Received: 20 September 2018 | Revised: 5 December 2018 | Accepted: 7 December 2018

DOI: 10.1111/mec.14986

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

The genomic basis of adaptation to high-altitude habitats in the eastern honey bee (*Apis cerana*)

Santiago Montero-Mendieta[1*] | Ken Tan[2*] | Matthew J. Christmas[3] | Anna Olsson[3] | Carles Vilà[1] | Andreas Wallberg[3] | Matthew T. Webster[3]

[1]Conservation and Evolutionary Genetics Group, Doñana Biological Station (EBD-CSIC), Seville, Spain

[2]Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Kunming, China

[3]Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

Correspondence
Matthew T. Webster, Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden.
Email: matthew.webster@imbim.uu.se

Funding information
Svenska Forskningsrådet Formas, Grant/Award Number: 2013-722; Vetenskapsrådet, Grant/Award Number: 2014-5096; china national research fund, Grant/Award Number: 31770420; Chinese Academy of Sciences, Grant/Award Number: 2017XTBG-T01 and Y4ZK111B01

Abstract

The eastern honey bee (*Apis cerana*) is of central importance for agriculture in Asia. It has adapted to a wide variety of environmental conditions across its native range in southern and eastern Asia, which includes high-altitude regions. eastern honey bees inhabiting mountains differ morphologically from neighbouring lowland populations and may also exhibit differences in physiology and behaviour. We compared the genomes of 60 eastern honey bees collected from high and low altitudes in Yunnan and Gansu provinces, China, to infer their evolutionary history and to identify candidate genes that may underlie adaptation to high altitude. Using a combination of F_{ST}-based statistics, long-range haplotype tests and population branch statistics, we identified several regions of the genome that appear to have been under positive selection. These candidate regions were strongly enriched for coding sequences and had high haplotype homozygosity and increased divergence specifically in highland bee populations, suggesting they have been subjected to recent selection in high-altitude habitats. Candidate loci in these genomic regions included genes related to reproduction and feeding behaviour in honey bees. Functional investigation of these candidate loci is necessary to fully understand the mechanisms of adaptation to high-altitude habitats in the eastern honey bee.

KEYWORDS
altitude adaptation, genetic differentiation, honeybees, local adaptation, positive selection, selective sweeps

1 | INTRODUCTION

Populations are exposed to different selection pressures across a species' range depending on the local environments. Populations that have higher fitness in their native sites than elsewhere are considered locally adapted (Kawecki & Ebert, 2004; Leimu & Fischer, 2008; Whitlock, 2015). The study of local adaptation is of fundamental importance for understanding how species evolved and diversified and to predict the effects of environmental change on species distributions. Identifying genetic regions and variants involved in local adaptation can help design conservation plans for endangered species, as well as providing insights into population structure, admixture and evolutionary history (Allendorf, Hohenlohe, & Luikart, 2010). Genomic data have the power to greatly advance our understanding of local adaptation, and whole-genome sequencing (WGS) on a population scale enables identification of specific genetic variants that govern adaptive differences among natural populations (Radwan & Babik, 2012).

High-altitude habitats present unique conditions to which organisms must adapt, such as lower temperatures, higher

*Authors contributed equally.

radiation, lower partial pressure of oxygen and different flora and fauna, compared to lowland areas (Cheviron & Brumfield, 2012). Populations living in these environments are often closely related to nearby lowland populations, which facilitates identifying adaptive genetic differences between them using genomic comparisons. Human populations living at high altitudes have been found to possess genetic variants that govern physiological adaptations to hypoxia. These variants differ between high-altitude populations from three continents, indicating differences in the molecular mechanisms underlying these adaptations (Bigham et al., 2010; Huerta-Sánchez et al., 2013; Ilardo & Nielsen, 2018; Peng et al., 2011; Xu et al., 2011).

Insects living at high altitudes often show numerous phenotypic adaptations (Dillon, 2006). However, few studies have investigated genomic regions underlying these adaptations. In a recent study, populations of the western honey bee (*Apis mellifera*) from the mountains of East Africa were shown to have two chromosomal inversions that are likely involved in adaptation to high-altitude habitats (Wallberg, Schöning, Webster, & Hasselmann, 2017). These inversions span 2.2 Mb (mega bases) in total and contain 88 genes. It is possible that multiple adaptive variants in these genes contribute to adaptation. Chromosomal inversions can facilitate local adaptation by protecting co-adapted alleles from recombination with maladaptive ones (Kirkpatrick & Barton, 2006). Candidate genes within these regions include some known to be associated with foraging behaviour in honey bees, which could indicate that the abundance and distribution of food supplies are an important factor mediating adaptation to high-altitude habitats. There are now many examples, from a wide range of species, of chromosomal inversions governing local adaptation, including several in insects (Wellenreuther & Bernatchez, 2018). However, it is unclear whether chromosome inversions are likely to govern adaptation to altitude in other related bee species, or if similar genes or molecular pathways are involved.

Pollination services provided by bees are essential for food production worldwide. Several bee species are kept in managed colonies, which are important both for commercial honey production and for crop pollination (Wang et al., 2012). The eastern honey bee *A. cerana* has a large distribution across East Asia and is adapted to a wide range of environmental conditions (Koetz, 2013). *A. cerana* is commonly used in beekeeping in Asia, but populations have declined since the early 20th century due to changes in local agriculture practices and the introduction of *A. mellifera* (Chen et al., 2017; Theisen-Jones & Bienefeld, 2016; Xu, Shi, & Chen, 2009). Understanding how populations of this species are adapted to particular environmental conditions can inform conservation efforts and population management (Parker et al., 2010). *Apis cerana* occurs naturally in deciduous forests, tropical rainforests, steppes and taigas, low-altitude grasslands and high-altitude mountain environments (Koetz, 2013). Across these habitats, populations of *A. cerana* exhibit different morphological traits. Phylogeographic analyses based on morphometric data and mtDNA data have partitioned this species into five or six clusters, which are related to climatic and geographical factors (Radloff et al., 2010; Smith, Villafuerte, Otis, & Palmer, 2000). Populations from China appear to all belong to the same cluster, although there is variation in morphology within this cluster.

In the Yunnan Province, China, populations of *A. cerana* occurring higher than 2,000 metres (m) are larger, darker and have longer body hair than lowland populations (Pereboom & Biesmeijer, 2003; Tan, Fuchs, Koeniger, & Ruiguang, 2003; Tan & Ling-juan, 2008). Similar phenotypic morphological differences are also observed in high-altitude populations of *A. mellifera* (Gruber, Schöning, Otte, Kinuthia, & Hasselmann, 2013; Smith, 1961) and likely represent adaptations for these habitats. Physiological and behavioural differences may also be involved in this local adaptation. In a recent study based on WGS, *A. cerana* populations sampled across China were found to have limited levels of population structure, with the highest levels of genome differentiation and lower genetic diversity observed in isolated mountain or island populations (Chen et al., 2018), indicating a greater degree of reproductive isolation. However, the molecular basis of the phenotypic differences exhibited by highland populations is unknown.

Honey bees are particularly amenable to population genomic studies due to their relatively small genome size. The draft genome of *A. cerana* (Park et al., 2015) spans approximately 230 Mb and is fragmented into 2,430 scaffolds with N50 equal to 1.4 Mb. Here, we compared the genomes of 60 *A. cerana* worker bees from six areas in China, covering a wide range of elevations, from 67 to almost 2,800 m. We explored the genomic differences between highland and lowland *A. cerana* populations using a combination of statistical approaches to scan the genome based on relative divergence, haplotype length and population branch statistics (PBSs). We aimed to identify regions in the genome that are divergent between highland and lowland populations, which may be candidates for local adaptation to high-altitude habitats.

2 | MATERIAL AND METHODS

2.1 | Sampling

We sampled 60 bees from six different localities within China: five areas in Yunnan province (Deqin, Lijiang, Kunming, Hekou and Xishuangbanna) and one area in Gansu province (Tianshui) (Figure 1). All the bees were female diploid workers from unrelated colonies. Sampling localities encompassed a wide range of altitudes and we classified them as "highland" populations if they were found >2,000 m above sea level, "intermediate" populations if they were found between 1,000 and 2,000 m, and "lowland" populations if they were found <1,000 m. Based on these ranges, our study includes: two highland populations, Deqin (2,714 m; $N = 10$ samples) and Lijiang (2,485 m; $N = 10$); two intermediate populations, Kunming (1,800 m; $N = 5$) and Tianshui (1,100 m; $N = 10$); and two lowland populations, Xishuangbanna (570 m; $N = 10$) and Hekou (67 m; $N = 15$). The abdomens of each bee were photographed under identical conditions and the photographs were assessed by eye for the degree of coloration in each sample, using a scale of 1 (light) to 5 (dark), in random order. For some of our analyses, we also included previously published data

