190.5 $\pm$ 11.31
wtar	107.26 $\pm$ 5.03	106.5 $\pm$ 4.61	105.1 $\pm$ 3.43	102.93 $\pm$ 5.35	104.47 $\pm$ 5.21	99 $\pm$ 3.51	104.67 $\pm$ 5.34	102.57 $\pm$ 5
lt3	191.17 $\pm$ 9.28	191.03 $\pm$ 9.79	171.1 $\pm$ 2.3	186.07 $\pm$ 6.67	183.53 $\pm$ 6.53	190.55 $\pm$ 6.83	182.33 $\pm$ 8.11	186.57 $\pm$ 3.52
lt4	187.62 $\pm$ 8.95	187.38 $\pm$ 10.4	169.32 $\pm$ 2.32	181.47 $\pm$ 6.12	182.47 $\pm$ 5.77	188.4 $\pm$ 6.91	178.07 $\pm$ 8.4	187.21 $\pm$ 4.35
lst3	244.49 $\pm$ 11.43	244.11 $\pm$ 11.82	230.4 $\pm$ 3.06	238.2 $\pm$ 9.44	240.07 $\pm$ 7.77	242.33 $\pm$ 9.04	244.8 $\pm$ 9.63	240.5 $\pm$ 9.53
lwm	110.16 $\pm$ 11.49	108.48 $\pm$ 9.96	93.71 $\pm$ 2.03	107.07 $\pm$ 7.97	100.27 $\pm$ 7.7	107.87 $\pm$ 8.98	108.93 $\pm$ 7.03	120.79 $\pm$ 7.13
wwm	207.99 $\pm$ 11.22	208.88 $\pm$ 11.67	186.63 $\pm$ 5.72	208.13 $\pm$ 15.86	196.6 $\pm$ 8.71	203.53 $\pm$ 14.36	198.13 $\pm$ 7.39	198.86 $\pm$ 8.47
dwm	28.56 $\pm$ 5.72	27.66 $\pm$ 5.72	34.1 $\pm$ 2.17	27.87 $\pm$ 3.83	25.47 $\pm$ 5.21	25.13 $\pm$ 5.53	27.87 $\pm$ 5.67	27.71 $\pm$ 5.7
lst6	230.2 $\pm$ 10.03	229.53 $\pm$ 10.19	219.21 $\pm$ 3.9	232.33 $\pm$ 7.91	223.67 $\pm$ 3.54	226 $\pm$ 7.74	231.27 $\pm$ 9.6	275.07 $\pm$ 6.67
wst6	280.38 $\pm$ 15.13	279.74 $\pm$ 15.29	256.86 $\pm$ 5.64	279.07 $\pm$ 8.94	264.13 $\pm$ 8.27	279 $\pm$ 9.8	274.93 $\pm$ 11.8	225.43 $\pm$ 4.6
lfw	845.16 $\pm$ 34.87	841.59 $\pm$ 33.77	791.56 $\pm$ 10.12	865 $\pm$ 8.79	816.93 $\pm$ 9.12	864.07 $\pm$ 17.98	846.6 $\pm$ 28.61	846.93 $\pm$ 13.25
wfw	297.06 $\pm$ 14.03	295.46 $\pm$ 13.64	280.09 $\pm$ 4.59	297.93 $\pm$ 7.12	284.07 $\pm$ 3.69	311.33 $\pm$ 20.78	292.27 $\pm$ 10.85	295.64 $\pm$ 6.44
cub1	50.06 $\pm$ 7.99	49.08 $\pm$ 8.44	47.84 $\pm$ 3.6	55.2 $\pm$ 3.69	49.73 $\pm$ 2.76	53.13 $\pm$ 2.61	49.87 $\pm$ 4.76	53.07 $\pm$ 3.52
cub2	12.03 $\pm$ 2.13	11.66 $\pm$ 2.07	13.69 $\pm$ 0.71	14.13 $\pm$ 1.96	14.07 $\pm$ 1.44	12.13 $\pm$ 1.68	14.4 $\pm$ 1.76	14.64 $\pm$ 1.86
