

MIT Open Access Articles

Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Sun, Cheng et al. "Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits." *Molecular Biology and Evolution* (September 2020): msaa240 © 2020 The Author(s)

As Published: <http://dx.doi.org/10.1093/molbev/msaa240>

Publisher: Oxford University Press (OUP)

Persistent URL: <https://hdl.handle.net/1721.1/129528>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution 4.0 International license



Title: Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits

Authors:

Cheng Sun^{1†*}, Jiaying Huang^{1†}, Yun Wang^{2†}, Xiaomeng Zhao^{1†}, Long Su^{1†}, Gregg W.C. Thomas^{3†}, Mengya Zhao^{4†}, Xingtang Zhang⁵, Irwin Jungreis^{6,7}, Manolis Kellis^{6,7}, Saverio Vicario⁸, Igor V. Sharakhov^{9,10}, Semen M. Bondarenko⁹, Martin Hasselmann¹¹, Chang N Kim¹², Benedict Paten¹², Luca Penso-Dolfin¹³, Li Wang¹⁴, Yuxiao Chang¹⁴, Qiang Gao¹⁵, Ling Ma¹⁵, Lina Ma¹⁶, Zhang Zhang¹⁶, Hongbo Zhang², Huahao Zhang¹⁷, Livio Ruzzante¹⁸, Hugh M. Robertson¹⁹, Yihui Zhu²⁰, Yanjie Liu¹, Huipeng Yang¹, Lele Ding¹, Quanguai Wang¹, Dongna Ma⁵, Weilin Xu¹, Cheng Liang²¹, Michael W. Itgen²², Lauren Mee²³, Gang Cao⁴, Ze Zhang², Ben M. Sadd²⁴, Matthew Hahn²⁵, Sarah Schaack²⁶, Seth M. Barribeau²³, Paul H. Williams²⁷, Robert M. Waterhouse^{18*} and Rachel Lockridge Mueller²²

Affiliations:

¹Institute of Apicultural Research, Chinese Academy of Agricultural Sciences, Beijing, China

²School of Life Sciences, Chongqing University, Chongqing, China

³Division of Biological Sciences, University of Montana, Missoula, Montana, USA

⁴State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan, China

⁵Fujian Provincial Key Laboratory of Haixia Applied Plant Systems Biology, Fujian Agriculture and Forestry University, Fuzhou, China

⁶MIT Computer Science and Artificial Intelligence Laboratory, Cambridge, Massachusetts 02139, USA

⁷Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA

⁸Institute of Atmospheric Pollution Research-Italian National Research Council C/O Department of Physics, University of Bari, via Orabona 4, 70125 Bari, Italy

⁹Department of Entomology, Virginia Polytechnic and State University, Blacksburg, VA, 24061, USA

¹⁰Department of Cytology and Genetics, Tomsk State University, Tomsk 634050, Russian Federation

- ¹¹Department of Livestock Population Genomics, Institute of Animal Science, University of Hohenheim, Stuttgart, Germany
- ¹²UC Santa Cruz Genomics Institute, University of California, Santa Cruz, USA
- ¹³Deutsches Krebsforschungszentrum, Heidelberg, Germany
- ¹⁴Shenzhen Branch, Guangdong Laboratory for Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China
- ¹⁵BGI Genomics, BGI-Shenzhen, Shenzhen, China
- ¹⁶China National Center for Bioinformation & Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China
- ¹⁷College of Pharmacy and Life Science, Jiujiang University, Jiujiang, China
- ¹⁸Department of Ecology and Evolution, University of Lausanne, and Swiss Institute of Bioinformatics, 1015 Lausanne, Switzerland
- ¹⁹Department of Entomology, University of Illinois at Urbana-Champaign, Champaign, IL, USA
- ²⁰Department of Medical Microbiology and Immunology, Genome Center, and MIND Institute, University of California, Davis, CA, USA
- ²¹Institute of Sericultural and Apiculture, Yunnan Academy of Agricultural Sciences, Mengzi, China
- ²²Department of Biology, Colorado State University, Fort Collins, CO, USA
- ²³Department of Ecology, Evolution and Behaviour, Institute of Integrative Biology, University of Liverpool, Liverpool, United Kingdom
- ²⁴School of Biological Sciences, Illinois State University, Normal, Illinois, USA
- ²⁵Department of Biology & Department of Computer Science, Indiana University, Bloomington, IN, USA
- ²⁶Department of Biology, Reed College, Portland, Oregon, USA
- ²⁷Department of Life Sciences, Natural History Museum, London, United Kingdom

†**Contributed equally.**

***Corresponding authors:** Cheng Sun (suncheng@caas.cn) & Robert M. Waterhouse (robert.waterhouse@unil.ch).

Keywords: *Bombus*, insect diversity, genome assembly, genome evolution, gene family evolution