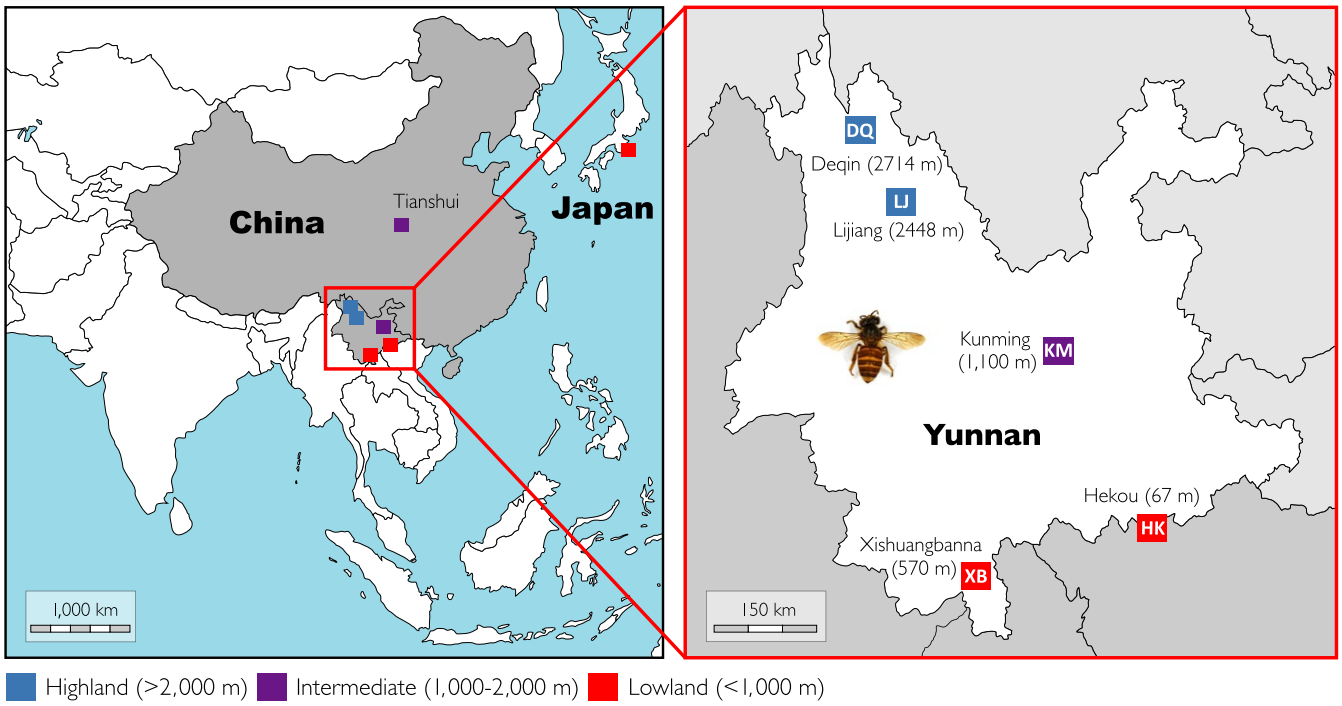

FIGURE 1 Sampling localities of *Apis cerana* in this study. The red square shows in more detail the five sampling localities in Yunnan province, China. The elevation in metres (m) is shown next to the name of each locality [Colour figure can be viewed at wileyonlinelibrary.com]

from 10 distantly related *Apis cerana* specimens sampled from multiple locations throughout Japan (Wallberg et al., 2014).

2.2 | DNA extraction, sequencing, SNP calling and haplotype phasing

We extracted genomic DNA from head tissue of bees using a salt-ethanol precipitation protocol as described in Nelson, Wallberg, Simões, Lawson, and Webster (2017). Genomic libraries were prepared with the TruSeq DNA PCR-free Library Prep Kit (Illumina Inc.) and sequenced by the SciLifeLab platform (Stockholm University) on a HiSeq2500 instrument (Illumina Inc.) using the v4 chemistry according to the manufacturer's protocols. We mapped all the reads against the *A. cerana* reference genome (Park et al., 2015) using the BURROWS–WHEELER ALIGNER (BWA-MEM) v0.7.5 (Li & Durbin, 2009). For outgroup data, we downloaded 75 bp SOLiD DNA fragment libraries of 10 Japanese *A. cerana* bees (Wallberg et al., 2014) from NCBI. We mapped the SOLiD data against the same *A. cerana* reference sequence with LIFESCOPE™ v2.5.0/v2.5.1 (Life Technologies). We sorted the resulting bam files using SAMTOOLS (Li et al., 2009).

We used PICARD v1.118 (http://broadinstitute.github.io/picard) to tag duplicate reads and assign read group tags to the mapped libraries. The GENOME ANALYSIS TOOLKIT (GATK v3.3.0) (McKenna et al., 2010) was used to perform joint indel realignments across all mapped data, including the outgroup. All these programs were used with default settings. We next called single-nucleotide polymorphisms (SNPs) across all samples using the haplotype-based variant detector FREEBAYES v0.9.20-16 (Garrison & Marth, 2012). We used the flags "-E 0," "-X" and "-u" to suppress construction of short multi-nucleotide haplotypes from closely positioned polymorphisms and to avoid making composite polymorphisms. We used "--theta 0.004" as the expected population mutation rate. We kept biallelic SNPs with a quality score of 50 or greater (QUAL ≥ 50) for further analysis. We used BEAGLE v3.3.2 (Browning & Browning, 2007) to phase haplotypes and infer missing genotypes. We used the flags "niterations=10" and "lowmem=true" to increase accuracy and reduce memory usage as recommended in the program manual. We performed a quality assessment to ensure that the combination of data from two sources, Illumina and SOLiD, did not lead to biases in our results (Cridland, Tsutsui, & Ramírez, 2017) (see Supporting Information Appendix S1, Figures S1–S3). All the commands used for data analysis are available on Supporting Information Appendix S2.

2.3 | Genetic diversity and population structure

To determine genetic diversity within populations, we computed Watterson's theta (θ_w; Watterson's estimator of the population mutation rate per base) (Watterson, 1975) for each population across the whole genome. This was estimated using the full set of SNP genotypes, with missing genotypes imputed, produced as the output of BEAGLE. An estimate of θ_w per base pair was calculated by dividing θ_w by the total number of base pairs in the reference sequence to which short reads were mapped. These calculations were performed using a custom Perl script. As our main aim was to compare highland and lowland honey bees, we also computed θ_w pooling the two high-altitude populations (Deqin and Lijiang) and the two lowland populations (Xishuangbanna and Hekou). We conducted a kinship assessment for each population using the program PLINK (Purcell et

al., 2007) with the "--genome" flag. As this calculation is not LD-sensitive, we did LD-based pruning using the same parameters as Chen et al. (2018) (–indep 50 5 1.1). We assessed population genetic structure across all populations using two different clustering programs, SNMF (Frichot, Mathieu, Trouillon, Bouchard, & François, 2014) and ADMIXTURE v1.3.0 (Alexander, Novembre, & Lange, 2009) in addition to a principal component analysis (PCA) with the function "glPca" from the ADEGENET R package (Jombart, 2008). These programs use different algorithms to cluster genetically similar individuals together. To determine the number of population clusters (*K*) in SNMF and ADMIXTURE, we ran 10 independent analyses using the flag "-K" from *K* = 2 to *K* = 7 with default parameters except for generating a pseudo-random seed using the system clock in ADMIXTURE (option "-s time"). We analysed the cross-validation errors (CV-errors) at each value of *K* to find the most likely *K* and combined the results from replicate runs using CLUMPP (Jakobsson & Rosenberg, 2007) with default settings.

2.4 | Scans for selection

To identify specific regions of the genome that could be under differential selection between highland (H: Deqin and Lijiang; *N* = 20) and lowland (L: Xishuangbanna and Hekou; *N* = 25) honey bees, we used a combination of allele frequency and haplotype structure analyses (outlined below). In order to determine whether outlier loci were associated with elevation rather than being associated with population genetic structure, we compared the previous results with those obtained for other possible combinations of the four populations: (a) Deqin (H) + Hekou (L) versus Lijiang (H) + Xishuangbanna (L); (b) Deqin (H) + Xishuangbanna (L) versus Lijiang (H) + Hekou (L).

2.4.1 | F_{ST} analysis

We calculated the fixation index (F_{ST}) at every SNP between highland and lowland bees using the Weir–Cockerham estimator (Weir & Cockerham, 1984) using a custom Perl script. This estimator is based on the allele frequencies at each SNP and sample sizes in each population, assuming random union of gametes as described in Weir and Cockerham (1984). We used the F_{ST} estimator of Reynolds, Weir, and Cockerham (1983) to infer the average genetic divergence across nonoverlapping 1-kbp windows, as well as the mean F_{ST} across the whole genome, using a custom Perl script. Estimates of F_{ST} were combined from multiple SNP loci in the same window (or across the whole genome) using the weighted average estimator described in Reynolds et al. (1983).

2.4.2 | XP-EHH analysis

To compare haplotype length and detect changes in haplotype diversity, we ran the cross-population extended haplotype homozygosity (XP-EHH) test (Sabeti et al., 2007) in the program SELSCAN v1.2.0 (Szpiech & Hernandez, 2014) for every SNP with a minor allele frequency ≥0.01 (option "--maf 0.01"). Since recombination is expected to disrupt linkage over time, this analysis allowed us to test whether F_{ST} outliers occurred on long haplotypes of linked variants, which would be suggestive of relatively recent selective sweeps. This test also indicates which of a pair of populations has the strongest signal of selection on a divergent SNP. In order to classify the two variants at every SNP as either ancestral or derived and format the data for SELSCAN, we first produced and analysed a pairwise alignment between the *A. cerana* (Park et al., 2015) and *Apis mellifera* (Elsik et al., 2014) reference sequences using the SATSUMA whole-genome synteny program (Grabherr et al., 2010) with default settings. We applied parsimony to classify the alleles: the variant that was shared between the sister species was taken as ancestral, whereas the other was taken as derived. Because a high-resolution genomic linkage map is missing for *A. cerana*, we applied a fixed recombination rate of 18 cM/Mbp across all data, which is close to the estimated average of this species (17.4 cM/Mbp; Shi et al., 2013). The XP-EHH score was not computed for short scaffolds with <10 SNPs.

2.4.3 | PBS analysis

To detect genomic regions that have exclusively accumulated changes on highland compared to lowland bees, we implemented the PBS as described in Yi et al. (2010). This statistic uses two focal populations and an outgroup population to find alleles with extreme frequencies that diverged rapidly within them. We used the Japanese samples, which were found to be highly diverged from the other samples (see below), as an outgroup. To do this, we first converted the 1-kb window Reynolds-F_{ST} estimates into divergence times (T) following the procedure in Yi et al. (2010) and Wallberg, Pirk, Allsopp, and Webster (2016). By comparing the divergence times between two Chinese populations to the corresponding divergence times between each population and the Japanese outgroup, the PBS allowed identification of divergent regions in which genetic changes are associated specifically with drift or selection in the highland bees (Wallberg et al., 2016).

2.4.4 | CSS analysis

We combined the F_{ST} and XP-EHH estimates into a single unbiased composite selection score (CSS) in order to identify genomic regions with the most comprehensive signals of selection, following Randhawa, Khatkar, Thomson, and Raadsma (2014). To compute the CSS, all SNPs were ranked for each statistic. For each SNP, the fractional rank positions for F_{ST} and XP-EHH, respectively, were converted into two *z*-statistics, the mean of which corresponded to a single joint rank for both statistics. The corresponding *p*-value was retrieved from a standard normal distribution. The CSS score was then taken as $-\log_{10}p$ as in Wallberg et al. (2016). We did not include PBS in this score as it is expected to correlate with F_{ST} under neutrality (Wallberg et al., 2016). However, a correlation between F_{ST} and XP-EHH is not expected under neutrality (Eden, Navon, Steinfeld, Lipson, & Yakhini, 2009).