Cubital index	4.23 $\pm$ 0.73	4.27 $\pm$ 0.74	3.5 $\pm$ 0.28	3.91 $\pm$ 0.73	3.58 $\pm$ 0.48	4.38 $\pm$ 1.55	3.46 $\pm$ 0.66	3.62 $\pm$ 1.89
A4	31.09 $\pm$ 1.36	31.34 $\pm$ 2.06	32.06 $\pm$ 0.68	32.75 $\pm$ 1.42	32.2 $\pm$ 1.58	31.34 $\pm$ 1.4	32.35 $\pm$ 1.75	31.86 $\pm$ 1.59
B4	109.73 $\pm$ 2.77	110 $\pm$ 2.94	105.83 $\pm$ 2.81	106.65 $\pm$ 2.62	108.69 $\pm$ 5.09	113.36 $\pm$ 3.61	109.51 $\pm$ 2.5	108.6 $\pm$ 4.14
D7	94.85 $\pm$ 1.59	94.67 $\pm$ 1.51	95.54 $\pm$ 1.75	90.82 $\pm$ 7.59	96.5 $\pm$ 1.37	95.93 $\pm$ 2.18	97.34 $\pm$ 4.1	97.98 $\pm$ 1.98
E9	19.94 $\pm$ 0.87	20.09 $\pm$ 1.25	19.52 $\pm$ 0.63	19.18 $\pm$ 0.74	19.57 $\pm$ 1.03	19.59 $\pm$ 1.05	18.09 $\pm$ 1.04	18.6 $\pm$ 4.14
G18	88.75 $\pm$ 2.57	88.45 $\pm$ 2.47	89.59 $\pm$ 3.28	87.37 $\pm$ 1.96	84.52 $\pm$ 2.68	87.78 $\pm$ 2.24	88.02 $\pm$ 2.66	80.82 $\pm$ 3.11
J10	46.69 $\pm$ 1.89	46.7 $\pm$ 2.02	49.64 $\pm$ 1.48	46.61 $\pm$ 2.87	51.12 $\pm$ 3.55	46.79 $\pm$ 2.45	43.81 $\pm$ 2.3	45.76 $\pm$ 3.54
J16	103.38 $\pm$ 2.33	103.61 $\pm$ 2.14	104.53 $\pm$ 1.28	106.93 $\pm$ 4.17	107.37 $\pm$ 4.14	105.87 $\pm$ 3.18	102.84 $\pm$ 2.29	104.1 $\pm$ 4.1
K19	79.05 $\pm$ 2.46	78.97 $\pm$ 2.55	78.36 $\pm$ 3.84	81.58 $\pm$ 1.89	78.6 $\pm$ 3.59	79.9 $\pm$ 2.15	80.4 $\pm$ 2.78	87.66 $\pm$ 4.46
L13	14.38 $\pm$ 1.03	14.87 $\pm$ 2.35	14.71 $\pm$ 0.74	16.34 $\pm$ 1.05	14.72 $\pm$ 1.23	16.56 $\pm$ 1.09	15.48 $\pm$ 0.93	16.87 $\pm$ 1.86
N23	81.94 $\pm$ 2.57	82 $\pm$ 2.44	84.64 $\pm$ 1.35	89.5 $\pm$ 3.79	89.19 $\pm$ 4.83	83.78 $\pm$ 3.76	82.62 $\pm$ 2.58	83.87 $\pm$ 4.77
O26	33.43 $\pm$ 2.1	33.23 $\pm$ 2.02	32.62 $\pm$ 0.86	31.28 $\pm$ 3.1	33.72 $\pm$ 3.44	31.44 $\pm$ 2.64	29.34 $\pm$ 2.74	35.17 $\pm$ 2.4
hooks	18.39 $\pm$ 0.8	17.23 $\pm$ 4.79	18.94 $\pm$ 0.73	18.93 $\pm$ 1.39	18.47 $\pm$ 1.46	17.8 $\pm$ 1.37	16.73 $\pm$ 1.39	18.36 $\pm$ 1.55