Gene reuse facilitates rapid radiation and independent adaptation to diverse habitats in the Asian honeybee

Yongkun Ji1*, Xingan Li2*, Ting Ji3†, Junbo Tang4‡, Lifei Qiu1, Jiahui Hu1, Jiangxing Dong1, Shiqi Luo1, Shanlin Liu1,5, Paul B. Frandsen6,7, Xuguo Zhou8, Sajad H. Parey9, Lianming Li10, Qingsheng Niu2†, Xin Zhou11

Animals with recent shared ancestry frequently adapt in parallel to new but similar habitats, a process often underlined by repeated selection of the same genes. Yet, in contrast, few examples have demonstrated the significance of gene reuse in colonization of multiple disparate habitats. By analyzing 343 genomes of the widespread Asian honeybee, *Apis cerana*, we showed that multiple peripheral subspecies radiated from a central ancestral population and adapted independently to diverse habitats. We found strong evidence of gene reuse in the Leucokinin receptor (Lkr), which was repeatedly selected in almost all peripheral subspecies. Differential expression and RNA interference knockdown revealed the role of Lkr in influencing foraging labor division, suggesting that Lkr facilitates collective tendency for pollen/nectar collection as an adaptation to floral changes. Our results suggest that honeybees may accommodate diverse floral shifts during rapid radiation through fine-tuning individual foraging tendency, a seemingly complex process accomplished by gene reuse.

INTRODUCTION

Genetic adaptation to novel habitats is a key process in speciation and diversification. When multiple populations rapidly radiate into new but similar habitats, natural selection has been shown to favor just a few genes responsible for increasing fitness (1) [e.g., sticklebacks radiated from marine to freshwater (2)]. However, it is less well known whether repeated selection on the same genes (a.k.a. gene reuse) can also facilitate adaptation to multiple but differing habitats. Honeybees are ideal models to address this question because they are widespread and founding populations often encounter diverse food resources, i.e., flowers with drastically different traits and phylogeny from those with which the bees are accustomed. Assuming honeybees use the same mechanism to explore varied food combinations, it is plausible that the same underlying genes may be under repeated selection among independently radiating populations.

Compared to the western honeybee (*Apis mellifera*), the Asian honeybee (*Apis cerana*) is less influenced by confounding signals derived from recent genetic admixtures caused by artificial breeding and beekeeping (3, 4). This is due to traditional beekeeping practices, which are exercised across its natural range, in which the Asian honeybee maintains a semiferal nature allowing populations to be less affected by admixture with domesticated colonies (4). This characteristic renders it an exemplar model system to investigate the genetic mechanisms underlying divergent adaptation.

The Asian honeybee exhibits the largest native range among all Asiatic *Apis* species and is one of Asia’s most important native pollinators within both natural and agricultural cropping systems (4). Through long-term natural selection, it has become well adapted to the local climatic conditions and flowering plants and thus plays irreplaceable and vitally important ecological roles relative to non-native honeybees (4). The species is ecologically diverse; it can survive both the harsh winters of northern China and the Russian Far East, as well as the hot and humid climates of tropical islands (4). In addition, it has penetrated the high-altitude Qinghai-Tibet (Q-T) Plateau through alpine canyons (5). This wide distribution pattern suggests that the Asian honeybee is capable of adjusting its foraging strategies to various floras and flowering phenologies, which are primarily shaped by local climates. Despite this, the Asian honeybee is facing unprecedented population and biodiversity decline, causing direct damage to pollination services in Asia and exacerbating the food deficit for more than half of the world’s humans (4, 6). Accelerating habitat changes due to climate change and anthropogenic influences are considered to be important causes of this decline (7). Together, the outsized importance of the Asian honeybee and the potential devastating impacts of its decline call for a better understanding of its genetic diversity, including its population structure, demographic history, and, importantly, mechanisms underlying its successful adaptation to diverse habitats.

A previous study based on mitochondrial DNA identified four main genetic lineages within the Asian honeybee, including mainland Asia, Sundaland, Palawan, and Luzon-Mindanao lineages (8). Among these, the mainland Asian lineage (hereinafter referred to as mainland Asian honeybee or mainland *A. cerana*) colonized the largest area, stretching from Northeast Asia to the Kra Isthmus of Thailand, where it is separated from the Sundaland lineage (8). Recent genomics work on the mainland Asian honeybee revealed several genetically distinct populations on a tropical island and in the canyons of the Q-T Plateau (9), some of which were previously proposed as subspecies (e.g., Hainan, Bomi, and Aba) (10). Given that these localities...
are geographically distant and/or isolated (5), local establishment is likely the result of independent colonization events from the same ancestral population, each being exposed to new and diverse habitats characterized by varied flowering phenologies and florals. However, the detailed population structure and differentiation processes of the mainland Asian honeybee still remain unclear yet are crucial prerequisites for understanding adaptive evolution and for formulating efficient conservation strategies for this vital pollinator in Asia.

The success of the Asian honeybee in colonizing diverse habitats is largely enabled by flexibility in colony foraging activities for pollen and nectar. Pollen serves as the main protein and lipid resource and is in the highest demand during the reproductive season (11). On the other hand, nectar provides energy for the entire colony throughout the year (11). Nectar-derived honey stores are particularly important for surviving winter (11), and large stores of this energy source during winter months have presumably enabled range expansion to temperate regions (4) and high-altitude canyons (5). For instance, A. cerana populations in alpine canyons of the Q-T Plateau (e.g., Aba and Qinghai populations) show strong nectar collection and honey storage capacity at the colony level (10, 12, 13), as an adaptation to a condensed flowering season (14) and prolonged winter. They primarily reproduce during the flowering season and swarm only once a year (10, 12, 13). In contrast, tropical A. cerana on Hainan island displayed reduced tendency to collect and store honey but strong colony-level pollen collection capacity (10, 12), reflecting the long flowering season (15) and lasting reproduction needs (10, 12). Note that these colony-level behavioral adaptations to local climates are persistent within A. cerana subspecies and ecotypes. For instance, the reluctance to build up large honey stores turned into a lethal disadvantage when Hainan colonies were moved to Beijing, where they failed to survive the harsh northern winter (12). The quantity and seasonal dynamics of nectar collection are largely heritable within A. cerana subspecies and ecotypes regardless of where they are reared (16).

Food preference at the colony level can be achieved via adjustments to foraging labor divisions within the colony (17). For example, a higher proportion of nectar foragers results in an increase to the honey/pollen ratio for the colony and vice versa (17, 18). This division of foraging labor is established at the individual bee level and is predetermined at an early stage in foragers via differential sensitivity to sucrose (19). For example, in A. mellifera, foragers that were 1 week old with low sucrose response thresholds tended to collect water, while those with medium or high thresholds tended to collect pollen or nectar, respectively, when they started to forage (19). Further work showed that these varied sensitivities are genetically based and affected by age and ambient environment (20). In addition, sucrose responsiveness was found to differ among workers of different patrilines (daughters of different males) within the same colony (21), resulting in disparate preferences in foraging tasks when exposed to the same stimulus (20). Thus, the differing task thresholds among patrilines can serve as the basis for natural selection, through which the colony’s collective food preference is modified (17) according to changes in flora. The evolution of the collective phenotype driven by colony-level selection was first proposed by Darwin (22). Nevertheless, the genes associated with foraging labor division, the important link that connects individual behaviors and colony adaptation, are still unknown for A. cerana. We hypothesize that, as a universal adaptive strategy, the genes underlying adjustments to a colony’s foraging tendency through labor division should be repeatedly selected across radiating descendant populations. Therefore, by identifying these genes, we will not only uncover the genetic basis of the behavioral adaptation that is essential to the survival of the honeybees but also improve our understanding concerning the role of gene reuse in facilitating adaptation into diverse environments.