Abstract

Bumblebees are a diverse group of globally important pollinators in natural ecosystems and for agricultural food production. With both eusocial and solitary life-cycle phases, and some social parasite species, they are especially interesting models to understand social evolution, behavior, and ecology. Reports of many species in decline point to pathogen transmission, habitat loss, pesticide usage, and global climate change, as interconnected causes. These threats to bumblebee diversity make our reliance on a handful of well-studied species for agricultural pollination particularly precarious. To broadly sample bumblebee genomic and phenotypic diversity, we *de novo* sequenced and assembled the genomes of 17 species, representing all 15 subgenera, producing the first genus-wide quantification of genetic and genomic variation potentially underlying key ecological and behavioral traits. The species phylogeny resolves subgenera relationships while incomplete lineage sorting likely drives high levels of gene tree discordance. Five chromosome-level assemblies show a stable 18-chromosome karyotype, with major rearrangements creating 25 chromosomes in social parasites. Differential transposable element activity drives changes in genome sizes, with putative domestications of repetitive sequences influencing gene coding and regulatory potential. Dynamically evolving gene families and signatures of positive selection point to genus-wide variation in processes linked to foraging, diet and metabolism, immunity and detoxification, as well as adaptations for life at high altitudes. Our study reveals how bumblebee genes and genomes have evolved across the *Bombus* phylogeny and identifies variations potentially linked to key ecological and behavioral traits of these important pollinators.

Introduction

Bumblebees (Hymenoptera: Apidae) are a group of pollinating insects comprising the genus *Bombus*, which are economically important for crop pollination (Velthuis and Van Doorn 2006; Garibaldi et al. 2013; Martin et al. 2019). Bumblebees are also ecologically important pollinators, serving as the sole or predominant pollinators of many wild plants (Fontaine et al. 2005; Goulson et al. 2008). They are particularly charismatic social insects that exhibit complex behaviors such as learning through observation (Alem et al. 2016) and damaging leaves to stimulate earlier flowering (Pashalidou et al. 2020). Global and local environmental changes have resulted in some species declining in range and abundance and others remaining stable or even increasing (Cameron et al. 2011; Bartomeus et al. 2013; Koch et al. 2015; Cameron and Sadd 2020). Decline in bumblebee abundance and distribution resulting from habitat loss, pathogen transmission, climate change, and agrochemical exposure is threatening pollination services to both wild plants and crops, raising concerns for bumblebees, the plant species they service, food security, and ecosystem stability (Grixti et al. 2009; Williams and Osborne 2009; Potts et al. 2010; Goulson et al. 2015; Cameron and Sadd 2020; Soroye et al. 2020) .

Bumblebees comprise ~250 extant species classified into 15 subgenera (Williams 1998; Cameron et al. 2007; Williams et al. 2018). The initial diversification of *Bombus* lineages occurred ~25–40 million years ago (Ma), near the Eocene-Oligocene boundary ~34 Ma (Williams 1998; Hines 2008). Bumblebees display considerable interspecific diversity in morphology, color patterning, food preference, pathogen incidence, and exhibit diverse life histories and ecologies (Williams 1994; Sikora and Kelm 2012; Persson et al. 2015; Arbetman et al. 2017; Cameron and Sadd 2020).

Members of the subgenus *Mendacibombus*, the sister group to all other extant bumblebees, are high-elevation specialists with distributions centered on the Qinghai-Tibetan plateau (Williams et al. 2018). Species in the subgenus *Psithyrus* exhibit social parasitism; they do not have a worker caste, and they feed on food collected by workers of their host species (Lhomme and Hines 2019). Bumblebees are distributed across the globe, from Greenland to the Amazon Basin and from sea level to altitudes of 5,640 m in the Himalayas, where they occupy diverse habitats, from alpine meadows to lowland tropical forest (Williams 1985; Williams et al. 2018). Much remains to be learned about bumblebees. For example, little is known about the

underlying genetic and genomic variation that gives rise to these diverse phenotypes, including their differential responses to changing environments.

About half of the ~250 extant species, representing 14 of the 15 *Bombus* subgenera, are found in China, making it a hotspot of bumblebee species richness (Williams 1998; Williams et al. 2017). To broadly sample the genomic and phenotypic diversity of bumblebees, we selected representative species from China for whole genome sequencing based on their phylogeny, ecology, behavior, geography, and specimen availability. To complete subgenus sampling, we additionally selected *B. polaris* from the subgenus *Alpinobombus*, which is endemic to arctic/subarctic regions. In total, we performed *de novo* sequencing and assembly of the genomes of 17 bumblebee species, representing all 15 subgenera within the genus *Bombus*. Integrating these datasets with two previously published bumblebee genomes, we performed comparative analyses of genome structures, genome contents, and gene evolutionary dynamics across the phylogeny. Our results characterize patterns of molecular and genomic evolution across the *Bombus* phylogeny and provide the first genus-wide quantification of genetic and genomic variation potentially underlying key eco-ethological traits.