2.4.5 | Genotype–environment association

To search for a possible association between environment and allele variation across all populations, we ran the program SAMBADA v0.5.3 (Stucki, Orozco-terWengel, Forester, Duruz, & Colli, 2016) that applies logistic regression models. We used altitude as the environmental variable and we considered the model as significant when both the *G*-score and Wald-score were ≤0.01 after Bonferroni correction (parameter file as follows: "HEADERS YES; NUMVARENV 2; NUMMARK 10000; NUMINDIV 60; DIMMAX 1; SAVETYPE END ALL 0.01").

2.5 | Analysis and identification of candidate genes

We used the original GFF gene models from the *A. cerana* reference genome (Park et al., 2015) to associate the SNPs with genes and annotate them as coding, UTRs, intronic or intergenic. To generate a reference gene list that could be used for gene ontology (GO) enrichment analyses, we searched the genome of the fruit fly (*Drosophila melanogaster*) from FLYBASE (dmel-all-genes-r6-17) for homologs of the candidate gene sequences using TBLASTN with an *E*-value ≤1×10^{-5}. As a target gene list, we selected the corresponding gene orthologs of *D. melanogaster* for the top 1,000 SNPs sorted by CSS score. We used WEBGESTALT (Wang, Vasaikar, Shi, Greer, & Zhang, 2017) with default settings to perform non-redundant GO enrichment between the target and reference gene lists. Our search was restricted to those genes with orthologs in *D. melanogaster* because GO information is not available for any species within the genus *Apis*.

3 | RESULTS

3.1 | Genetic diversity, population structure and differentiation

We detected 5,871,875 SNPs across the 60 individuals corresponding to an average genetic diversity level of θ_w = 0.0047. The level of genetic diversity within each population was similar between all the populations (θ_w = 0.0028–0.0039; Table 1). These levels of variation are similar to European, but lower than African populations of *Apis mellifera* (Wallberg et al., 2014). The lowest genetic diversity was found in the population at the highest altitude (Deqin).

A neighbour-joining tree based on F_{ST} distances indicated that the Chinese populations were very differentiated from the Japanese one, with no clear indication of substructure within the Chinese samples (Figure 2a). The two lowland populations from Yunnan (Xishuangbanna and Hekou) have the lowest genetic differentiation (F_{ST} = 0.043) (Figure 2b). The population from Deqin has the largest genetic differentiation from other populations (Figure 2b). The average kinship relatedness as estimated by the PLINK's "PI_HAT" statistic was very low within each population (DQ = 0.049; LJ = 0; KM = 0.052; TS = 0.026; HK = 0; XB = 0.029).

Population structure analyses with SNMF and ADMIXTURE showed that the data for the samples from China were most consistent with the presence of two clusters (*K* = 2), as inferred by the mean cross-validation error across 10 replicate runs (Figure 3a; Supporting Information Figure S4). The Deqin highland samples (DQ) formed a separate cluster compared to all other populations at this value of *K*. At *K* = 4, other clusters appear clearly differentiated as well. A PCA (Figure 3b) also showed that the Deqin (DQ) population was the most divergent, separating from the other populations along the first axis (PC1). However, the second axis (PC2) accounted for a similarly large proportion of the total variance and showed that the individuals from Tianshui (TS) and some of the individuals from Xishuangbanna (XB) were also clearly separated from the rest. These patterns of differentiation do not appear to correspond with altitude or geographical proximity. The other highland population, Lijiang (LJ), which is located at an altitude just 229 m lower and 119 km away from Deqin (DQ), was very different and seemed genetically more similar to lowland populations. Similarly, the individuals from the lowland populations XB and HK did not form a unique cluster in the PCA analysis. These results imply that the highland and lowland populations sampled here do not represent separate evolutionary lineages.

We evaluated the shade of coloration of the abdomen of each sequenced specimen by eye. There was a significant trend for specimens to have darker coloration in the higher altitude populations (*t* test comparison of highland vs. lowland and intermediate altitude populations, *p* = 0.002), although there was also a large amount of

TABLE 1 Summary of *A. cerana* populations included and their genetic diversity (θ_w)

Code	Locality	Altitude (m)	Number of samples	Variable SNPs	θ_w
DQ	Deqin (Yunnan)	2,714	10	2,286,040	0.0028
LJ	Lijiang (Yunnan)	2,485	10	2,850,497	0.0035
KM	Kunming (Yunnan)	1,800	5	1,945,203	0.0030
TS	Tianshui (Gansu)	1,100	10	2,570,782	0.0032
XB	Xishuangbanna (Yunnan)	570	10	2,880,465	0.0036
HK	Hekou (Yunnan)	67	15	3,569,169	0.0039
DQ + LJ	HIGHLAND	–	20	3,544,415	0.0036
XB + HK	LOWLAND	–	25	4,322,498	0.0042
Total	ALL POPULATIONS	–	60	5,871,875	0.0047

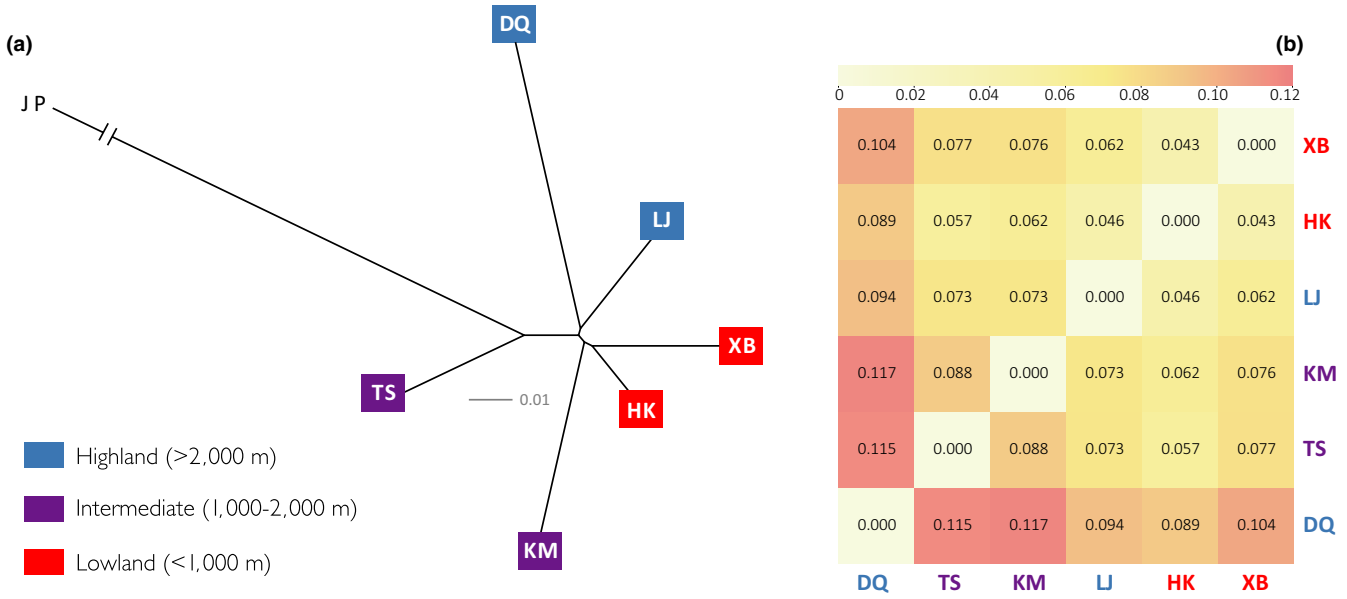

FIGURE 2 (a) Neighbour-joining tree based on F_{ST}. (b) Genetic differentiation (F_{ST}) of *Apis cerana* populations. In both plots, populations are defined as follows: JP: Japan; DQ: Deqin; LJ: Lijiang; KM: Kunming; TS: Tianshui; XB: Xishuangbanna; HK: Hekou

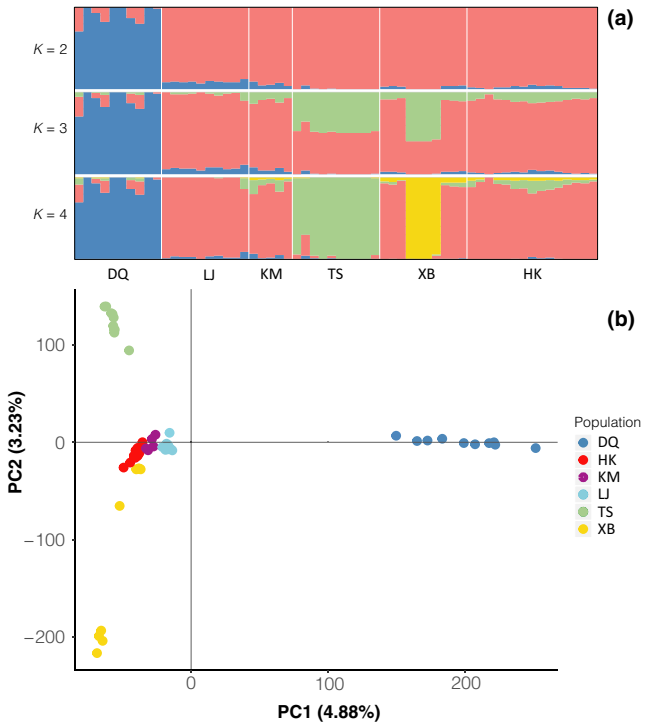

FIGURE 3 (a) Population structure for $K = 2$ to $K = 4$ for the 60 *Apis cerana* samples based on 5.8 million SNPs using the program SNMF. (b) Principal component analysis (PCA) based on 5.8 million SNPs using the ADEGENET package in R. In both plots, populations are defined as follows: DQ: Deqin; LJ: Lijiang; KM: Kunming; TS: Tianshui; XB: Xishuangbanna; HK, Hekou

variation in coloration within populations (Supporting Information Figure S5 and Table S5). Genetic structure is observed in the lowland XB population, which appears to correlate with coloration. At $K = 4$, two distinct blocks are seen in this population (Figure 3a, Supporting Information Figure S4). One of these blocks comprises four samples (01:16, 01:17, 01:18 and 01:20) that are darker than the other six sampledones, which could indicate substructure in this population.