In the present study, we aim to (i) ascertain the population structure and evolution history of the mainland Asian honeybee through whole genome resequencing and geometric morphometric evaluation and subsequently use this system to (ii) identify the prevalence of gene reuse among genetically distinct population groups (subspecies) inhabiting diverse local habitats and (iii) explore the functions of the key gene under consistent selection across groups to confirm its role in foraging labor division.

### RESULTS

#### Sampling and sequencing

We sequenced a total of 306 A. cerana worker bees across nearly its full distribution range in China and adjacent areas (Fig. 1A). Thirteen individuals collected on flowers were identified as full- or half-sisters to other individuals based on the kinship coefficient (data S1) and were subsequently excluded from further analyses. As a result, genomic sequences representing 293 colonies (one worker per colony) were retained in our study (Fig. 1A and data S1). The average coverage depth of clean data was approximately 15x. After adding public data (50 workers from multiple areas; data S2) and filtering the data, about 4.5 million single-nucleotide polymorphisms (SNPs) from 343 workers were retained for the following analyses.

#### Multiple peripheral groups were independently derived from a single Central group

Following linkage disequilibrium (LD) pruning, about 1.8 million SNPs were used to infer population structure. Eight genetically distinct clusters were detected in the NetView network when the number of mutual nearest neighbors (mk) was 130 (Fig. 1B), corresponding with the point where the number of detected clusters reached an asymptote (fig. S1A). Among these, one cluster was referred to as the Malay group (ML; including Malaysia and Singapore) and belongs to the Sundaland A. cerana (8), while all other clusters are part of the mainland Asian honeybee.

Within mainland A. cerana populations, one cluster in the NetView network was defined as the Central group, including populations widely spread in the central region of the study area. ADMIXTURE produced two ancestry components in populations of the Central group, whose relative composition formed a transition from southwest to northeast (Fig. 1, A and C). However, they were widely admixed without clear genetic boundaries throughout the central range (Fig. 1A). Thus, we defined the Central group as a single genetically distinctive group. The population in the north temperate zone within the Central group roughly reflects the North ecotype of the Chinese A. cerana (10) (see the Supplementary Materials for details), which was embedded within the southern populations in the neighbor-joining (NJ) tree (fig. S2) and was, therefore, inferred as a derived ecotype from the southern populations.

The six remaining clusters in the network were primarily located in the peripheral range (Fig. 1A) and are referred to as “peripheral groups” hereafter. These six peripheral groups contained populations from Hainan (HN; including Hainan Island and Yongxing Island), Taiwan (TW), Bomi (BM; southeastern Tibet), Aba (AB; including Maerkang and Jinchuan), Qinghai (QH), and Northeast Asia (NE;
including Heilongjiang and Jilin in China, North Korea, South Korea, and Japan). This population structure was further supported by ADMIXTURE (Fig. 1C), principal components analyses (PCAs) based on whole genome SNPs (fig. S3), haplotype networks of the mitochondrial cytochrome oxidase subunit I (COI) gene (fig. S4), and the geometric morphometric evaluation of forewing venation of worker bees (fig. S5 and table S1) (see Supplementary Materials for details).

The pairwise fixation index ($F_{st}$) (fig. S6 and table S2) between each peripheral group and the Central group ($0.165 \pm 0.068$; mean $\pm$ SD) was significantly lower than those between peripheral groups ($0.344 \pm 0.095$) ($P < 0.001$; two-tailed Mann-Whitney test), indicating that the Central group was the most genetically similar group to each of the six peripheral groups. The level of average genetic differentiation (average $F_{st} = 0.163$) between peripheral groups and the Central group was similar to that between subspecies of *A. mellifera* (average $F_{st} = 0.10$) (23), implying their divergences at a subspecies level. When the Central group was divided into multiple populations according to provinces or countries (for non-Chinese populations), the $F_{st}$ values between populations within the Central group were low ($0.039 \pm 0.018$). This pattern revealed limited differentiation within the Central group despite its wide distribution range, which may result from a recent rapid expansion or frequent gene flow. The nucleotide diversity ($\theta_s$), observed heterozygosity ($H_o$), and expected heterozygosity ($H_e$) of the Central group were all significantly higher than those of peripheral groups (adjusted $P < 0.001$; two-tailed Mann-Whitney test) (fig. S6). Thus, a single widespread Central group and multiple peripheral groups can be identified genetically within the mainland Asian honeybee. Although these units represent subspecies-level divergences, formal proposal of subspecies is out of the scope of this study. Thus, these genetically distinct groups were still referred to as population groups hereinafter.

Phylogenetic reconstructions recovered the ML group as sister to all other populations sampled (Fig. 1, D and E), implying an early division between the Sundaland and mainland Asian honeybees. Within the mainland bees, the peripheral groups were nested within the Central group in both the NJ (Fig. 1D and fig. S7) and the maximum likelihood phylogeny trees (Fig. 1E) and were spread throughout the phylogeny. Unlike peripheral groups, evolutionary changes quantified by drift parameters were comparatively negligible within the Central group (Fig. 1E). Thus, the combined results from phylogeny, genetic similarity, and genetic diversity indicated that each of the peripheral groups was independently derived from the Central group.

**Mainland *A. cerana* radiated rapidly to diverse habitats**

On the basis of the best-fit isolation with the migration model, the divergence times between peripheral groups and the Central group...
were estimated as ca. 124 to 98 thousand years (ka) ago (Fig. 1F and table S3), corresponding with the time when the effective population size \((N_e)\) of the Central group had reached a peak and began to decrease, in line with the downward temperature trajectory (indicated by a small arrow in Fig. 1G). The \(N_e\) of the Central group decreased during the glacial periods and increased during warm periods (Fig. 1G). This pattern was also supported by ecological niche modeling, which showed that the potential range of the Asian honeybee had retracted toward the south during the last glacial maximum (LGM) (fig. S8), and then again expanded northward after the LGM. Thus, it can be inferred that the Central group likely underwent multiple expansions and contractions throughout its history. The gene flow between each peripheral group and the Central group was inferred to be bidirectional after divergence (fig. S9 and table S3), suggesting a reciprocal introgression.

On the basis of the above results, we conclude that the peripheral populations colonized new habitats during the expansion phase of the Central group before ca. 100 ka ago. Subsequently, the Q-T Plateau canyons may have served as effective refugia for the peripheral populations (Fig. 2A) during the glacial period when the Central group retracted (Fig. 1G). The alterations to the environmental conditions in the peripheral habitats would have likely posed strong selective pressure on the resident populations. Geographic barriers (e.g., long and narrow canyons with confined openings; Fig. 2A) would have constrained genetic exchanges between the peripheral and Central groups through secondary contacts during reexpansion of the Central group during the warm period. While the AB, QH, BM, and NE groups have probably followed this evolutionary history (Fig. 2B), the island groups (TW and HN) may have migrated to their current regions during the glacial period when the islands were connected with the continent and became isolated when the sea level rose during the warm period.

The bioclimates (e.g., precipitation, temperature, etc.) of peripheral habitats show notable deviations from that of the Central group, as well as from each other, at both the present time and during the LGM (Fig. 3A). These distinct bioclimates were likely an active driver of shifts in floral composition (24) and flowering phenology (25). Our analyses based on data from long-term phenological records in the Chinese range of \(A. cerana\) (14) suggest that the length of the flowering season is highly variable among different climatic regions (Fig. 3B). The flowering season in the tropical region lasts all year round, and its duration is successively shortened in the subtropical and temperate regions. The highland canyons at the eastern edge of Q-T Plateau have the shortest flowering season, where plants only bloom three to four months/year (Fig. 3B).