Results

High quality genomic resources for all 15 *Bombus* subgenera

Sequencing and assembly strategies resulted in high quality genomic resources with 12 scaffold-level and five chromosome-level genome assemblies (Table 1). Criteria including phylogenetic position, species traits, and geographic distribution were applied to select species for whole genome sequencing from across the genus. For the five species for which sufficient samples could be collected, high-throughput chromatin conformation capture (Hi-C) (Belton et al. 2012) was used to produce chromosome-level genome assemblies (Table 1). A total of 17 species were selected (Supplementary tables Table S1; Supplementary figures Figure S1), which span all 15 subgenera of the genus *Bombus* (Williams et al. 2008). Among these, two species (*B. superbus* and *B. waltoni*) are from *Mendacibombus*, the earliest split in the *Bombus* phylogeny; four species (*B. superbus*, *B. waltoni*, *B. skorikovi*, and *B. difficillimus*) inhabit high elevations (> 4000 m above sea level); two species (*B. turneri* and *B. skorikovi*) exhibit social parasitism; three species (*B. pyrosoma*, *B. picipes* and *B. superbus*) are endemic to China; and one species (*B. polaris*) is endemic to

Arctic/subarctic regions (Williams et al. 2019). In addition, species traits including range size, tongue length, parasite incidence, and decline status vary across the selected species (Williams 1994; Arbetman et al. 2017; Cameron and Sadd 2020). Sequencing and assembly strategies included generating two Illumina sequencing datasets for each species: (i) overlapping paired-end reads (2×250 bp) from one small-insert fragment library using one single haploid drone per species (insert size: 400 or 450 bp); and (ii) paired-end reads (2×150 bp) from four large-insert jump libraries using 3-5 individuals per species (insert sizes: 4 kb, 6kb, 8kb and 10 kb, respectively; Supplementary tables Table S2). Whole-genome overlapping paired-end reads from fragment libraries were assembled into continuous sequences (contigs) using the software DISCOVAR *de novo* (Love et al. 2016), then scaffolded with reads from jump libraries using the software BESST (Sahlin et al. 2014). The resulting assemblies have a mean contig N50 of 325 Kb, ranging up to 579 Kb for *B. breviceps*; the mean scaffold N50 is 4.0 Mb, ranging up to 6.9 Mb for *B. superbus* (Table 1). Genome assembly quality in terms of expected gene content was evaluated by Benchmarking Universal Single-Copy Ortholog (BUSCO) analysis (Waterhouse et al. 2018), which showed high BUSCO completeness scores (average 99.0%, from 97.5 to 99.6%; Supplementary figures Figure S2) for all genomes. Genome annotation resulted in total protein-coding gene predictions per species ranging from 14,027–16,970 (mean = 15,838, standard deviation = 908; Supplementary tables Table S3). These were annotated using the MAKER pipeline (Cantarel et al. 2008), based on *ab initio* gene predictions, transcript evidence, and homologous protein evidence. Gene counts are similar to those of 12 drosophilid species (mean = 15,361, sd. = 852; Clark et al., 2007), but are higher than those of 19 anophelines (mean = 13,110, sd. = 1,397) (Neafsey et al. 2015), and they do not correlate significantly with assembly contiguity ($p = 0.1757$; Supplementary figures Figure S3). Between 7,299–8,135 genes were assigned at least one Gene Ontology (GO) term and 9,431–10,578 genes were annotated with at least one protein domain (Supplementary tables Table S3). BUSCO analysis of the annotated genes also showed high completeness scores for all species (Supplementary figures Figure S4). Furthermore, comprehensive miRNA, tRNA, and lncRNA gene prediction revealed an average of 93, 306, and 3,353 genes, respectively (Supplementary tables Table S3). Finally, transposable element (TE) annotation showed that the total TE content ranged

from 9.66% (22.2 Mb) in *B. superbis* to 17.88% (46.9 Mb) in *B. sibiricus* (Supplementary tables Table S4).

Genome-scale phylogeny of bumblebees

The species-level molecular phylogeny (Figure 1A) estimated from maximum-likelihood analysis with IQ-TREE (Minh et al. 2020b) is largely consistent with previously inferred phylogenetic relationships of the 15 subgenera based on five genes (Cameron et al. 2007; Williams et al. 2008), showing only two topological differences. The results support previous conclusions that: (i) subgenus *Mendacibobus* (labeled *Md* in Figure 1A) is the sister group to all the other subgenera; and (ii) species of *Psithyrus* (labeled *Ps* in purple in Figure 1A) are within the genus *Bombus*, arguing *Psithyrus* should not be named as an independent genus. The species phylogeny was built from the concatenated aligned protein sequences of 2,918 universal single-copy orthologs from 19 bumblebee species (17 from the current study, two published previously: *B. terrestris* and *B. impatiens* (Sadd et al. 2015)) and four honeybee species (*A. florea*, *A. dorsata* (Oppenheim et al. 2020), *A. cerana* (Park et al. 2015), and *A. mellifera* (Weinstock et al. 2006)), with orthologous groups delineated using the OrthoDB software (Kriventseva et al. 2015).

Complementary analysis with ASTRAL based on maximum likelihood gene trees (Zhang et al. 2018) resulted in an identical species tree with the exception of the placement of *B. pyrosoma*, which no longer forms a monophyletic pairing with *B. breviceps*, but rather forms an asymmetrical four-taxa clade with *B. breviceps*, *B. sibiricus*, and *B. cullumanus* (Supplementary figures Figure S5). The tendency for maximum likelihood concatenation to return a symmetrical four-taxa topology while ASTRAL returns an asymmetrical topology (as observed here) is a known shortcoming of maximum-likelihood concatenation in the presence of incomplete lineage sorting (ILS) (Kubatko and Degnan 2007; Mendes and Hahn 2018), implying that the ASTRAL topology is likely the correct topology.