3.2 | Evidence of selection in highland populations

To search for evidence of directional selection in high-altitude honey bee populations, we compared the two highland (H) populations (Deqin and Lijiang) with the two lowland (L) populations (Xishuangbanna and Hekou). In average, each 1-kb window had ~23 SNPs (from 1 to 181 SNPs) and each scaffold had ~2,113 SNPs (from 0 to 172,088 SNPs) (Supporting Information Tables S1 and S2). Across all SNPs, average F_{ST} was low (mean $F_{ST} = 0.027$) and the majority of SNPs had an F_{ST} of <0.1 ($N = 4{,}796{,}897$; 93.37%). A total of 723 SNPs (0.014%) had $F_{ST} \geq 0.5$. Of those, 126 SNPs had $F_{ST} \geq 0.8$ and only 5 SNPs had $F_{ST} \geq 0.9$ (Figure 4a). In our control comparisons (where each highland population was paired with a lowland population), we found substantially fewer SNPs with high F_{ST}. For Deqin (H) + Hekou (L) versus Lijiang (H) + Xishuangbanna (L), there were only two SNPs with $F_{ST} \geq 0.5$ (mean $F_{ST} = 0.008$). For Deqin (H) + Xishuangbanna (L) (the two most divergent populations in the PCA analyses in Figure 3b) versus Lijiang (H) + Hekou (L), there were 12 SNPs with $F_{ST} \geq 0.5$ (mean $F_{ST} = 0.010$). There are therefore substantially more outlier SNPs in comparisons between highland and lowland populations compared to mixed control comparisons. Despite the observation that the Deqin population is relatively more diverged from the others, there is still a strong signal of SNP variants associated with highland populations, shared between Deqin (H) and Lijiang (H).

Genomic regions subject to selection are expected to contain not only SNPs with high F_{ST}, but also high XP-EHH and PBS scores. We therefore estimated these statistics for all SNPs in the highland

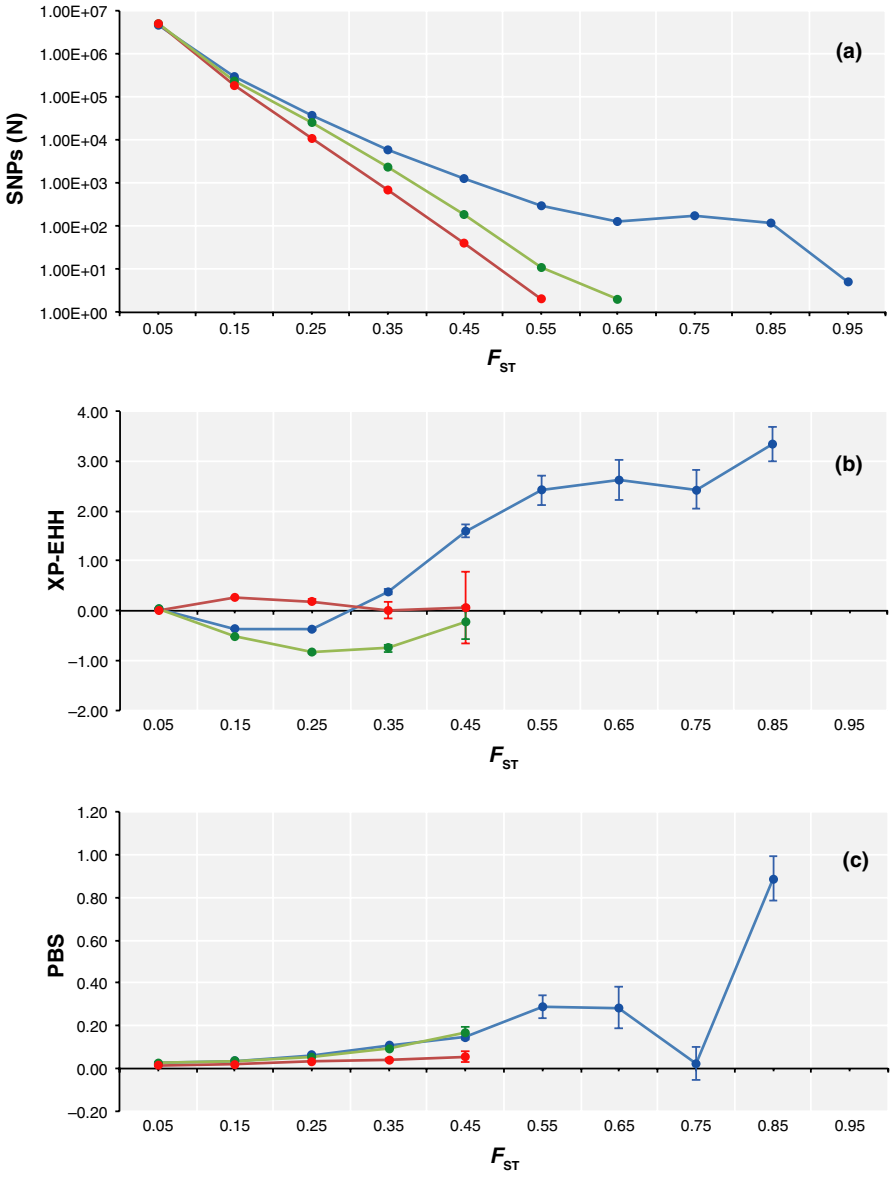

FIGURE 4 Selection signals are greater on SNPs showing increasing differentiation, as measured by Weir & Cockerham's F_{ST}, in the comparison between highland and lowland bees. (a) The number of SNPs (log-scale) for F_{ST} bins, colour-coded for each comparison as follows: Deqin (H) + Lijiang (H) versus Hekou (L) + Xishuangbanna (L) (highland vs. lowland, blue); Deqin (H) + Xishuangbanna (L) versus Lijiang (H) + Hekou (L) (green); Deqin (H) + Hekou (L) versus Lijiang (H) + Xishuangbanna (L) (red). (b) Mean XP-EHH (cross-population extended haplotype homozygosity) score and (c) mean PBS (population branch statistic) value for F_{ST} bins. Categories with <10 SNPs are not included in (b) and (c). Populations are defined as follows: DQ: Deqin; LJ: Lijiang; KM: Kunming; TS: Tianshui; XB: Xishuangbanna; HK: Hekou

versus lowland comparison. At SNPs with low F_{ST}, XP-EHH and PBS scores were close to zero and not biased towards any population. The mean XP-EHH was 0.0054 with upper and lower 99.9% percentiles of 4.39 and −3.15, respectively. Similarly, the mean PBS across all 1 kb windows was 0.035 with upper and lower 99.9% percentiles of 1.76 and −0.48, respectively. For both XP-EHH and PBS scores, positive values imply selection in highland populations while negative values indicate selection in the background population (lowland bees). The mean XP-EHH score was significantly elevated in high F_{ST} regions (Figure 4b), implying that high F_{ST} SNPs were associated with long haplotypes specifically in highland bees. The control comparisons contained very few SNPs at high F_{ST} and no indication of elevated XP-EHH values with higher F_{ST}. The PBS estimates also had positive values in high F_{ST} regions, although these were not significantly positive for the F_{ST} ranges of 0.7–0.8 (N = 170 SNPs) (Figure 4c). High F_{ST} values were not associated with high values in the control comparisons.

We also analysed the functional categories of SNPs according to F_{ST} value in the highland versus lowland comparison. SNPs within the lowest F_{ST} class (0–0.2) between highland and lowland bees were almost exclusively located in intergenic regions (48%) and introns (44%) (Figure 5a). With increasing F_{ST}, a higher proportion of SNPs were found in coding regions. At high F_{ST} categories (0.6–1), SNPs were more frequently found within UTRs and coding sequences (6.9-fold increase) compared to SNPs in low F_{ST} categories (0–0.6) (χ^2 = 485.31, p-value < 2.20^{-16}). In addition, for variable positions in coding sequences, SNPs with high F_{ST} (0.6–1) were more often associated with nonsynonymous mutations (3.1-fold increase) (i.e., that change the amino acid sequence of the resulting proteins) than SNPs with low F_{ST} (0–0.6) (χ^2 = 22.47, p-value < 2.13^{-6}) (Figure 5b). Hence, SNPs with highly divergent allele frequencies between highland and lowland populations were more likely to have functional effects and are likely targets of selection.

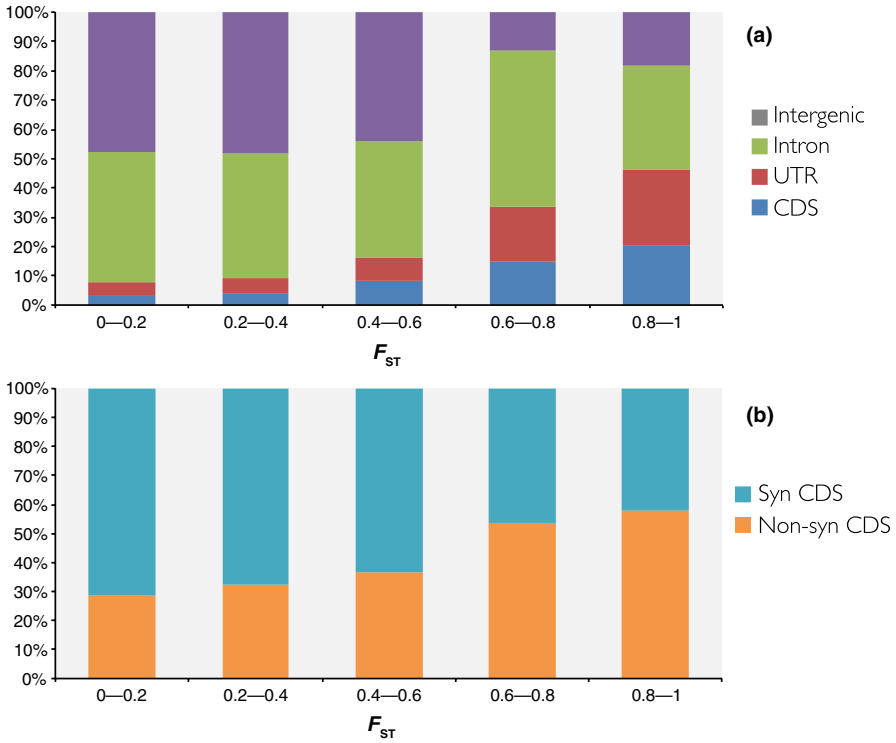

FIGURE 5 Genome annotation changes according to F_{ST}. (a) Proportion of SNPs located across the different parts of the genome (intergenic, intron, UTR and CDS-coding sequences) in relation to F_{ST}. (b) Proportion of SNPs that produce synonymous (Syn CDS) or non-synonymous mutations (Nonsyn CDS) in coding sequences in relation to F_{ST}

3.3 | Identification of outlier SNPs and candidate regions of selection

We calculated a CSS for all 5,137,335 SNPs, which had a mean CSS of 0.36 with upper and lower 99.9% percentiles of 2.34 and 0.006, respectively. In order to identify candidate genomic regions of selection, we applied a series of conservative cut-offs to restrict our data set to only the SNPs with the strongest signals. We used the top 1,000 SNPs sorted by CSS score (CSS from 5.31 to 3.10; mean 3.50; Supporting Information Table S3). These outlier SNPs were associated with 144 unique *Apis cerana* genes. In order to search for potential shared functions among these genes, we identified the corresponding *Drosophila* orthologs, resulting in 94 unique *D. melanogaster* gene accessions. Using WEBGESTALT, we found that 93 IDs were unambiguously mapped to unique Entrez Gene IDs. These were used to detect high-level GO terms across nonredundant biological process (BP), cellular component and molecular function categories (Supporting Information Figure S6). However, no statistically significant enrichments of functions were observed (false discovery rate; FDR >0.05; Supporting Information Table S4), indicating that potential gene candidates have diverse roles.