Thus, these new peripheral habitats (Figs. 1A and 2) differ from the central ancestral habitats, both in flowering phenology and floral composition. The length of flowering season was inferred to be shorter for the QH, AB, BM, and NE groups but longer for the HN group, compared to that of the central ancestral habitats. These variations would have presumably posed selective pressure on pollen/nectar collection tendency at the colony level, which coincides with the varied capacities in pollen/nectar collection and honey storage among respective \(A. cerana\) populations (4, 10) (see the Supplementary Materials for details).

**Gene reuse was prevalent during rapid radiation and independent habitat adaptation**

To detect selective sweeps during the radiation of the peripheral groups from their central ancestry, we scanned the whole genome using both population differentiation \((\theta_{st} \text{ ratio and } F_{st})\) between the Central group and each peripheral group) and composite likelihood ratio (CLR) methods. Only genes detected by both methods were considered as candidates for selective sweeps (179 genes in total; see data S3 for the full gene list). Overall, the frequency distribution of the selected genes was strongly skewed toward lineage-specific genes (Fig. 4A), indicating that the majority of genes under selection were unique to each of the genetically distinct peripheral groups. Some of these genes might be relevant to local adaptations, such as those associated with olfactory reception, cold hardiness, food selection and consumption, and learning and memory (see data S3 for details). On the other hand, 26 genes (ca. 14% of the 179 selected genes) (data S3) were under selection in at least two peripheral groups (i.e., repeated selection), significantly more than expected by chance \((P < 0.001; \text{exact multinomial test for goodness of fit})\) (Fig. 4B). This result implies that some genes in \(A. cerana\) tended to undergo selection repeatedly during population radiation, providing strong evidence for adaptive evolution. These repeatedly selected genes are likely a reflection of shared selective factors, where many of them are associated with appetite, circadian rhythm, phototaxis, optesthesia, mating, immunity, and DNA repair (see data S3 for details).

Gene scanning using LASSI (likelihood-based approach for selective sweep inference) software suggested that, for all genes under repeated selection, natural selection preferred to act upon standing variations (i.e., soft sweep) and noncoding regions, when compared with the uniquely selected genes (fig. S10; see the Supplementary Materials for details). Such a mechanism would enable the bees to take advantage of existing genetic variation and to create regulatory flexibility while still maintaining core gene functions (26), thus increasing the likelihood of gene reuse.
A consistently reused gene was associated with sucrose sensitivity and social division of foraging labor

Among the genes under repeated selection, the Leucokinin receptor (Lkr; LOC107998395) gene was detected in all peripheral groups by \( F_{st} \), \( \theta_{rs} \), and CLR methods (Fig. 4C and figs. S11 and S12), except in the TW group, where it was only detected by CLR (fig. S11; see the Supplementary Materials for details for gene identification). Gene scanning using LASSI revealed that specific positions of selective peaks on Lkr varied among peripheral groups (Fig. 4D), suggesting that these selective sweeps were result of independent evolutionary events rather than plesiomorphic traits inherited from their common ancestry. In addition, within the Central group, Lkr was also under selection in the North ecotype (derived) (fig. S13) but not in the southern populations (ancestral). Given the evolutionary history of the mainland A. cerana revealed by this study, the gene selection results suggested that Lkr was a crucial gene involved in almost all major habitat transitions during range expansion.

The relative expressions of the Lkr gene were significantly higher in both antennae and brain, centers for signal perception and behavioral regulation, than those in thorax and gut (\( P < 0.01 \)) (Fig. 5A and fig. S14A). Furthermore, the Lkr was expressed at significantly higher levels in the antennae and brains of foragers than those of nurse bees (\( P < 0.01 \)) (Fig. 5A), indicating its elevated importance in foraging bees.
To test the specific involvement of Lkr in sucrose responsiveness, we performed proboscis extension response (PER) experiments on foragers of *A. cerana*. Facilitated by nanocarriers, RNA interference (RNAi) knockdown of Lkr effectively reduced its expression in the brain of foragers (fig. S14) and led to a significant increase in the sucrose response threshold when sucrose solution was applied directly to (Fig. 5C) or placed closely to but not touching the antennae (Fig. 5D). Thus, these provide evidence that the expression level of Lkr in the brain of *A. cerana* regulates its sensitivity to sucrose. Given that differences in sucrose responsiveness predetermine collection
tasks (water, pollen, and nectar) in honeybee foragers (19), our results indicate that Lkr can directly influence division of foraging labor. This is further supported by our finding that nectar foragers showed significantly lower Lkr expression than pollen foragers \( (P < 0.001) \) (Fig. 5B), confirming Lkr’s involvement in *A. cerana* foraging division.

Despite the findings of these experiments, it remained unclear how Lkr expression was regulated in *A. cerana*. However, fine signal scanning on Lkr suggested that the first intron (intron-1), which might be involved in Lkr regulation, was repeatedly under selection. The regions (50 SNP sliding window) bearing peak \( T \) values in the LASSI results all fell within intron-1 (Fig. 4D), suggesting that beneficial mutations had most likely occurred in this region. In addition, haplotype diversity in the Lkr intron-1 decreased in all peripheral groups when compared to the Central group and radiated rapidly to diverse habitats around 100 ka ago. The Central group served as a reservoir with high potential for diversification, pumping out genetically unique populations through range expansion and contraction. Therefore, these findings indicate that not only peripheral populations adapted to local habitats deserve conservation consideration but also the Central group, which serves as the source of new descendants.

**DISCUSSION**

Our study has revealed the population structure and rapid radiation history of the mainland Asian honeybee in unprecedented detail. Six genetically distinct peripheral population groups surrounding the widespread Central group were recognized. These peripheral groups all derived independently from the Central group and radiated rapidly to diverse habitats around 100 ka ago. The Central group served as a reservoir with high potential for diversification, pumping out genetically unique populations through range expansion and contraction. Therefore, these findings indicate that not only peripheral populations adapted to local habitats deserve conservation consideration but also the Central group, which serves as the source of new descendants.
should also be subject to particular protection. Furthermore, this model system grants us an unprecedented opportunity to test whether gene reuse (repeated selection on the same gene) can facilitate adaptation to diverse habitats.

Young descendant populations derived from a common ancestor often have a high probability of gene reuse (1). However, this pattern has been reported almost exclusively in organisms that are repeatedly exposed to new habitats of the same type (1), e.g., from marine to freshwater in three-spine sticklebacks (2). One possible exception is the study on Darwin’s finches, which suggested that the ALX1 gene might be related to the variation in beak morphology, thereby facilitating their adaptation to a variety of seeds as food (29). In A. cerana, we showed that the Lkr gene associated with social division of foraging labor was essential in the adaptation to diverse habitats with various differences in flora and phenology. This behavioral strategy could be further strengthened by the involvement of additional genes associated with general foraging and social behaviors. For instance, a number of repeatedly selected genes (e.g., 5-HTR1, aop, and Klf7) in this study are associated with appetite, vision, learning and memory, and circadian rhythms (data S3). Our results thus advance our knowledge that gene reuse can also be a common pattern in a radiation to diverse habitats.