However, inspecting the gene trees reveals extreme levels of discordance: none of their topologies match the topology of the tree inferred from concatenation (Supplementary tables Table S5 and Table S6), and nearly every gene tree has a unique topology (Supplementary tables Table S7). Such extreme levels of discordance have been observed previously in birds (Jarvis et al. 2014) and tomatoes (Pease et al. 2016), and have been attributed to a variety of sources, such as ILS and introgression (Maddison 1997). A lack of informative sites, only 24%, compared to 47% in a

similar dataset of 25 drosophilids (Da Lage et al. 2019), possibly due to the relatively recent diversification of bumblebees (Hines 2008), may also cause discordance. Gene and site concordance factor analysis (Minh et al. 2020a) was employed to quantify the amount of discordance between gene trees and the IQ-TREE species tree (node labels in Figure 1A). For each node in the IQ-TREE species tree, gene concordance factors (gCF) reflect the percentage of gene trees that contain that node as defined by its descendant taxa, and site concordance factors (sCF) reflect the percentage of informative sites that support that node via parsimony. On average across the *Bombus* phylogeny, nodes in the IQ-TREE species tree show a gCF of 38.4%, indicating that on average a node is present in only two fifths of gene trees. More stringent filtering to use gene trees with the highest bootstrap support results in higher gCF values for all nodes (Supplementary figures Figure S6). Average sCF across *Bombus* nodes is 53.6%, meaning a little over half of informative sites in the gene alignments support the nodes of the IQ-TREE species tree (node labels in Figure 1A). These site concordance factors, the short internal branches of the species tree, and the strong correlation between them (Supplementary figures Figure S7), are consistent with ILS driving the observed gene tree discordance. The possible contribution of introgression to the observed discordance among gene trees was examined using tree topologies to calculate Δ as described in (Huson et al. 2005, Vanderpool et al. 2020) for each branch in the IQ-TREE and the ASTRAL species tree that showed a gCF of less than 95%. Using bootstrap sampling of gene trees to provide a null distribution (Supplementary figures Figure S8), no lineages in either species tree showed significantly high values of Δ , ruling out introgression as a source of discordance (Supplementary figures Figure S9). Because of the high levels of discordance, gene-level phylogenies are therefore used in all subsequent gene-based molecular evolution analyses because such discordance can bias inferences of substitutions when mapped onto a species tree (Mendes and Hahn 2016).

Major genomic rearrangements in social parasites

The five Hi-C genome assemblies indicate that four of the five subgenera have 18 chromosomes (Figure 2A and 2C; Supplementary figures Figure S10A-B), consistent with previous karyotypic analysis that inferred the ancestral chromosome number is 18 (Owen et al. 1995). However, the social parasite bumblebee, *B. turneri*, subgenus *Psithyrus*, has 25 chromosomes (Figure 2B), consistent with previous cytological work (Owen 1983). Despite the higher chromosome number, its genome size is within

the range of other bumblebees (Figure 1A; Table 1). Investigating macrosynteny relationships between *B. turneri* and the other species with chromosomal-level assemblies revealed three major processes that explain how a 25-chromosome karyotype was derived from the ancestral karyotype of 18 chromosomes. First, some chromosomes descended, structurally unchanged, from ancestral chromosomes (e.g., chromosome 5; Figure 2D in blue). Second, some originated by fission of an ancestral chromosome (e.g., 11 and 25 of *B. turneri* originated by the fission of ancestral chromosome 11; Figure 2D in red). Lastly, some are derived from fusions of two or more ancestral chromosome segments (e.g., *B. turneri* chromosome 22 was derived from the fusion of segments of ancestral chromosomes 7, 8, 10, and 16 (Figure 2D in gold). Pairwise comparisons between *Psithyrus* and members of other subgenera reveal similar results, and support the inference that the 25 chromosomes of the social parasite bumblebee result from a combination of fission, fusion, and retention of ancestral chromosomes (Supplementary figures Figure S10).

Rates of chromosome evolution, in terms of rearrangements relative to *B. terrestris*, were investigated for each of the five species with chromosome-level assemblies. Rearrangement rates in bumblebees range from 0.0016–0.0075 inversions/Mb/My (Supplementary tables Table S8), which is much lower than those of drosophilids (0.013–0.159 inversions/Mb/My) and anophelins (0.052–0.068 inversions/Mb/My) (von Grotthuss et al. 2010; Neafsey et al. 2015). Thus, although bumblebee genomes have a high recombination rate (Wilfert et al. 2007), their rates of chromosome evolution are relatively slow, which is further supported by the observed high synteny contiguity across species (average 88%, from 80–95%; Supplementary tables Table S9).