For the purpose of qualitatively characterizing the most prominent candidate genes, we chose to focus on the top 10 regions in the genome (labelled A–J; Figure 6; Table 2) with regard to the CSS scores. These regions were defined as segments within scaffolds with at least 10 outlier SNPs with CSS scores equal to or above 3.50 (N = 272). The 10 candidate regions are confined to particular locations within 10 different scaffolds and in total span 37 genes. The number of outlier SNPs was not evenly distributed among these scaffolds: 53.31% of the SNPs (N = 145) are found within just three scaffolds (41, 15 and 65) corresponding to candidate regions "F," "B" and "G" (Figure 6, Supporting Information Figure S7). However, the amount of outlier SNPs is not related to differences in sequence length as sweep region "A" (44,930 bp, in scaffold 14) contains 10 outlier SNPs, while sweep region "J" (4,337 bp, in scaffold 1,417) includes 25 outlier SNPs (Table 2).

Candidate region "F" contains the largest number of SNPs (N = 62), with the majority located within two genes, ACSNU04790 (N = 38) and ACSNU04792 (N = 17), which code for prisilkin-39-like and leucokinin receptor, respectively. In region "B," there are 52 SNPs distributed across five genes, but most SNPs are found again within two genes, ACSNU02553 (N = 27) and ACSNU02549 (N = 13) coding for esterase FE4-like and msta, isoform A-like, respectively. Finally, region "G" has 29 SNPs of which 16 are found in ACSNU06152 that codes for potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 4-like. Other candidate genes with large numbers of SNPs are: ACSNU04458 (N = 19) in region "D" that codes for UDP-glucuronosyltransferase 2C1-like and ACSNU10297 (N = 22) in region "J" that codes for glutamate NMDA receptor 1 and has the highest average CSS score (Table 2).

Using SAMBADA, we identified 402 SNPs (0.0068%) corresponding to 143 genes that were strongly correlated with altitude across all six populations (p-value ≤ 0.01). In order to compare these SNPs with those identified in the highland versus lowland comparisons, we looked at the F_{ST} values for these SNPs from the highland and lowland comparisons. We found 249 SNPs (61.94%) with F_{ST} ≥0.5 located within 40 genes and within these, 120 SNPs (29.85%) with F_{ST} ≥0.8 in 16 genes. Strikingly, 23 out of 37 (62.16%) of the candidate genes identified using the CSS approach were found to contain SNPs that had significant associations with altitude (Table 2).

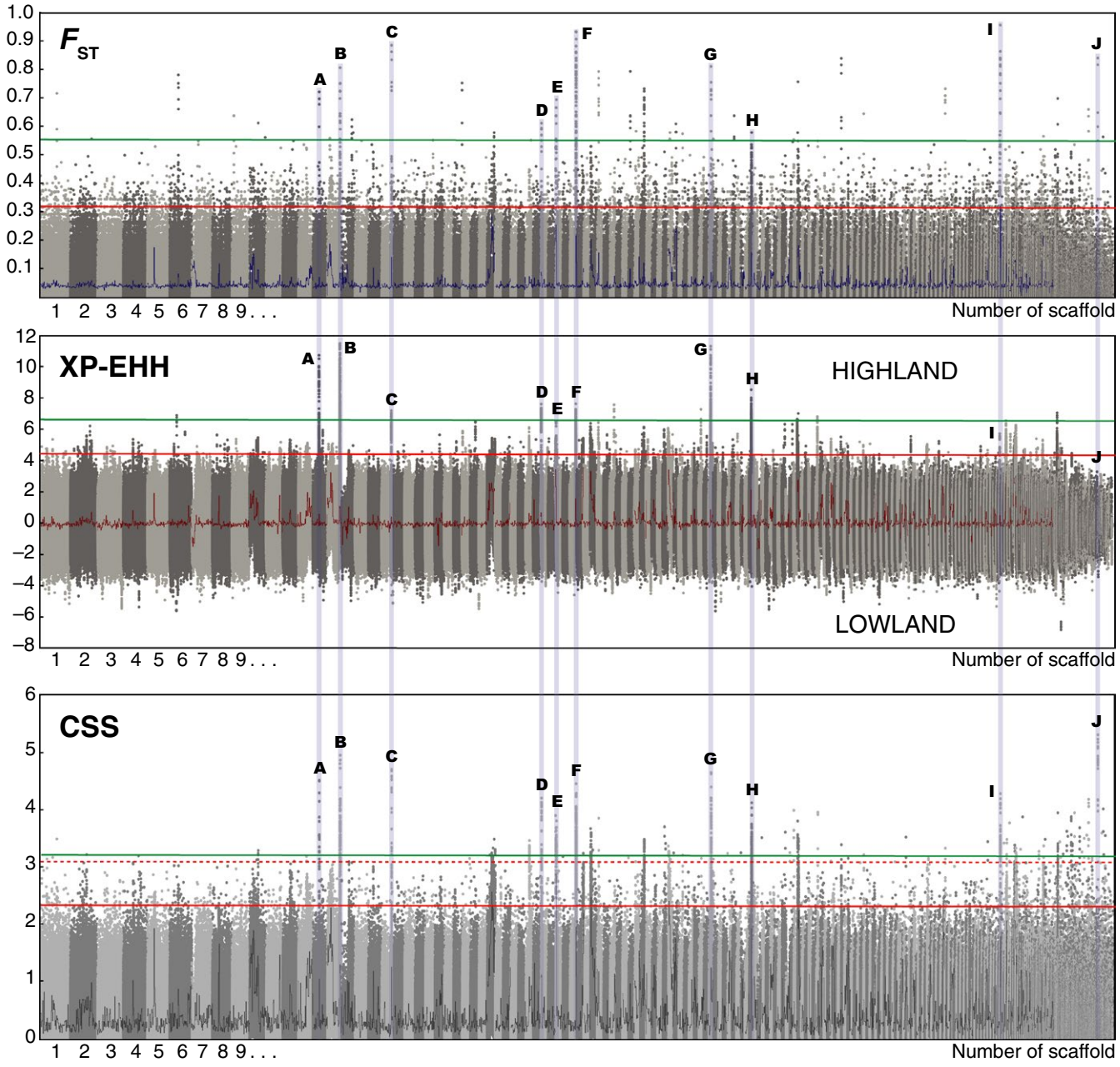

FIGURE 6 Candidate genes for selection signals based on Weir & Cockerham's F_{ST} and XP-EHH (cross-population extended haplotype homozygosity) score. Green lines are the 99.99% percentiles. Red lines are the 99.9% percentiles. Red dashed line shows the minimum CSS level of the top 1,000 SNPs

4 | DISCUSSION

In this study, we compared the genome sequences of 60 eastern honey bees (*Apis cerana*) from highland and lowland habitats across China. Our use of a combination of F_{ST} statistics, long-range haplotype tests (XP-EHH) and PBSs led to insights into the evolutionary history and genomic basis of adaptation in these populations. First, we found that population differentiation is low across our sampled populations, with an average pairwise F_{ST} of 0.027 although a degree of structure exists between populations. Second, we showed that very few SNPs are highly differentiated between highland and lowland bees in Yunnan, but that there are distinct regions of the genome containing a large proportion of the SNPs of heightened F_{ST}. There were substantially more F_{ST} outliers in the comparison of highland and lowland populations relative to control comparisons, where highland and lowland populations were grouped together. Third, we found that F_{ST} outlier SNPs were enriched within gene coding regions, and in particular for nonsynonymous variants. This indicates that F_{ST} outlier SNPs have functional consequences and are likely to have been selected upon. Finally, the observation that XP-EHH and PBS scores were elevated in the vicinity of SNPs with high F_{ST} demonstrates that these SNPs were associated with signatures of

TABLE 2 Summary of candidate genes of selection as inferred by the CSS approach

Accession	Locus	Product	Scaffold	Outlier SNPs	Avg. CSS	Sweep region
ACSNU02426	LOC107992823	hypothetical protein	14	8	4.149	A
ACSNU02429	LOC107992813	7,8-dihydro-8-oxoguanine triphosphatase-like	14	2	3.541	A
ACSNU02549*	LOC107993222	msta, isoform A-like	15	13	3.888	B
ACSNU02550*	LOC107993208	tetratricopeptide repeat protein 37-like	15	2	3.655	B
ACSNU02551	LOC107993213	integrin-linked protein kinase-like	15	3	3.554	B
ACSNU02552*	LOC107993212	carboxylesterase [esterase FE4-like]	15	9	3.728	B
ACSNU02553*	LOC107993168	esterase FE4-like	15	27	3.954	B
ACSNU02968*	LOC107994344	rolling stone-like isoform 1	19	4	3.754	C
ACSNU02969*	LOC107994344	rolling stone-like isoform 2	19	10	4.494	C
ACSNU02970*	LOC107994334	hypothetical protein [zinc finger protein 184-like]	19	1	4.386	C
ACSNU04458*	LOC107997718	UDP-glucuronosyltransferase 2C1-like	37	19	4.018	D
ACSNU04459*	LOC107997715	hypothetical protein	37	7	3.921	D
ACSNU04627	LOC107998006	choline/ethanolaminephosphotransferase 1-like	39	1	3.901	E
ACSNU04628	LOC107998008	CD9 antigen	39	3	3.807	E
ACSNU04635	LOC107997937	vinculin-like isoform 2	39	2	3.600	E
ACSNU04637	LOC107998040	transcription initiation factor TFIID subunit 8-like	39	1	3.637	E
ACSNU04649	LOC107998043	hypothetical protein [zinc finger protein 62-like isoform X1]	39	1	3.559	E
ACSNU04652	LOC107998004	hippocampus abundant transcript 1 protein-like isoform 1	39	1	3.533	E
ACSNU04653	LOC107998005	TBC1 domain family member 7-like	39	2	3.751	E
ACSNU04790*	LOC107998402	hypothetical protein [prisilkin-39-like]	41	38	3.867	F
ACSNU04792*	LOC107998395	leucokinin receptor [neuropeptide Y receptor type 1-like]	41	17	3.669	F
ACSNU04793*	LOC107998398	aveugle	41	4	3.672	F
ACSNU04794	LOC107998397	40S ribosomal protein S26	41	3	3.529	F
ACSNU06146	LOC108000944	hypothetical protein [abl interactor 2 isoform X1]	65	3	3.859	G
ACSNU06147	LOC108000928	cytochrome P450 304a1-like	65	5	3.974	G
ACSNU06148*	LOC108000929	signal peptidase complex catalytic subunit SEC11C	65	1	4.022	G
ACSNU06151*	LOC108000943	phosphatidylinositol 4-kinase alpha-like isoform 2	65	4	3.846	G
ACSNU06152*	LOC108000937	potassium/sodium hyperpolarization-activated cyclic nucleotide-gated channel 4-like	65	16	3.986	G
ACSNU06519*	LOC108001634	growth hormone secretagogue receptor type 1	74	1	3.529	H
ACSNU06522*	LOC108001633	hypothetical protein	74	19	3.714	H
ACSNU06523	LOC108001632	tyrocidine synthase 3 [mycosubtilin synthase subunit C]	74	5	3.588	H