Our findings related to A. cerana add important knowledge to the understanding of habitat adaptation in honeybees. In A. mellifera, several quantitative trait loci influencing foraging division have been identified (20), within which the octopamine receptor genes were strong candidates for adaptation to highland habitats in the East African honeybee (A. m. monticola) (30). Along with the evidence of repeated selection on Lkr from the present study, these results thus imply that colony-level adaptation through individual foraging division might be a common adaptive mechanism to habitat changes in eusocial honeybees and that the selection pressure posed by habitat changes may have led to convergent traits related to foraging division in these two honeybee species. In social bees, colony-level foraging behavior is typically mediated by reproductive status, a trait that is probably derived from an ancestral ground plan that functions to synchronize dietary preferences with reproductive needs in solitary insects (31). Given the facts that pollen is used primarily as a protein and lipid source for reproduction and that nectar primarily provides energy for growth and survival (11), the ratio of task division in A. cerana foragers essentially balances the trade-off between growth/survival and reproduction. This is especially important when A. cerana ventures into new territories, such as range expansion into diverse habitats.

Classic studies have reported multiple functions for the Leucokinin (Lk) and Lkr genes in Drosophila and other insects, including feeding, meal size, desiccation and starvation resistance, and circadian rhythms (32), suggesting that they are involved in diverse but coordinated processes. Our study demonstrates that Lkr is involved in regulation of foraging labor division in honeybees, a phenomenon linked to sucrose sensitivity. The consistent selection on this same gene across peripheral groups indicates that colony-level adjustment of food composition is a vitally important adaptive trait, which helps the honeybee to survive when exposed to a new environment. This previously unidentified function in social bees appears to be rooted from highly conserved traits. For instance, Lk and Lkr are homologous to vertebrate Tachykinin and its receptor (33), implying that the roles of the tachykinin system in regulating food intake might be evolutionarily conserved between insects and vertebrates. How Lkr affects sucrose responsiveness remains to be clarified. However, the mechanism is likely connected to the insulin pathway. Lkr is closely related to the insulin pathway in Drosophila (32), while at least some genes associated with division of labor and foraging in the social insects are also involved in the insulin pathway (34).

Our analyses of the distribution of selective signal along the Lkr gene also provide some suggestions of the potential regulatory mechanism. We speculate that different Lkr alleles may control Lkr expression levels through possible cis-acting (e.g., enhancer) or trans-acting regulation (e.g., IncRNA) (fig. S17) on intron-1. Thus, repeated selection on intron-1 might lead to differential Lkr expression, which is then associated with colony-level pollen/nectar foraging activities, mediated through individual adjustments to sugar sensitivity and foraging tendency (Fig. 6). Consistent with this theory, the selected haplotype(s) carrying the beneficial allele in Lkr intron-1 was found

---

Fig. 6. A proposed model showing how Lkr under selection translates to a colony-level adaptive phenotype.
in high frequency in each peripheral group (fig. S16), as a result of independent selective sweeps. The modification of Lkr allele frequency in each peripheral population group could change proportions of patrilines within a colony, a unique characteristic in honeybees caused by polyandry. Variation in sucrose thresholds among patrilines would translate into a shift in the proportion of each foraging task in a colony (fig. S18 and Fig. 6), that is, in the colony-level tendency for pollen or nectar collection and honey storage (17, 20). Such plasticity in colony-level phenotype would enable rapid adaptation to diverse habitats with varied florars primarily shaped by bioclimates, such as flowering season length (Fig. 3) and relative abundance of pollen/nectar source plants.

In summary, our study suggests that gene reuse is a crucial mechanism for A. cerana to adapt to diverse habitats. In the Asian honeybee, the repeated selection on the Lkr gene has enabled flexible adjustment of foraging labor division at the colony level, as a common adaptive strategy to climatic and floral changes. Our findings also highlight the possibility that this colony-level adaptation might be a common strategy for other eusocial honeybees, as a shared derived trait or convergent character. Further research is necessary to understand the specific molecular mechanism of Lkr regulation and its impact on colony adaptation.

**METHODS**

**Sample collection**

Specimens of A. cerana were collected across its nearly full distribution range in China and adjacent areas (Fig. 1A and data S1). They were preserved in 95% ethanol after collection and stored at −20°C. To avoid the influence of potential genetic admixture caused by introduction of exotic populations, bee samples were collected from endemic populations. For regions that have been under the risk of population introgression, specimens collected before intense beekeeping activities were used, including those from 18 years ago (data S1). A single worker was randomly selected from each colony for sequencing, and a total of 327 workers were used for sequencing.

Specimens from China, North Korea, and South Korea were collected mainly from managed beehives populated with native feral swarms, except for a few caught on wild flowers (data S1; see the “Relatedness of bees caught on flowers” section for kinship identification). Our sampling covered all nine ecotypes of Asian honeybee in China (a.k.a. Chinese honeybee), which were mainly classified on the basis of geographical distributions and morphology (10), including Aba, Tibet, Changbaishan, Hainan, Southern China, Central China, Northern China, Yun-Gui Plateau, and South Yunnan ecotypes. Specimens from Thailand, Vietnam, Malaysia, and Singapore were mainly collected on flowers in the field (data S1).

**DNA preparation and sequencing**

Genomic DNA was extracted from the head and thorax of each worker bee using the DNeasy Blood and Tissue Kit (Qiagen, Germany) following the manufacturer’s instructions or using the phenol-chloroform extraction method. Insert size libraries of 200 to 400 bp were constructed at BGI-Shenzhen and sequenced on the BGISEQ-500 platform with 100-bp paired-end (PE) reads following standard procedures. Samples were sequenced to a target raw depth of around 15 to 20× (ca. 3.5 to 4 Gb of raw data per sample). The raw sequence data of all 327 worker bees from this study have been submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database (accession number PRJNA592293).

**Public data**

We downloaded public sequence data of 10 A. cerana workers from Japan and 10 A. mellifera workers from the NCBI SRA database (data S2). In addition, we also included published sequences of 40 workers of A. cerana representing four complementary Chinese populations, including Diqing, Yulin, Jiuzaizi, and Sining from a recent study (9). The first three were proposed as tentative distinct populations, and the last one was complementary to our sampling in the Sichuan Basin. The detailed information of all public data is listed in data S2.

**Mapping, variant calling, and filtering**

Before mapping, we carried out quality control for raw sequence data using fastp 0.13.1 (35) with parameters “–q 10 -n 10 -u 40.” Both paired reads were filtered out if either one contained >40% of low-quality bases (Q < 10) or >10% Ns. Adapter contaminations were also trimmed.

We mapped clean reads to the A. cerana reference genome (ACSNU-2.0, GCF_001442555.1) (36) using the Burrows-Wheeler Aligner (BWA)–MEM algorithm (v 0.7.17-r1188) (37), with default settings and an additional “–M” parameter to reach compatibility with Picard. Public data sequenced on SOLiD (data S2) were mapped separately using SHIMP 2.2.3 (38) and default settings, because the format of these data was color space. Read duplicates were marked using Picard MarkDuplicates 2.18.9 (http://broadinstitute.github.io/picard/) after mapping.

The Genome Analysis Toolkit 4.0.4 (39) was used to call short variants. We used HaplotypeCaller in the GVCF mode to call variants for each sample. The indel realignment had been integrated into HaplotypeCaller via reassembly of haplotypes in active regions. All of the per-sample GVCFs were then passed to the joint genotyping tool GenotypeGVCFs to produce a set of jointly called SNP and indel calls that were ready for filtering (including ca. 25 million raw SNPs and 7 million indels).