Transposable elements drive genome size variation

Genome assembly sizes (haploid) range from 230 Mb in *B. superbus* to 262 Mb in *B. sibiricus* (Figure 1A). Ancestral genome size inference of bumblebees produced an estimate of 230–231 Mb, similar to that of members of the subgenus *Mendacibombus*, but smaller than the genomes of all other extant bumblebees surveyed (Supplementary figures Figure S11). Comparing genome size differences with relative content of TEs, simple sequence repeats (SSRs), and coding DNA sequences (CDS) shows that TE content explains a majority of the differences across bumblebees (Pearson correlation $R = 0.92$, $P = 1.9e-08$, $R^2=0.85$; Figure 1C, Figure 1D, Supplementary figures Figure S12). *Mendacibombus* species have a smaller genome size than other species (Figure

1A), and TEs that transposed in non-*Mendacibombus* species after divergence from *Mendacibombus* show copy numbers ranging from 1,992–4,755 (Supplementary figures Figure S13), supporting the contribution of TEs to genome size evolution. Furthermore, TE proliferation history analysis indicated that all non-*Mendacibombus* species have more recent TE amplification peaks (Supplementary figures Figure S14), consistent with increased TE activity driving genome size increases. The genomic distributions of TEs include 1,074–1,786 TE loci that overlap with the coding regions of protein-coding genes (Supplementary tables Table S10). 352 of these genes are universal single-copy orthologs across the 19 bumblebees whose overall dN/dS values are all < 1 (Supplementary tables Table S11), suggesting long-term functional constraints. One case of a putative ancient and maintained chimeric TE-gene fusion involves a gene with single-copy orthologs across the 19 bumblebees where the C-terminus of the proteins match the sequence of a reverse transcriptase of an R1 retrotransposon (Figure S15). Aligned reads from RNA-sequencing data continue with similar coverage levels into the putatively TE-derived region at the 3'-end of the gene, supporting the prediction and the expression of the full chimera. TE activity has therefore contributed to the evolution of the bumblebee protein-coding gene repertoire. In addition, there are thousands of TEs located within 1 kb of a gene in each species (Supplementary tables Table S10), and, in *B. terrestris*, 278 such TEs co-locate with open chromatin regions detected by ATAC-seq (Supplementary tables Table S12), suggesting those TEs may have become incorporated into regulatory sequences.

Gene content evolution reflects foraging and diet diversity

Orthology delineation results indicate that a majority of genes are found in one or more copies in nearly all species across bumblebees (Figure 1B). These include 53 orthologous groups specific to the *Bombus* genus, which are present in all 19 bumblebees but absent in all four honeybees (Figure 1B; Supplementary tables Table S13), and may play roles in lineage-specific traits. Functional annotation suggests that five of these *Bombus*-specific genes are associated with protein metabolism and transport (Supplementary tables Table S13), potentially linked to the higher protein content of pollen collected by bumblebees than honeybees (Leonhardt and Blüthgen 2011) or the importance of proteins for insect diapause, which is a critical step in the bumblebee life cycle (Denlinger 2002; Colgan et al. 2011). Orthologous groups with the broadest species representation are functionally enriched for core biological

processes such as protein transport, signal transduction (e.g. Wnt pathway), (de)ubiquitination, and cytoskeleton organization (Supplementary tables Table S14). In contrast, those with sparse or lineage-restricted species representation are enriched for processes including olfactory and gustatory perception, amino acid biosynthesis, and oxidation-reduction (Supplementary tables Table S14). On average, 465 species-specific genes (those without an ortholog in any other species) were identified in each bumblebee species (range 137–767) (Supplementary tables Table S15), which may contribute to species-specific traits but whose functional roles remain to be explored. Turnover analysis (gains and losses) of gene repertoires across the *Bombus* phylogeny (15 species, one per subgenus) using CAFE v3.0 (Han et al. 2013) identified expansions and contractions among 13,828 gene families and quantified variations in gene turnover rates across species (Supplementary figures Figure S16). After error correction, the overall rate of gene turnover in *Bombus* genomes is 0.0036/gene/million years, similar to an analysis of 18 anopheline species and 25 drosophilids (Supplementary tables Table S16) (Neafsey et al. 2015; Da Lage et al. 2019). However, these genus-specific gene turnover rates are 2-3 times higher than order-wide rates, which average 0.0011 (Supplementary tables Table S16) (Thomas et al. 2020), possibly due to the denser sampling in genus-level studies that allow more events to be captured. Gene gain and loss events, along with the number of rapidly evolving gene families, are summarized for each species (Supplementary tables Table S17), with a total of 3,797 rapidly changing gene families. The most dynamic gene families are enriched for processes including smell and taste perception, chitin metabolism, microtubule-based movement, and methylation (Supplementary tables Table S18). Complementary analysis using three measures of gene copy number variation also identifies these processes as enriched among the most variable gene families, in contrast to the most stable that are involved in processes related to translation, adhesion, and transport (Supplementary tables Table S19). In terms of protein domain copy number evolution, the most highly variable genes are those with protein-protein interaction mediating F-box domains, putatively DNA-binding SAP motifs, and phosphate-transferring guanylate kinases (Supplementary tables Table S20).