(Continues)

TABLE 2 (Continued)

Accession	Locus	Product	Scaffold	Outlier SNPs	Avg. CSS	Sweep region
ACSNU08974*	LOC107993893	cyclin-related protein FAM58A-like isoform 1	171	2	3.726	I
ACSNU08975*	LOC107993894	alpha-methylacyl-CoA racemase-like	171	10	3.855	I
ACSNU08976*	LOC107993871	serine/threonine-protein phosphatase 2A activator	171	1	3.946	I
ACSNU08977*	LOC107993873	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A-like protein 1-like	171	2	3.874	I
ACSNU10297*	LOC107992915	hypothetical protein [glutamate NMDA receptor 1]	1,417	22	4.711	J
ACSNU10298*	LOC107992915	NMDA receptor 1	1,417	3	4.356	J

Candidate genes that were also found by SAMBADA are indicated with an asterisk.

selection in bees from high altitudes. Taken together, our results provide evidence that selection on multiple genes has led to local adaptation in high elevation populations.

By comparing highland and lowland populations, we found little differentiation across the genome in general. However, we found a number of SNPs with high F_{ST} values, showing that allele frequencies differ quite markedly between highland and lowland bees at these loci. Control comparisons involving pairings of highland and lowland populations generated substantially fewer F_{ST} outlier SNPs, which demonstrates that the signals we observed were most likely due to selection rather than random genetic drift. A potential issue with our analysis is the presence of population structure affecting the proportion of high F_{ST} SNPs. However, our results are unlikely to be affected by this because population structure does not correlate with the location of populations in highland and lowland habitats (see Figure 3), but rather occurs within these habitats. Therefore, our analysis demonstrates that there are alleles shared in both highland populations that are implicated in high-altitude adaptation, which are not shared with lowland populations.

Further support for the hypothesis that selection has driven adaptive divergence between highland and lowland bees comes from the strong association we found between the genomic location of SNPs and their F_{ST} values. Low F_{ST} SNPs were mainly located in noncoding regions (intergenic and intron sequences), while high F_{ST} SNPs were significantly enriched within UTRs and coding sequences. In coding sequences, we observed that high F_{ST} SNPs more often caused nonsynonymous mutations compared to low F_{ST} SNPs. This enrichment for functional parts of the genome strongly suggests that selection and local adaptation in high-altitude bees have led to divergence between lowland and highland populations of *A. cerana* from Yunnan. Genetic drift can be discounted as an explanation for most of the divergence we observe, as drift would not have resulted in the enrichment of divergent SNPs within coding regions and nonsynonymous positions.

During the course of local adaptation, natural selection is expected to increase the frequency of beneficial alleles within a population. Selection also affects the frequency of linked variants due to hitchhiking or background selection, which generates genomic divergence between populations experiencing different environments and leads to EHH around alleles under selection (Garner, Goulet, Farnitano, Molina-Henao, & Hopkins, 2018; Sabeti et al., 2007). In the comparison of highland and lowland bees, we find that highly differentiated genomic regions tend to have long haplotypes and elevated divergence in highland bees as shown by the XP-EHH and PBS results, which is a sign of recent selection in these populations. The observation that these statistics are elevated in the highland populations indicates that it is specifically these populations that have undergone recent selection rather than the lowland populations. We expect such recent selection to be detectable if it occurred more recently than the average time to coalescence of alleles within the population, which is approximated by 2*Ne* (Wakeley, 2009). Based on the mutation rate estimated for *Apis mellifera* (Wallberg et al., 2014) and the level of genetic variation (θ_w) estimated here for *A. cerana* (Table 1), we estimate this value as approximately 600,000 generations.

Among high-CSS SNPs, we identified 10 candidate genomic regions with evidence for selection in highland bees. These candidate regions span 37 genes and include 272 SNPs in total. There was also clear overlap in the positions of the outlier SNPs detected by SAMBADA and these candidate regions. These loci are widespread across multiple scaffolds and their gene products are involved in a diverse range of functions. A large number of outlier SNPs are found in two genes that code for esterase FE4-like ($N = 27$) and carboxylesterase ($N = 9$) proteins. In insects, esterase proteins are involved in many BPs, such as detoxification of environmental toxins (e.g., plant materials), development, courtship behaviour and reproduction (Bloch, Heifetz, & Wolfner, 2003; Mackert, do Nascimento, Bitondi, Hartfelder, & Simões, 2008; Mane, Tompkins, & Richmond, 1983). Interestingly, it has been shown that low temperature (4°C) affects esterase enzyme activity in *A. cerana* (Ma et al., 2018). In *D. melanogaster*, the reproductive fitness of males is strongly affected by temperature and the esterase-6 alleles

they carry. At low temperatures, active esterase-6 leads to earlier mating, shorter copulation and the production of more offspring (Gilbert & Richmond, 1982). It is therefore possible that selection on esterase genes in highland *A. cerana* has mediated its activity at lower temperatures in these populations.

In mountain populations of *A. mellifera* in East Africa, octopamine receptor genes, which are also involved in foraging behaviour through the octopamine pathway, have been identified as potential candidates for involvement in adaptation to highland habitats (Barron, Schulz, & Robinson, 2002; Chen et al., 2016; Giray, Galindo, & Oskay, 2007; Wallberg et al., 2017). Altered foraging behaviour may ensure that bees are able to cope with a different distribution and abundance of food resources compared to lowland areas. We did not identify selection signals on octopamine receptor genes in highland *A. cerana*. However, we identified signals of selection in other genes that may mediate foraging. For example, we identified signals of selection in leucokinin receptor genes (17 outlier SNPs), which are known to be involved in feeding and sucrose responses in flies and mosquitoes (Al-Anzi et al., 2010; Kwon et al., 2016). Leucokinins are neuropeptides that negatively regulate feeding behaviour, and their mutations cause an increase in meal size and a compensatory reduction in meal frequency in flies (Al-Anzi et al., 2010; Taghert & Veenstra, 2003). We also detected a high number of outlier SNPs in NMDA receptor genes ($N = 25$), which have been related to enhanced olfactory learning and memory activity, as well as foraging behaviour in honey bees (Behrends & Scheiner, 2012; Kim, Kim, Kim, Hong, & Lee, 2018; Si, Helliwell, & Maleszka, 2004). The ability to find and memorize the location of food resources in high-altitude environments is likely essential for highland populations of *A. cerana*.

Our analysis indicates that multiple loci likely drive high-altitude adaptation in *A. cerana*. A classic "hard" selective sweep refers to the rapid increase in frequency of a new mutation to fixation and the elimination of surrounding variation in flanking regions due to genetic hitchhiking (Maynard Smith & Haigh, 1974). However, adaptation may also result from soft sweeps on standing variation or polygenic adaptation due to more subtle allele frequency changes at multiple loci (Hermisson & Pennings, 2005; Pritchard, Pickrell, & Coop, 2010). In our data set, there are a handful of SNPs with extreme allele frequency differences that may belie hard sweeps but this does not appear to have been the dominant process involved in adaptation. This pattern contrasts with the mechanism of adaptation in high-altitude populations of *A. mellifera* in East Africa, where two ancient chromosome inversions in highland bees appear to house the adaptive variants required for high-altitude living (Christmas et al., 2018; Wallberg et al., 2017). We did not find any evidence for the presence of structural rearrangements between high- and low-altitude populations of *A. cerana*, suggesting that inversions do not always contribute to highland adaptation in honey bees.

High altitudes present a series of unique environmental conditions, such as lower temperatures, reduced levels of oxygen, higher ultraviolet radiation, as well as differences in vegetation and food availability compared to lowland areas. For insects, altitude challenges thermoregulation, flying capabilities and reproduction (Hodkinson, 2005). Highland populations of *A. cerana* are larger, darker and have longer body hair compared to lowland bees (Pereboom & Biesmeijer, 2003; Tan et al., 2003; Tan & Lingjuan, 2008). Consistent with this, we found a significant tendency for the highland specimens we sampled to be darker than samples from lower altitudes. Similarly, *A. mellifera* honey bees in the mountain forests of East Africa tend to be darker and larger compared to the surrounding lowland savannah populations (Ruttner, 1988). Body size in insects may be related to cold resistance (Scharf, Filin, Ben-Yehoshua, & Ovadia, 2009; Scharf, Sbilordo, & Martin, 2014). Larger body size in mountain *A. cerana* honey bees may result in an increased capability of nectar collection, as nectar glands of flowers serving as food sources tend to become deeper as the elevation increases in this region of China (Zhu et al., 2017). In addition to physiological constraints, the spatial and temporal distribution and abundance of food sources in highland habitats may favour different optimal foraging strategies compared to lowland habitats.