We then filtered the variant file and retained only high-quality, biallelic SNPs sites that met all of the following criteria: (i) average depth > 1 / 3 and ≤ 2× of the mean depth of the whole dataset; (ii) quality score > 20; (iii) average genotype quality > 20; (iv) minor allele frequency > 0.01; (v) proportion of missing genotypes < 50%.

**Relatedness of bees caught on flowers**

Forty-six workers in our samples were collected on flowers in the field (data S1), when managed colonies were not available. To best represent the diversity of local populations, only one worker was allowed to represent each colony in our dataset, because those from the same colony may potentially have more homogeneous genomic features. Therefore, we calculated kinship coefficients between individual pairs using the KING-robust method (40) to identify the relatedness of worker bees caught on flowers. Because a queen can mate with multiple haploid drones, the kinship coefficient is 0.375 (1 / 8) between full-sisters and is 0.125 (1 / 8) between half-sisters (same mother and different drone fathers) in a colony (note that, by definition, the kinship coefficient is half of the coefficient of relationship) (11). Thus, the kinship coefficient between workers from the same colony is expected to be greater than or equal to 0.125. Therefore, workers with kinship coefficient ≥ 0.125 were considered as belonging to the same colony, and only one worker was retained for each colony.
Population structure and phylogenies
To avoid the potential influence of LD on analyses of population structure and admixture, we generated a subset of pruned SNPs using the function “snpqgsLDpruning” in R package SNPRelate (41). The correlation coefficient method was used with a window size of 50 SNPs and an LD threshold of 0.2. The PCA and identity-by-state (IBS) distance (also called allele sharing distance) matrix were performed and constructed using functions “snpqgsPCA” and “snpqgsIBS” in SNPRelate, respectively. A NJ tree was reconstructed on the basis of the IBS distance matrix using the function “nj” in R package Ape (42). Node support values were obtained after 1000 bootstrap replicates. The population structure was also visualized using the R package NetView 1.1 (43) based on IBS distance matrix. This package generated networks using a minimum spanning tree and mutual k-nearest neighbor graphs. As this method is sensitive to missing data (43), individuals with a missing rate ≥ 10% (26 individuals) were removed. The “Infomap” algorithm was used to detect clusters in a network. The cluster is a structure containing individuals that are more densely linked to each other than to the others in a network, which is equivalent to the genetically distinct group or lineage in population structure. The program ADMIXTURE 1.3.0 (44) was used to infer ancestries. Fifteen independent runs for K = 1 to 15 were performed.

TreeMix 1.13 (45) was used to reconstruct the maximum likelihood population tree. Because SNPs closely distributed on the genome may not be independent, they were grouped together in a window size of 500 SNPs, using the parameter “-k 500.” Bee populations were partitioned according to genetically distinct groups identified in this study, and the Central group was further divided into multiple populations based on provinces or countries (for non-Chinese populations). Populations represented by only a single individual were not included in TreeMix analyses.

We also extracted the full-length sequences of mitochondrial COI gene from our dataset. SNPs were called using the pipeline congruent with that for the nuclear genome, except for the parameter “--sample-ploidy 1” in HaplotypeCaller. The full-length COI sequences for each individual were obtained using “bcftools consensus” tool (46). Haplotype networks were constructed using PopART 1.7 (47) with minimum spanning and TCS methods.

Genetic diversity and Fst statistics
θe, H0, and He for each group or population were calculated using the tool “populations” in software Stacks 2.2 (48), with a window size of 10,000 SNPs. Multiple comparisons were performed with the two-tailed Mann-Whitney test, and P values were adjusted using the “holm” method. The pairwise Fst matrix was calculated using the R package SNPRelate (41) and Weir and Cockerham’s method, which is unbiased to unbalanced sample sizes. The software Beagle 5.0 (49) was used to phase with the parameter “niterations = 10,” and the haplotype diversity was conducted using the R package haplotypes 1.1 (https://cran.r-project.org/package=haplotypes).

Historical effective population sizes
We used SMC++ 1.13.1 (50) to estimate changes in the Nₑ over the past 3 million years (Ma). The program only considered variant sites from the input variant call format (VCF) file, while those not included in the VCF file (including not only invariant but also uncalled and filtered sites) were all treated as homozygous references across all samples. Thus, the level of homozygosity could be overestimated and the Nₑ in the past might be misleading (50). Therefore, we masked the filtered and uncalled regions in the VCF file using the “-m” parameter. The mutation rate was set to 5.27 × 10⁻⁹ per base pair per generation, following the divergence estimation of 7 Ma between A. mellifera and A. cerana (23). The generation time was assumed as 1 year. The polarization error was set to 0.5 as the identity of the ancestral allele was unknown.

Demographic models and divergence time
To estimate divergence times and gene flow between each peripheral group and the Central group, we used two demographic models and conducted maximum likelihood estimations using the Python module momi2 (51) based on site frequency spectrum (SFS). The null model was a clean split model, which assumed no gene flow after population divergence. The divergence time and Nₑ were set as free parameters for the null model. Alternatively, the isolation with the migration model assumed the presence of gene flow after divergence, in which case three extra free parameters, including two interactive migration rates and occurrence time, were added. The null and alternative models were compared using the likelihood ratio (LR) test according to the formula LR = 2(lnL₁ − lnL₀), where lnL₀ and lnL₁ were log-likelihood of the null and alternative models, respectively. The null model would be rejected if the P value of the LR test (approximate χ²-distribution) was less than 0.01. To expedite computation for likelihoods, we down-sampled each group to only include five representative individuals from distant locations that also had relatively high sequencing and mapping quality (data S1). We conducted 100 independent runs to identify the global optimum. Of these runs, we chose the one with the highest likelihood as the preferred point estimate for all parameters. One hundred bootstraps were conducted by resampling blocks of the SFS to calculate confidence intervals and to estimate the uncertainty of parameters.

To examine whether the divergence times were congruent with the contraction period of the Central group, we constructed three divergence models (A, B, and C; fig. S19). In these models, the Central group had undergone a contraction followed by an expansion. The peripheral group diverged from the Central group in the expansion period of the latter in both models A and C, but in the contraction period in model B. The likelihood of each model was estimated using the Python module momi2 (51) based on SFS. The Akaike information criterion (AIC) was used to choose the best-fit model according to the formula AIC = −2lnL + 2 K, where lnL was the log-likelihood and K was the number of free parameters of the model.

Geometric morphometrics
We conducted geometric morphometrics and statistical analyses using forewing traits of worker bees. A set of 211 worker bees collected from 19 sampling sites (table S7) was used for the geometric morphometrics analyses. Most of these specimens were newly sampled because many of the bees used for genome resequencing had been stored in 95% ethanol for years (data S1), which causes distortion and damage to wing characters. Bees were mainly collected from the same apiaries that provided samples for sequencing or those adjacent. The right forewings of worker bees were fixed with two slides and photographed using a Nikon D700 camera at the Institute of Zoology, Chinese Academy of Sciences. A total of 20 landmarks were identified on the forewings (fig. S5). All landmarks were digitized using tpsDig 2.0 and tpsUtil 1.40 (http://life.bio.sunysb.edu/morph). The landmark configurations were then superimposed
using Generalized Procrustes Analysis in the software Coordgen 6.0 (https://life.bio.sunysb.edu/morph).

Canonical variable analyses (CVAs) were performed using the size-corrected shape (residuals from multivariate regression of shape on size) to show the degree of morphological discrimination among distinct groups that were identified in the genomic analyses. The CVA was conducted in Morphoj 1.06 (52) with the thin plate spline method.