Stable intron-exon structures with abundant stop-codon readthrough

Protein-coding potential analysis using *B. terrestris* as the reference species identified 851 candidate readthrough stop codons (Supplementary figures Figure S17;

Supplementary tables Table S21), i.e. where translation likely continues through stop codons to produce extended protein isoforms. Coding potential was assessed using PhyloCSF (Lin et al. 2011) on whole genome alignments of all 19 bumblebees and four honeybees. The false discovery rate was estimated using enrichment for the TGA-C stop codon context, which is favored in readthrough genes, to infer that no more than 30% of the 200 highest-scoring candidates are false positives, and that at least 306 of our 851 candidates undergo functional readthrough. While readthrough is rare beyond Pancrustacea, hundreds of *Drosophila* and *Anopheles* genes undergo readthrough (Jungreis et al. 2011; Dunn et al. 2013; Jungreis et al. 2016; Rajput et al. 2019) and our whole-genome-alignment-based results support the prediction (Jungreis et al. 2011) that insect species have abundant stop-codon readthrough. In contrast, intron-exon boundaries within bumblebee genes are relatively stable. Examining evolutionary histories of intron gains and losses revealed few changes, representing only 3-4% of ancestral intron sites, with more gains than losses (Supplementary figures Figure S18; Supplementary tables Table S22), unlike drosophilids and anophelines where losses dominate (Neafsey et al. 2015), suggesting that bumblebee gene structure has remained relatively stable over the 34 million years since their last common ancestor.

Divergence and selective constraints of protein-coding genes

Bumblebee genes with elevated sequence divergence and/or relaxed constraints include processes related to smell perception, chitin metabolism, RNA processing, DNA repair, and oxidation-reduction (Figure 3). Measures of evolutionary rate (amino acid sequence divergence measured as the mean of normalized inter-species ortholog protein sequence identities) and selective constraint (dN/dS) showed similar trends among different functional categories of genes. Most genes are strongly constrained, with median estimates of dN/dS much lower than one. Assignment of GO terms and InterPro domains is usually biased towards slower-evolving, well-conserved genes (Supplementary figures Figure S19). Nevertheless, functional categories with the fastest-evolving genes are further supported and complemented by examining molecular function GO terms (Supplementary figures Figure S20A) and InterPro domains (Supplementary figures Figure S20B), which show elevated rates for odorant binding, olfactory receptor activity, chitin binding, oxidoreductase activity, serine-type endopeptidase activity, and olfactory receptor domains. GO term enrichment analysis of the slowest and fastest evolving subsets of genes, bottom and

top 20% respectively (Supplementary figures Figure S21), showed genes with the slowest evolutionary rates and the lowest dN/dS ratios were enriched for essential house-keeping biological processes and molecular functions (Supplementary tables Table S23 and Table S24). In contrast, genes with the fastest evolutionary rates were enriched for processes linked to polysaccharide biosynthesis, tRNA aminoacylation, drug binding and RNA methyltransferase activity (Supplementary tables Table S23). Genes with the highest dN/dS ratios were enriched for processes and functions including proteolysis, translation, ncRNA processing, and chitin metabolism (Supplementary tables Table S24).

Codon usage bias driven by AT content

Analysis of codon usage bias showed no evidence for selection on optimal codons, in contrast to drosophilids but similar to anophelines (Vicario et al. 2007; Neafsey et al. 2015). Instead, codon usage bias in bumblebees seems to be driven mainly by AT content, consistent with previous reports in Hymenoptera (Behura and Severson 2012). Optimal codons were estimated in each species and correlation coefficients were computed between relative synonymous codon usage (RSCU) and effective number of codons (ENC) per gene. All species have a similar preference and intensity of preference; for each amino acid, there was a consistently highly preferred codon and often a secondarily preferred one, all ending in A/T (Supplementary figures Figure S22). To test if codon usage could largely be explained by mutation bias, a linear model was used to predict Fop (frequency of optimal codon) from overall gene AT content and amino acid use. The model explained 99.2% of the Fop variation without the need to include the species origin of each gene. The AT content alone explained 81% of the variation (Supplementary figures Figure S23). Moreover, a strong correlation was observed between codon AT content and the correlation between RSCU and ENC across all species (Supplementary figures Figure S24).

Evolution of genes associated with bumblebee eco-ethology

Many ecological and environmental factors—for example, shortage of food, pathogen emergence, pesticide exposure, and climate change—are contributing to the overall decline of bumblebees worldwide (Williams et al. 2009; Goulson et al. 2015; Cameron and Sadd 2020). To begin to explore the complement of genes likely to be involved in bumblebee interactions with their environment, we examined the evolution of gene families associated with their ecology and life histories. Sampling across the *Bombus* genus enabled the first survey of natural gene repertoire diversity

of such families that are likely to be important for bumblebee adaptability and success.