Despite the observations of differences in size and colour of highland *A. cerana*, we did not specifically uncover genes with well-known functions in morphology or pigmentation that might underlie these differences, although we cannot rule out that any of these genes we identified have such effects. No such genes were observed in chromosomal inversions connected to highland adaptation in mountain populations of *A. mellifera* either (Wallberg et al., 2017). Furthermore, in *A. mellifera*, no correlation was observed between pigmentation and presence of the highland inversion haplotype, which is close to fixation in highland populations. The adaptive advantages encoded within variants under selection in both highland *A. cerana* and *A. mellifera* may not be related to pigmentation and instead relate to other factors important for adaptation to high-altitude habitats in this species, such as foraging strategy. The degree to which coloration is genetically controlled in honey bees is unclear, and studies in other insects indicate phenotypic plasticity mediated by temperature (Marriott & Holloway, 1998).

Although local adaptation to high-altitude environments in *A. cerana* has a different genomic basis to that in *A. mellifera*, involving selection at multiple loci, it is possible that selection for similar functions is involved in highland adaptation for both species. In highland *A. cerana*, we have also identified signals of selection in genes with potential function in feeding behaviour, indicating that behavioural differences may be more important for highland adaptation than morphological differences. However, further work including functional experiments would be needed to fully address the relevance of these genes in honey bees and the effects of the variants under selection.

5 | CONCLUSIONS

We have identified several extremely differentiated genomic regions between highland and lowland populations of the eastern

honey bee. These regions are biased towards coding sequences and contain a higher proportion of nonsynonymous mutations compared to the rest of the genome. They also tend to have high haplotype homozygosity in the highland bees, indicating selective sweeps in these populations. We have found that selection on multiple loci could be the source of local adaptation in highland *Apis cerana* bees. Genes associated with these selected regions have diverse functionality and are likely related to adaptive traits present in highland bee populations.

ACKNOWLEDGEMENTS

We would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, NGI, and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure. We also thank Julia Jones for her feedback on the manuscript. The Spanish Government (Ministerio de Economía y Competitividad) supported this work with a FPI (Formación de Personal Investigador) fellowship (BES-2014-069006) as well as three travel grants (EEBB-I-16-10576, EEBB-I-17-12168 and EEBB-I-18-12878) to SMM. This research was supported by grants 2013-722 from the Swedish Research Council Formas and 2014-5096 Swedish Research Council Vetenskapsrådet to MTW. The CAS Key Laboratory of Tropical Forest Ecology, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Science, also supported this work with the CAS 135 programme (No. 2017XTBG-T01), China National Research Fund (31770420), as well as the Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Grant No. Y4ZK111B01 to KT. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

A.W. and M.T.W. conceived and designed the experiments. S.M.-M., K.T., M.J.C., A.O., A.W. and M.T.W. performed the experiments. S.M.-M., M.J.C., A.W. and M.T.W. analysed the data. K.T. and C.V. contributed reagents, materials or analysis tools. S.M.-M., M.J.C., C.V., A.W. and M.T.W. wrote the paper. S.M.-M., K.T., M.J.C., A.O., C.V., A.W. and M.T.W. revised and approved the final version of the manuscript.

DATA ACCESSIBILITY

All data from this study have been deposited at the NCBI Sequence Read Archive (SRA) under BioProject PRJNA480840.

ORCID

Carles Vilà https://orcid.org/0000-0002-4206-5246

Matthew T. Webster https://orcid.org/0000-0003-1141-2863

REFERENCES

Al-Anzi, B., Armand, E., Nagamei, P., Olszewski, M., Sapin, V., Waters, C., … Benzer, S. (2010). The leucokinin pathway and its neurons regulate meal size in *Drosophila*. *Current Biology*, *20*, 969–978. https://doi.org/10.1016/j.cub.2010.04.039

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*, 1655–1664. https://doi.org/10.1101/gr.094052.109

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*, 697–709. https://doi.org/10.1038/nrg2844

Barron, A. B., Schulz, D. J., & Robinson, G. E. (2002). Octopamine modulates responsiveness to foraging-related stimuli in honey bees (*Apis mellifera*). *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, *188*, 603–610.

Behrends, A., & Scheiner, R. (2012). Octopamine improves learning in newly emerged bees but not in old foragers. *Journal of Experimental Biology*, *215*, 1076–1083. https://doi.org/10.1242/jeb.063297

Bigham, A., Bauchet, M., Pinto, D., Mao, X., Akey, J. M., Mei, R., … Shriver, M. D. (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. *PLoS Genetics*, *6*, e1001116. https://doi.org/10.1371/journal.pgen.1001116

Bloch, Q. M. C., Heifetz, Y., & Wolfner, M. F. (2003). The developments between gametogenesis and fertilization: Ovulation and female sperm storage in *Drosophila melanogaster*. *Developmental Biology*, *256*, 195–211. https://doi.org/10.1016/S0012-1606(02)00125-2

Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *The American Journal of Human Genetics*, *81*, 1084–1097. https://doi.org/10.1086/521987

Chen, C., Liu, Z., Luo, Y., Xu, Z., Wang, S., Zhang, X., … Shi, W. (2017). Managed honeybee colony losses of the Eastern honeybee (*Apis cerana*) in China (2011–2014). *Apidologie*, *48*, 692–702. https://doi.org/10.1007/s13592-017-0514-6

Chen, C., Liu, Z., Pan, Q., Chen, X., Wang, H., Guo, H., … Shi, W. (2016). Genomic analyses reveal demographic history and temperate adaptation of the newly discovered honey bee subspecies *Apis mellifera sinisxinyuan* n. ssp. *Molecular Biology and Evolution*, *33*, 1337–1348.

Chen, C., Wang, H., Liu, Z., Chen, X., Tang, J., Meng, F., & Shi, W. (2018). Population genomics provide insights into the evolution and adaptation of the Eastern honey bee (*Apis cerana*). *Molecular Biology and Evolution*, *35*, 2260–2271. https://doi.org/10.1093/molbev/msy130

Cheviron, Z. A., & Brumfield, R. T. (2012). Genomic insights into adaptation to high-altitude environments. *Heredity*, *108*, 354–361. https://doi.org/10.1038/hdy.2011.85

Christmas, M. J., Wallberg, A., Bunikis, I., Olsson, A., Wallerman, O., & Webster, M. T. (2018). Chromosomal inversions associated with environmental adaptation in honeybees. *Molecular Ecology*. https://doi.org/10.1111/mec.14944

Cridland, J. M., Tsutsui, N. D., & Ramírez, S. R. (2017). The complex demographic history and evolutionary origin of the western honey bee, *Apis mellifera*. *Genome Biology and Evolution*, *9*, 457–472. https://doi.org/10.1093/gbe/evx009

Dillon, M. E. (2006). Into thin air: Physiology and evolution of alpine insects. *Integrative and Comparative Biology*, *46*, 49–61. https://doi.org/10.1093/icb/icj007

Eden, E., Navon, R., Steinfeld, I., Lipson, D., & Yakhini, Z. (2009). GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics*, *10*, 48. https://doi.org/10.1186/1471-2105-10-48

Elsik, C. G., Worley, K. C., Bennett, A. K., Beye, M., Camara, F., Childers, C. P., … Gibbs, R. A. (2014). Finding the missing honey bee genes: Lessons learned from a genome upgrade. *BMC Genomics*, *15*, 86.

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and efficient estimation of individual ancestry coefficients. *Genetics*, *196*, 973–983. https://doi.org/10.1534/genetics.113.160572

Garner, A. G., Goulet, B. E., Farnitano, M. C., Molina-Henao, Y. F., & Hopkins, R. (2018). Genomic signatures of reinforcement. *Genes*, *9*, 191. https://doi.org/10.3390/genes9040191

Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. Preprint at https://arxiv.org/abs/1207.3907.

Gilbert, D. G., & Richmond, R. C. (1982). Esterase 6 in *Drosophila melanogaster*: Reproductive function of active and null males at low temperature. *Proceedings of the National Academy of Sciences of the United States of America*, *79*, 2962–2966. https://doi.org/10.1073/pnas.79.9.2962

Giray, T., Galindo, A., & Oskay, D. (2007). Octopamine influences honey bee foraging preference. *Journal of Insect Physiology*, *53*, 691–698. https://doi.org/10.1016/j.jinsphys.2007.03.016

Grabherr, M. G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., Di Palma, F., & Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics*, *26*, 1145–1151. https://doi.org/10.1093/bioinformatics/btq102

Gruber, K., Schöning, C., Otte, M., Kinuthia, W., & Hasselmann, M. (2013). Distinct subspecies or phenotypic plasticity? Genetic and morphological differentiation of mountain honey bees in East Africa. *Ecology and Evolution*, *3*, 3204–3218. https://doi.org/10.1002/ece3.711

Hermisson, J., & Pennings, P. S. (2005). Soft sweeps: Molecular population genetics of adaptation from standing genetic variation. *Genetics*, *169*, 2335–2352. https://doi.org/10.1534/genetics.104.036947

Hodkinson, I. D. (2005). Terrestrial insects along elevation gradients: Species and community responses to altitude. *Biological Reviews of the Cambridge Philosophical Society*, *80*, 489–513. https://doi.org/10.1017/S1464793105006767

Huerta-Sánchez, E., Degiorgio, M., Pagani, L., Tarekegn, A., Ekong, R., Antao, T., … Nielsen, R. (2013). Genetic signatures reveal high-altitude adaptation in a set of ethiopian populations. *Molecular Biology and Evolution*, *30*, 1877–1888. https://doi.org/10.1093/molbev/mst089

Ilardo, M., & Nielsen, R. (2018). Human adaptation to extreme environmental conditions. *Current Opinion in Genetics & Development*, *53*, 77–82. https://doi.org/10.1016/j.gde.2018.07.003

Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, *23*, 1801–1806. https://doi.org/10.1093/bioinformatics/btm233

Jombart, T. (2008). Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*, 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kawecki, T. J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecology Letters*, *7*, 1225–1241. https://doi.org/10.1111/j.1461-0248.2004.00684.x

Kim, K., Kim, J. H., Kim, Y. H., Hong, S. E., & Lee, S. H. (2018). Pathway profiles based on gene-set enrichment analysis in the honey bee *Apis mellifera* under brood rearing-suppressed conditions. *Genomics*, *110*, 43–49. https://doi.org/10.1016/j.ygeno.2017.08.004

Kirkpatrick, M., & Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics*, *173*, 419–434. https://doi.org/10.1534/genetics.105.047985

Koetz, A. (2013). Ecology, behaviour and control of *Apis cerana* with a focus on relevance to the Australian incursion. *Insects*, *4*, 558–592. https://doi.org/10.3390/insects4040558

Kwon, H., Ali Agha, M., Smith, R. C., Nachman, R. J., Marion-Poll, F., & Pietrantonio, P. V. (2016). Leucokinin mimetic elicits aversive behavior in mosquito *Aedes aegypti* (L.) and inhibits the sugar taste neuron. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 6880–6885.