**Ecological niche modeling**

To infer the possible distribution range of *A. cerana* in different climatic periods, we downloaded 19 bioclimatic variables of the current climate (v2.0) and the LGM (MIROC-ESM; v1.4) at the resolution of 2.5 min from WorldClim (www.worldclim.org) as predictors for niche modeling. To reduce multicollinearity in the bioclimatic variables, the R package raster (https://cran.r-project.org/package=raster) was used to calculate the Pearson correlation index among them, and seven variables (bio02, bio04, bio05, bio06, bio15, bio16, and bio19) with Pearson correlation index < 0.75 were retained for further analyses.

We downloaded occurrence records of *A. cerana* from the Global Biodiversity Information Facility (GBIF occurrence download; https://doi.org/10.15468/dl.rpjgqw). These data were filtered to remove duplicates and records with geo-referencing errors. A total of 900 unique occurrence records were obtained. In addition, only a single record was retained for those within 5 km of each other to be compatible with the resolution of bioclimatic variables and to reduce potential sampling bias. At last, 765 occurrence records were used for niche modeling.

The program MaxEnt (53) was used to model ecological niche and to predict potential distributions for *A. cerana*. The R package ENMeval (54) was used to calculate the preferred combination of feature types and regularization multiplier with the minimum AICc (corrected AIC). We randomly partitioned 75% of occurrence data for modeling and the remaining 25% for model tests. “Maximum test sensitivity plus specificity” was applied as the threshold rule, with other settings as defaults.

**Bioclimatic differentiation and flowering phenology**

To explore the ecological differentiation across the distribution area of our sampling sites, the values of bioclimatic variables for all sampling sites were extracted by DIVA-GIS (www.diva-gis.org) and were z-score normalized. We performed PCA using the “prcomp” function in R to obtain the first two principal components. The results of both present conditions and those in LGM were shown as scatter plots. The length of flowering season in different locations was explored on the basis of the plant phenological observation dataset of the Chinese Ecosystem Research Network (14). We used the phenological observation data from 78 sample plots across the distribution range of Asian honeybee in China (table S9), including observation data of more than 480 flowering plant species during 2003–2015.

**Detection of selective sweeps**

We used both population differentiation and CLR methods to detect genome regions under selective sweep. In the population differentiation method, $F_{st}$ and $\theta_{st}$ were calculated in 10-Kb sliding windows and a 5-Kb step size using VCFTools (55). Only regions containing at least 10 SNPs in a sliding window were used. Candidate regions were extracted if they fell in the intersection of the top 5% of both Z-transformed $F_{st}$ and log value of $\theta_{st}$ ratio (ratio of the Central group to each peripheral group). Genes overlapping with these regions were annotated on the basis of the reference genome of *A. cerana* (GCF_001442555.1; NCBI Annotation Release 101). In the CLR method, SweeD 3.1 (56) was used with a 1-Kb grid size. The CLR method does not need to compare between each peripheral group and the Central group but only needs SNPs of each peripheral group as input. CLR was calculated by comparing the selective sweep model to the neutral model. Positions with CLR value above the top 1% cutoff level were extracted. Genes including these positions were identified. The genes detected by both methods were considered as candidate genes under selection. Expected intersection sizes among peripheral groups were calculated using the Monte Carlo method of random sampling, and one million replicates were conducted. Because the expected values were very small, exact multinomial tests for goodness of fit were performed between observed and expected values using the “multinomial.test” function in the R package EMT (https://CRAN.R-project.org/package=EMT). Gene Ontology (GO) enrichment was conducted using the enrichGO function in the R package clusterProfiler (57).

We further used the program LASSI (58) to differentiate hard and soft sweeps, with the window size and step size of 50 and 5 SNPs, respectively. Selective sweeps were detected from an elevated value of the $T$-statistic, which is defined as the log CLR of a sweep model compared to the genome-wide background (taken as a model of neutrality). Regions with the peak point of $T$ value were regarded as the locations where the beneficial variation most likely occurred. For regions with elevated $T$ values, sweeps were inferred to be hard when the model parameter $m$ (number of sweeping haplotypes) = 1 or soft when $m > 1$.

**Gene tree and intron phase**

To explore the phylogenetic position of LOC107998395 (annotated as RYamide receptor–like or the RYa–R–like gene in NCBI), we downloaded 625 protein sequences of the G protein–coupled receptor (GPCR) family from 15 bilaterian species (http://neuroevo.org) (59) and added six protein sequences annotated as RYa–R–like from both the *A. cerana* and *A. mellifera* genomes. Sequence were aligned using MAFFT 7.310 (60) using the FFT-NS-2 method, and ambiguous sites were trimmed using trimAl 1.4 (61) with default parameters. The gene tree was reconstructed using FastTree 2.1 (62) with maximum likelihood and default parameters.

We also identified positions and phases of intron insertions in protein sequences of LOC107998395 and eight Lkrs from seven bilaterian species (59). An intron can be categorized to one of three phases based on which codon position the intron/exon boundary falls in: phase 0 (between two consecutive codons), phase 1 (between the first and second nucleotides of a codon), and phase 2 (between the second and third nucleotides of a codon).

**Tissue and labor specificity of selected genes**

To understand tissue specificity of identified genes under selection, we downloaded public transcriptome sequence data generated from various tissues of *A. cerana* workers from the NCBI SRA database, including brain (accession number SRR1380970), antennae (SRR1380976), hypopharyngeal gland (SRR1380979), gut (SRR1380984), and fat body (SRR1388774) (36). Reads were aligned to the *A. cerana* reference genome (ACSNU2.0; GCF_001442555.1) (36) using HISAT2 2.1.0 (63). Before aligning, the index of exons and splice
We conducted real-time quantitative polymerase chain reaction (qPCR) experiments to compare the expression levels of the target LOC107998395 gene in various tissues of nurse and foraging bees. Samples of *A. cerana* were collected from a colony at the apiary in Miyun District, Beijing, a representative of the Central population group. Nurse bees were collected from brood combs while nursing larvae. Foragers carrying pollen on their corbiculae were collected at hive entrances and were dissected immediately. Dissected antennae, thorax, gut, and abdomen were stored in liquid nitrogen on site. Brains were dissected in RNALater (Invitrogen) on an ice bag. These tissues were transferred into TRIzol reagent (Invitrogen) to extract total RNA according to the manufacturer’s protocols. cDNA was prepared using the HiScript III RT SuperMix for qPCR (Vazyme, China). qPCR was performed using the ChamQ Universal SYBR qPCR Master Mix (Vazyme, China) and QuantStudio 1 Real-Time PCR Instrument (Applied Biosystems, USA) in a standard 96-well block (20-μl reactions; incubation at 95°C for 3 min, 40 cycles of denaturation at 95°C for 10 s, annealing/extension at 60°C for 20 s). The primers of the candidate gene are shown in table S8. The *A. cerana* Actin gene was chosen as the control, and relative expression was analyzed using the 2^ΔΔCT method.

We further conducted qPCR experiments to compare the expression levels of the LOC107998395 gene in foragers with different collection tasks. Five managed colonies of *A. cerana* maintained by the Southwest Biodiversity Laboratory, Chinese Academy of Sciences in Kunming, Yunnan, were used in the experiment. Three plastic containers, each with a single food source [sucrose solution (30%, wt/vol), water, or pollen], were placed next to each other and 5 m away from the beehives. Bees caught on the sucrose solution were marked with nontoxic yellow paint on the tergum of the mesothorax; those foraging on water were marked with nontoxic red paint. Pollen collectors were identified by the presence of pollen pellet in corbiculae. Labeled bees were released and observed in revisits to confirm those foraging on water were marked with nontoxic red paint. Pollen collectors were identified by the presence of pollen pellet in corbiculae. Labeled bees were released and observed in revisits to confirm task division. Pollen, nectar, and water collectors were captured 1 day later to test for gene expression. RNA was extracted from antennae and the brain immediately after dissection. The qPCR experiments followed the same protocols described above.