Chemosensory receptor diversity: Chemosensation plays a critical role in locating food and nests, communicating with nestmates, and identifying other environmental cues (Ayasse and Jarau 2014). A search of the three major chemosensory receptor gene families—odorant receptors (ORs), gustatory receptors (GRs), and ionotropic receptors (IRs)—in the sequenced bumblebee genomes identified 3,228 genes, with mean (minimum-maximum) counts of 150 (133-165) ORs, 18 (13-22) GRs, and 22 (21-22) IRs (Supplementary tables Table S25). Only complete genes were used for gene gain and loss analysis. Despite the similarities in total OR gene counts, examples of gene gain/loss were observed in specific lineages. There was a net loss of 15 ORs in the common ancestor of the subgenus *Mendacibombus* (*Md*) (Figure 4A; Supplementary figures Figure S25). Species in *Mendacibombus* mainly inhabit high mountains including the Qinghai-Tibetan plateau, with relatively low floral diversity (Williams et al. 2018), which may be linked to OR loss in this subgenus. A net loss of 11 ORs was observed in the common ancestor of subgenus *Psithyrus* (*Ps*) (Figure 4A; Supplementary figures Figure S25). For ORs shared across bumblebees, seven showed evidence of positive selection in a subset of species, including putative pheromone receptors (Supplementary tables Table S26). Compared with ORs, GRs and IRs have much lower and more stable gene counts (Supplementary figures Figure S25). However, despite overall conservation of gene number and widespread evidence for purifying selection, there is evidence that some GR and IR genes experienced positive selection in a subset of species, including receptors putatively involved in sensing fructose and temperature (Supplementary tables Table S26).

Detoxification capacity: Detoxification genes are used to neutralize xenobiotics, such as toxic plant secondary metabolites and pesticides. Repertoires of carboxyl/cholinesterases (CCEs), cytochrome P450 monooxygenases (P450s), and glutathione S-transferases (GSTs) in the 17 genomes are much smaller than in drosophilids and anophelines (Supplementary tables Table S27), indicating a genus-wide deficit of this gene category, previously observed in two bumblebees (Sadd et al. 2015). There are 88 detoxification genes on average in bumblebees, with little variation across species (Supplementary tables Table S27). Despite overall conservation of gene number and widespread evidence for purifying selection (mean dN/dS is 0.26), a total of 19 detoxification genes, including CCEs, P450s, and GSTs,

showed evidence of positive diversifying selection in a subset of species (Supplementary tables Table S28).

Immune defense: Immune genes are involved in recognition of and defense against pathogens. Similar to detoxification genes, counts in the 17 sequenced genomes are much lower than in drosophilids and anophelins (Supplementary tables Table S29), showing that the previously noted paucity in two bumblebees (Barribeau et al. 2015; Sadd et al. 2015) extends to the whole genus. Bumblebee genomes contain components of all major immune pathways described in insects, and gene counts are fairly conserved across species (Supplementary tables Table S29). For example, all species have two genes encoding Gram-negative bacteria binding-proteins, while peptidoglycan-recognition proteins are more variable with between four and six gene copies. Comparing dN/dS ratios between immune genes and all single-copy orthologous genes in bumblebees showed that immune genes exhibit slightly higher dN/dS ratios ($P = 0.04$, Wilcoxon rank sum test), and among immune genes, recognition and signaling genes have higher dN/dS ratios than effector genes (Figure 4B). In addition, despite widespread evidence for purifying selection, a total of 52 immune genes showed evidence of positive selection in a subset of bumblebee species (Supplementary tables Table S30).

Genes involved in high-elevation adaptation: *Bombus superbus*, *B. waltoni*, *B. difficillimus*, and *B. skorikovi* are four species collected at elevations $> 4,000$ m that represent three subgenera (Figure 1). No genes show signatures of positive selection in all high-elevation species but none of the low-elevation species. However, six genes show evidence of positive selection in species representing two of the three high-elevation subgenera, but none of the low-elevation species (Supplementary tables Table S31). One encodes CPAMD8, which is involved in eye development (Cheong et al. 2016). As bumblebees detect flowers visually (Meyer-Rochow 2019), signatures of selection might be related to fine tuning eye development for optimal foraging in high altitude light conditions. Three genes encode a histone deacetylase, synaptotagmin-12, and a heterogeneous nuclear ribonucleoprotein, which are involved in maintaining muscle integrity and keeping “flight state”, which is critical for undertaking long-distance food-searching (Liu et al. 2001; Manjila et al. 2019; Pigna et al. 2019). Two genes encode a sodium-coupled monocarboxylate transporter, and a glycosyltransferase family protein, which are believed to be involved in

metabolic adaptation to hypoxia (Véga et al. 2006; Shirato et al. 2010)

(Supplementary tables Table S31).