Leimu, R., & Fischer, M. (2008). A meta-analysis of local adaptation in plants. *PLoS ONE*, *3*(12), e4010.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *2516*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Ma, M., Jia, H., Cui, X., Zhai, N., Wang, H., Guo, X., & Xu, B. (2018). Isolation of carboxylesterase (esterase FE4) from *Apis cerana cerana* and its role in oxidative resistance during adverse environmental stress. *Biochimie*, *144*, 85–97. https://doi.org/10.1016/j.biochi.2017.10.022

Mackert, A., do Nascimento, A. M., Bitondi, M. M. G., Hartfelder, K., & Simões, Z. L. P. (2008). Identification of a juvenile hormone esterase-like gene in the honey bee, *Apis mellifera* L.—Expression analysis and functional assays. *Comparative Biochemistry and Physiology. Part B: Biochemistry and Molecular Biology*, *150*, 33–44. https://doi.org/10.1016/j.cbpb.2008.01.004

Mane, S. D., Tompkins, L., & Richmond, R. C. (1983). Male esterase 6 catalyzes the synthesis of a sex pheromone in *Drosophila melanogaster* females. *Science*, *222*, 419–421. https://doi.org/10.1126/science.222.4622.419

Marriott, C. G., & Holloway, G. J. (1998). Colour pattern plasticity in the hoverfly, *Episyrphus balteatus*: The critical immature stage and reaction norm on developmental temperature. *Journal of Insect Physiology*, *44*, 113–119. https://doi.org/10.1016/S0022-1910(97)00097-8

Maynard Smith, J., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, *23*, 23–35. https://doi.org/10.1017/S0016672300014634

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., … DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303. https://doi.org/10.1101/gr.107524.110

Nelson, R. M., Wallberg, A., Simões, Z. L. P., Lawson, D. J., & Webster, M. T. (2017). Genomewide analysis of admixture and adaptation in the Africanized honeybee. *Molecular Ecology*, *26*, 1–15. https://doi.org/10.1111/mec.14122

Park, D., Jung, J. W., Choi, B.-S., Jayakodi, M., Lee, J., Lim, J., … Kwon, H. W. (2015). Uncovering the novel characteristics of Asian honey bee, *Apis cerana*, by whole genome sequencing. *BMC Genomics*, *16*, 1. https://doi.org/10.1186/1471-2164-16-1

Parker, R., Melathopoulos, A. P., White, R., Pernal, S. F., Guarna, M. M., & Foster, L. J. (2010). Ecological adaptation of diverse honey bee (*Apis mellifera*) populations. *PLoS ONE*, *5*(6), e11096.

Peng, Y., Yang, Z., Zhang, H., Cui, C., Qi, X., Luo, X., … Su, B. (2011). Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. *Molecular Biology and Evolution*, *28*, 1075–1081. https://doi.org/10.1093/molbev/msq290

Pereboom, J. J. M., & Biesmeijer, J. C. (2003). Thermal constraints for stingless bee foragers: The importance of body size and coloration. *Oecologia*, *137*, 42–50. https://doi.org/10.1007/s00442-003-1324-2

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, *20*, R208–R215. https://doi.org/10.1016/j.cub.2009.11.055

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: A toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, *81*, 559–575.

Radloff, S. E., Hepburn, C., Randall Hepburn, H., Fuchs, S., Hadisoesilo, S., Tan, K., … Kuznetsov, V. (2010). Population structure and classification of *Apis cerana*. *Apidologie*, *41*, 589–601.

Radwan, J., & Babik, W. (2012). The genomics of adaptation. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 5024–5028. https://doi.org/10.1098/rspb.2012.2322

Randhawa, I. A., Khatkar, M., Thomson, P., & Raadsma, H. (2014). Composite selection signals can localize the trait specific genomic regions in multi-breed populations of cattle and sheep. *BMC Genetics*, *15*, 34. https://doi.org/10.1186/1471-2156-15-34

Reynolds, J., Weir, B. S., & Cockerham, C. C. (1983). Estimation of the coancestry coefficient: Basis for a short-term genetic distance. *Genetics*, *105*, 767–779.

Ruttner, F. (1988). *Biogeography and taxonomy of honeybees*. Spinger-Verlag Berlin Heidelberg.

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., … Stewart, J. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *49*, 913–918. https://doi.org/10.1038/nature06250

Scharf, I., Filin, I., Ben-Yehoshua, D., & Ovadia, O. (2009). Phenotypic plasticity and variation in morphological and life-history traits of antlion adults across a climatic gradient. *Zoology*, *112*, 139–150. https://doi.org/10.1016/j.zool.2008.05.004

Scharf, I., Sbilordo, S. H., & Martin, O. Y. (2014). Cold tolerance in flour beetle species differing in body size and selection temperature. *Physiological Entomology*, *39*, 80–87. https://doi.org/10.1111/phen.12050

Shi, Y. Y., Sun, L. X., Huang, Z. Y., Wu, X. B., Zhu, Y. Q., Zheng, H. J., & Zeng, Z. J. (2013). A SNP based high-density linkage map of *Apis cerana* reveals a high recombination rate similar to *Apis mellifera*. *PLoS ONE*, *8*, e76459. https://doi.org/10.1371/journal.pone.0076459

Si, A., Helliwell, P., & Maleszka, R. (2004). Effects of NMDA receptor antagonists on olfactory learning and memory in the honeybee (*Apis mellifera*). *Pharmacology Biochemistry and Behavior*, *77*, 191–197. https://doi.org/10.1016/j.pbb.2003.09.023

Smith, D. R., Villafuerte, L., Otis, G., & Palmer, M. R. (2000). Biogeography of *Apis cerana* F. and *A. nigrocincta* Smith: Insights from mtDNA studies. *Apidologie*, *31*, 265–279.

Smith, F. G. (1961). The races of honeybees in Africa. *Bee World*, *42*, 255–260. https://doi.org/10.1080/0005772X.1961.11096896

Stucki, S., Orozco-terWengel, P., Forester, B. R., Duruz, S., & Colli, L. (2016). High performance computation of landscape genomic models including local indicators of spatial association. *Molecular Ecology Resources*, *17*, 1072–1089. https://doi.org/10.1111/1755-0998.12629

Szpiech, Z. A., & Hernandez, R. D. (2014). Selscan: An efficient multithreaded program to perform EHH-based scans for positive selection. *Molecular Biology and Evolution*, *31*, 2824–2827. https://doi.org/10.1093/molbev/msu211

Taghert, P. H., & Veenstra, J. A. (2003). *Drosophila* neuropeptide signaling. *Advances in Genetics*, *49*, 1–65.

Tan, K., Fuchs, S., Koeniger, N., & Ruiguang, Z. (2003). Morphological characterization of *Apis cerana* in the Yunnan Province of China. *Apidologie*, *34*, 553–561.

Tan, K., & Ling-juan, L. (2008). Morphology and taxonomy of *Apis cerana* in Deqin. *Journal of Yunnan Agricultural University*, *23*, 230–232.

Theisen-Jones, H., & Bienefeld, K. (2016). The Asian honey bee (*Apis cerana*) is significantly in decline. *Bee World*, *93*, 90–97.

Wakeley, J. (2009). *Coalescent theory: An introduction* (1st ed.). Greenwood Village, CO: Roberts and Company.

Wallberg, A., Han, F., Wellhagen, G., Dahle, B., Kawata, M., Haddad, N., … Webster, M. T. (2014). A worldwide survey of genome sequence variation provides insight into the evolutionary history of the honeybee *Apis mellifera*. *Nature Genetics*, *46*, 1081–1088. https://doi.org/10.1038/ng.3077

Wallberg, A., Pirk, C. W., Allsopp, M. H., & Webster, M. T. (2016). Identification of multiple loci associated with social parasitism in honeybees. *PLoS Genetics*, *12*, 1–30. https://doi.org/10.1371/journal.pgen.1006097

Wallberg, A., Schöning, C., Webster, M. T., & Hasselmann, M. (2017). Two extended haplotype blocks are associated with adaptation to high altitude habitats in East African honey bees. *PLoS Genetics*, *13*, 1–30. https://doi.org/10.1371/journal.pgen.1006792

Wang, J., Vasaikar, S., Shi, Z., Greer, M., & Zhang, B. (2017). WebGestalt 2017: A more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Research*, *45*, W130–W137. https://doi.org/10.1093/nar/gkx356

Wang, Z. L., Liu, T. T., Huang, Z. Y., Wu, X. B., Yan, W. Y., & Zeng, Z. J. (2012). Transcriptome analysis of the Asian honey bee *Apis cerana cerana*. *PLoS ONE*, *7*(10), e47954.

Watterson, G. A. (1975). On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, *7*, 256–276. https://doi.org/10.1016/0040-5809(75)90020-9

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*, 1358.

Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution*, *33*, 427–440. https://doi.org/10.1016/j.tree.2018.04.002

Whitlock, M. C. (2015). Modern approaches to local adaptation. *The American Naturalist*, *186*, S1–S4. https://doi.org/10.1086/682933

Xu, P., Shi, M., & Chen, X. (2009). Antimicrobial peptide evolution in the Asiatic honey bee *Apis cerana*. *PLoS ONE*, *4*, e4239. https://doi.org/10.1371/journal.pone.0004239

Xu, S., Li, S., Yang, Y., Tan, J., Lou, H., Jin, W., … Jin, L. (2011). A genome-wide search for signals of high-altitude adaptation in Tibetans. *Molecular Biology and Evolution*, *28*, 1003–1011. https://doi.org/10.1093/molbev/msq277

Yi, X., Liang, Y., Huerta-Sanchez, E., Jin, X., Cuo, Z. X., Pool, J. E., & Wang, J. (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, *329*, 75–78. https://doi.org/10.1126/science.1190371

Zhu, X., Zhou, S., Xu, X., Wang, J., Yinglong, Y., Yang, K.-c., … Zhou, B. (2017). Morphological differentiation in Asian honey bee (*Apis cerana*) populations in the basin and highlands of southwestern China. *Journal of Apicultural Research*, *56*, 203–209.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Montero-Mendieta S, Tan K, Christmas MJ, et al. The genomic basis of adaptation to high-altitude habitats in the eastern honey bee (*Apis cerana*). *Mol Ecol*. 2019;28:746–760. https://doi.org/10.1111/mec.14986