**RNAi and PER to sucrose**

We carried out RNAi experiments to examine the role of the LOC107998395 gene in sensing sucrose. The partial fragment of the gene amplified using specific primers from cDNA was cloned into the pCE2-TA-Blunt-Zero vector (Vazyme, China) and was verified by Sanger sequencing. The 373-bp-long fragment was amplified from the plasmid using specific primers with a T7 promoter and then used for double-stranded RNA (dsRNA) synthesis using the MEGAscript RNAi Kit (Invitrogen, USA). A 385-bp-long fragment amplified from the green fluorescent protein (GFP) gene plasmid (MH423581.1) was used as the control. The primers used above are shown in table S8. Nanocarriers were mixed gently with dsRNA at a recommended mass ratio of 1:1 to improve the efficiency of RNAi in feeding.

Solution mixtures were made for different treatments: (i) mixture I (Lkr-dsRNA/nanocarrier/sucrose solution) as the RNAi treatment; (ii) mixture II (GFP-dsRNA/nanocarrier/sucrose solution) as an intact dsRNA control; (iii) sucrose solution only as the untreated control. The final concentrations for dsRNA + nanocarrier and sucrose were 500 ng/μl and 50% (wt/vol), respectively. Bees in each treatment were fed on the mixture in an incubator at 35°C for 7 days and were subsequently used for the following tests. Each bee ingested about 15 μg of dsRNA per day.

Foragers used for PER experiments were collected from a colony at the Miyun apiary, Beijing. Before the test, bees were starved for 12 hours by removing sugar syrup from the incubator. Bees were then anesthetized at 4°C, fixed to modified 2-ml centrifuge tubes with Parafilm M (Bemis), and tested for sucrose responses. In the first experiment, a drop of sucrose solution contained in a syringe touched the antennae of the tested bee. Sucrose solutions following a concentration series at 0.1, 0.3, 1, 3, 10, 30, and 50% (wt/vol) were applied from low to high concentration. Distilled water was used between each sucrose application to avoid sensitization effects. Consecutive tests were made at 4-min intervals. In the second experiment, a toothpick dipped in sucrose solution was carefully positioned closest to the antennae for 5 s without touching them. Distilled water and sucrose solutions were applied from low to high concentration, following the same concentration series. The lowest sucrose concentration eliciting PER was considered as the response threshold.

**Strand-specific lncRNA sequencing and analysis**

Two replicate bee samples were prepared for IncRNA sequencing. Each sample contained 20 nurse bees of *A. cerana* collected from two hives in Miyun, Beijing, in August 2019. Guts were removed and the remaining bodies were pooled for RNA sequencing. Total RNA was extracted with TRIzol (Thermo Fisher Scientific), and ribosomal RNA (rRNA) was removed with a Globin-Zero Gold rRNA removal kit (Epicenter). The remaining RNA was fragmented, and a strand-specific library was built with the NEBNext Ultra Directional RNA Library Prep Kit (New England Biolabs). Sequencing was carried out on an Illumina HiSeq X Ten (insert size 240 bp, 150 PE reads) at Novogene, Beijing, which produced ca. 25 Gb and 21 Gb data for the two samples, respectively. Reads were aligned to the *A. cerana* reference genome (ACSNV-2.0, GCF_001442555.1) using HISAT2 2.1.0 with parameter “--rna-strandness RF” (63).

**SUPPLEMENTARY MATERIALS**

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/6/51/eabdb390/DC1

View/request a protocol for this paper from Bio-protocol.

**REFERENCES AND NOTES**

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

2. F. C. Jones, M. G. Grabherr, Y. F. Chan, P. Russell, E. Mauceli, J. Johnson, R. Swofford, M. Pirun, M. C. Zody, S. White, E. Birnies, S. Searle, J. Schmutz, J. Grimmerwood, M. C. Dickson, R. M. Myers, C. T. Miller, B. R. Summers, A. K. Knecht, S. D. Brady, H. Zhang, A. A. Pollen, T. Howes, C. Amemiya; Broad Institute Genome Sequencing Platform; Whole Genome Assembly Team, E. S. Lander, F. D. Palma, K. Lindblad-Toh, D. M. Kingsley, The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484, 55–61 (2012).

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

4. D. P. Abrol, Asiatic Honeybee Apis cerana: Biodiversity Conservation and Agricultural Production (Springer, 2013).

5. Y. Yu, S. Zhou, X. Zhu, X. Xu, W. Wang, L. Zha, P. Wang, J. Wang, K. Lai, S. Wang. Genetic differentiation of eastern honey bee (Apis cerana) populations across Qinghai-Tibet plateau-valley landforms. *Front. Genet.* 10, 483 (2019).
20. N. J. Lemanski, C. N. Cook, B. H. Smith, N. Pinter-Wollman, A multiscale review of behavioral variation in collective foraging behavior in honey bees.

21. M. L. Winston, The Biology of the Honey Bee (Harvard Univ. Press, 1991).

22. C. Darwin, On the Origin of Species by means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life (John Murray, 1860).

23. A. Wallberg, F. Han, G. Wellhagen, B. Dahle, M. Kawata, N. Haddad, Z. L. P. Simões, M. H. Alssopp, I. Kanderim, P. De la Rúa, C. W. Pirk, M. T. Webster, A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honey bee Apis mellifera. Nat. Genet. 46, 1081–1088 (2014).

24. M. D. Jennings, G. M. Harris, Climate change and ecosystem composition across large landscapes. Landsc. Ecol. 32, 195–207 (2017).

25. E. E. Cleland, I. Chune, A. Menzel, H. A. Mooney, M. D. Schwartz, Shifting plant phenology in response to global change. Trends Ecol. Evol. 22, 357–365 (2007).

26. G. A. Wray, The evolutionary significance of cis-regulatory mutations. Nat. Rev. Genet. 8, 206–216 (2007).

27. I. Gonzalez, R. Munita, E. Agirre, T. A. Dittmer, K. Gysling, T. Mielst, R. F. Luco, A IncRNA regulates alternative splicing via establishment of a splicing-specific chromatin signature. Nat. Struct. Mol. Biol. 22, 370–376 (2015).

28. R.-W. Yao, Y. Wang, L.-L. Chen, Cellular functions of long noncoding RNAs. Nat. Cell Biol. 21, 542–551 (2019).

29. S. Lamicz, L. Jiang, B. Sim, M. Almén, K. Maqbool, M. Grabher, A. Martinez-Barrio, M. Promerová, C.-J. Rubin, C. Wang, N. Zarnami, B. R. Grant, P. R. Grant, M. T. Webster, L. Andersson, Evolution of Darwin’s finches and their beaks revealed by genome sequencing. Nature 518, 371–375 (2015).

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

31. G. V. Amadon, K. Norberg, M. K. Fondrk, R. E. Page Jr., Reproductive ground plan may mediate colony-level selection effects on individual foraging behavior in honey bees. Proc. Natl. Acad. Sci. U.S.A. 101, 11350–11355 (2004).

32. D. R. Nääs, M. Zandawala, Recent advances in neuropeptide signaling in Drosophila, from genes to physiology and behavior. Prog. Neurobiol. 79, 101607 (2019).

33. A. A. Anzai, A. Emard, P. Nagamei, M. Ozszeski, V. Sapin, C. Waters, K. Zinn, R. J. Wyman, S. Benaz, The Leucokinin pathway and its neurones regulate meal size in Drosophila. Curr. Biol. 20, 969–978 (2010).

34. C. R. Smith, A. L. Tóth, A. V. Suarez, G. E. Robinson, Genetic and genomic analyses of the division of labour in insect societies. Nat. Rev. Genet. 9, 735–748 (2008).

35. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: An ultra-fast all-in-one FASTQ preprocessing tool. Bioinformatics 34, 884–890 (2018).

36. D. Park, J. W. Jung, B.-S. Choi, M. Jayakodi, J. Lee, J. Lim, Y. Yu, Y.-S. Choi, M.-L. Lee, Y. Park, J.-Y. Choi, T.-J. Yang, O. R. Edwards, G. Nah, H. W. Kwon, Uncovering the novel characteristics of Asian honey bee, Apis cerana, by whole genome sequencing. BMC Genomics 16, 1 (2015).

37. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint arXiv:1303.3997 (2013).

38. M. David, M. Dzamba, D. Lister, L. Ilie, M. Brudno, SHRIMP2: Sensitive yet practical short read mapping. Bioinformatics 27, 1011–1012 (2011).

39. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altschuler, S. Gabriel, M. Daly, M. A. De Pristo, The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010).

40. A. Manichakul, J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale, W.-M. Chen, Robust relationship inference in genome-wide association studies. Bioinformatics 26, 2867–2873 (2012).

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

42. E. Paradis, K. Schliep, ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528 (2018).

43. M. Neuvelitschko, M. S. Khattar, H. W. Raadsma, NetView: A high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. PLOS ONE 7, e48375 (2012).

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

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

46. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetic parameter estimation from sequencing data. Bioinformatics 27, 2987–2993 (2011).

47. J. W. Leigh, D. Bryant, poptar: Full-feature software for haplotype network construction. Methods Ecol. Evol. 6, 1110–1116 (2015).

48. J. Catchen, P. A. Hohenlohe, S.巴斯, A. Amores, W. A. Cresko, Stacks: An analysis tool set for population genomics. Mol. Ecol. Res. 11, 353–357 (2011).

49. S. J. Phillips, R. P. Anderson, R. E. Schapire, Maximum entropy modeling of species geographic distributions. Ecol. Model. 190, 231–259 (2006).

50. R. Muscarella, P. J. Galante, M. Soley Guardia, R. A. Boria, J. M. Kass, M. Uriarte, R. P. Anderson, ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. Methods Ecol. Evol. 5, 1198–1205 (2014).

51. P. Danecak, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. De Pristo, R. E. Handsaker, G. Lunter, G. T. Marsh, S. T. Sherry, G. M. Vean, R. Durbin, 1000 Genomes Project Analysis Group, The variant call format and VCFtools. Bioinformatics 27, 2156–2158 (2011).

52. P. Pavlidis, D. Živković, A. Stamatakis, N. Lachiotis, Sweed: Likelihood-based detection of selective sweeps in thousands of genomics. Mol. Biol. Evol. 30, 2224–2234 (2013).

53. G. Yu, L.-G. Wang, Y. Han, Q.-Y. He, clusterProfiler: An R package for comparing biological themes among gene clusters. OMICS 16, 284–287 (2012).

54. A. M. Harris, M. DeGiorgio, A likelihood approach for uncovering selective sweep signatures from haplotype data. bioRxiv 678722 (2019).

55. O. Mirabeau, J.-S. Joly, Molecular evolution of peptide signaling systems in bilaterians. Proc. Natl. Acad. Sci. U.S.A. 110, E2028–E2037 (2013).

56. K. Kato, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30, 772–780 (2013).

57. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAL: A tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973 (2009).

58. M. N. Price, P. S. Dehal, P. A. Arkin, FastTree 2—Approximately maximum-likelihood trees for large alignments. PLOS ONE 5, e94940 (2010).

59. D. Kim, B. Langmead, S. L. Salzberg, HISAT: A fast spliced aligner with low memory requirements. Nat. Methods 12, 357–360 (2015).
acknowledgments: We acknowledge sample contributions from a great number of collaborators, including Fengchen Ge, Yunbo Xue, Zhiyong Li, Zhi Wang, Jialong Huang, John J. Wilson, Shixiao Luo, Yueshiong Luo, Yanqiong Peng, Zhisheng Zhang, Chaodong Zhu, Dan Zhang, Qingtao Wu, Shunhai Wang, Hu Li, Syed I. A. Shah, and Sing Kong Wah among many others. Thanks to Weewei Li and Kaikin Li of the Kunming Natural History Museum of Zoology, Kunming Institute of Zoology, CAS, who have made important specimens available to this study. Thanks to Wei Zhang of Peking University for the constructive suggestions in the early stage of the study. Ming Bai of the Institute of Zoology, Chinese Academy of Sciences, provided guidance on wing imaging and analyses. Jie Shen and Shuo Yan of China Agricultural University lent materials and expertise on nanocarriers used in RNAi experiments. Changsheng Ma assisted in Lkr expression profiling of foraging division. Colleagues from BGI-Qingdao provided important technical support on DNA preparation and sequencing on the BGI-seq platform. Thanks to the two anonymous reviewers for constructive comments that have helped improve the manuscript significantly. A special dedication to Professor Fengchen Ge, a pioneer researcher on the Chinese honeybee and former head of the Institute of Jilin Apiary Research. Funding: The work was supported by the Program of Ministry of Science and Technology of China (2018FY100403), National Special Support Program for High-Level Talents (Ten-Thousand Talents Program), National Natural Science Foundation of China (No. 31772493), and funding from the Beijing Advanced Innovation Center for Food Nutrition and Human Health through China Agricultural University grant to Xin Zhou. Sample collection was also supported by the NSF of China (no. 31470123) and Jilin Science and Technology Program (no. 200330561) to X.L. S.H.P. is thankful to the National Mission on Himalayan Studies (NMHS)–Almora, Ministry of Environment, Forest and Climate Change, Government of India, for providing financial assistance with reference no. GBPNI/NMHS-2017-18/MG-12 dated 26 February 2018. Author contributions: Xin Zhou, T.J., Q.N., and X.L. designed the study. T.J., X.L., Q.N., and Xin Zhou organized sampling. Y.J. led the analyses. J.T. conducted qPCR and RNAi experiments. L.Q. performed morphometrics analysis. J.H. constructed the historical habitat range. J.D. built the gene tree for the GPCR families. S. Liu contributed to pipeline optimization and data interpretation. P.B.F. contributed in sample collection, revised the manuscript, and helped provide interpretation of the results. S. Luo supervised the literature review of relevant functional genes and conducted analysis on gene regulation. S.H.P. contributed knowledge on A. cerana populations from southeastern Asia. L.L. organized the sampling in Aba. Y.J., Xin Zhou, and Xuguo Zhou wrote the first drafts, and all authors contributed to and proofed the manuscript. Competing interests: The authors declare that they have no competing interests. Data and materials availability: All A. cerana specimens used in this study were collected from legitimate resources, abiding local laws. All the sequences reported in this study are deposited into the NCBI SRA under the accession number PRJNAS92293. All software used is already published and/or available in the public domain and cited in the manuscript.

Submitted 17 June 2020
Accepted 30 October 2020
Published 18 December 2020
10.1126/sciadv.abd3590

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