Sex-determination: Evolutionary analysis of sex-determination genes in bumblebees and related species indicated that all bumblebee genomes share a duplicated copy of *feminizer* (*fem*), named *fem 1* (Figure 4C). Compared to *fem*, *fem 1* shows a higher level of divergence among bumblebees ($fem_{Bombus} dN/dS = 0.24$; $fem 1_{Bombus} dN/dS = 0.77$; Figure 4C). These ratios are close to the range observed for *Apis*, in which *fem* has evolved under purifying selection and the paralogous gene *complementary sex determiner* (*csd*) has evolved by neo-functionalization (Figure 4C) (Hasselmann et al. 2008). A hypothesis branch-site testing framework (RELAX; Wertheim et al. 2015), identifies evidence for relaxation of selection in *fem 1_{Bombus}* compared to *fem_{Bombus}* ($P < 0.001$, LR = 36.34). Moreover, the spurious action of diversifying selection on branches was predominantly found in *fem 1_{Bombus}* (Figure 4C). A mixed effect model of evolution (MEME; Murrell et al. 2012) was applied to identify individual sites that were subject to episodic diversifying selection, and at least 15 sites ($p < 0.05$) were found to be under positive selection, with some being located in known motifs (Supplementary figures Figure S26). The results of these selection analyses suggest that both *fem* and *fem 1* contribute to the bumblebee sex determination pathway. For the *transformer 2* (*tra-2*) gene, consistent amino acid changes between *Bombus* and *Apis* were found within the RNA recognition domain (Supplementary figures Figure S27), supporting a previous hypothesis of a regulatory modification between honeybees and bumblebees (Biewer et al. 2015).

Discussion

Comparative analysis of multiple genomes in a phylogenetic framework substantially improves the precision and sensitivity of evolutionary inference and provides robust results identifying stable and dynamic features. In this study, we performed comparative analyses of genome structures and contents, as well as global and family-targeted gene evolutionary dynamics across the phylogeny of the genus *Bombus*, using 17 annotated *de novo* assemblies and two previously published genomes.

Many attributes of bumblebee genomes are highly conserved across species. For example, overall genome size and genome structure, the number of protein-coding genes and non-coding RNAs, gene intron-exon structures, and the pattern of codon usage are all very similar across these 19 genomes. However, other aspects of genome

biology are dynamically evolving. TEs are a major contributor to genome size variation (Figure 1) as well as a potential source of coding and regulatory sequences (Supplementary tables Table S10-12). Differential gene gain and loss also contribute to gene content variation across bumblebees and lead to lineage-specific gene repertoires (Figure 4A; Supplementary figures Figure S25; Supplementary tables Table S17). Finally, for genes shared by all species, the action of positive selection is different across species (Supplementary tables Table S26, Table S28, Table S30, and Table S31), which can lead to gene functional divergence possibly reflecting key eco-ethological differences.

An exception to the otherwise overall conserved genome structure is the set of species in the subgenus *Psithyrus*. These bumblebees exhibit social parasitism; they do not have a worker caste, and it is not necessary for them to forage for nectar and pollen to provision developing larvae (Lhomme and Hines 2019). Originally, this subgenus was argued to be a separate genus due to distinct behavior and higher chromosome number, however subsequent phylogenetic analysis placed *Psithyrus* within the genus *Bombus* (Cameron et al. 2007; Williams et al. 2008). Here, based on a much larger genomic dataset, we confirm that species in the subgenus *Psithyrus* (labeled *Ps* in purple in Figure 1A) form a monophyletic group within the genus *Bombus*. In addition, we show that, although *Psithyrus* species have an increased chromosome number, their genome sizes are within the range of those of the other bumblebees (Figure 1A), and their 25 chromosomes reflect a mix of fission, fusion, and retention of the 18 ancestral bumblebee chromosomes (Figure 2; Supplementary figures Figure S10). Chromosome rearrangements (e.g., fissions, fusions, and inversions) have been posited to play roles in speciation (Ayala and Coluzzi 2005), and thus may explain the diversification and social parasitic behavior of *Psithyrus*. In addition to genome structure variation, we identified a net loss of 11 odorant receptor genes in the common ancestor of *Psithyrus* species (Figure 4), which could be a cause or consequence of their socially parasitic behavior.

Bumblebee species exhibit different food preferences (Goulson and Darvill 2004; Sikora and Kelm 2012; Somme et al. 2015), but the genetic basis underlying such variation is unknown. Like in other insects, smell and taste are used to distinguish different food sources (Kunze and Gumbert 2001; Ruedenauer et al. 2015). In this study, we found out that genes involved in olfactory and gustatory perception are

among the fastest evolving gene categories, both in copy number variation and in sequence divergence (Figure 3; Supplementary figures Figure S20; Supplementary tables Table S18-19). Therefore, the dynamic evolution of genes involved in olfactory and gustatory perception likely relates to different food preferences, improved understanding of which could inform the use of new species in agricultural settings. The importance of chemical perception for social Hymenoptera in mating (Ayasse et al. 2001), determination of reproductive status (Monnin 2006), and recognition of kin (Page and Breed 1987) could also contribute to rapid evolution of genes in this category.

Bumblebees exhibit rich morphology differences across species including mandible, labrum, tibia, and basitarsus structures, as well as patterns of keels, ridges, and grooves formed by the cuticle (Williams 1994), and they show species-specific responses to insecticides (Baron et al. 2017). Chitin is a major component of the insect cuticle and peritrophic matrix, and chitin metabolic processes are related to morphogenesis, resistance to insecticides, and the tolerance of toxins in food (Barbehenn 2001; Merzendorfer and Zimoch 2003; Zhu et al. 2016; Erlandson et al. 2019). Genes related to chitin metabolism are also among the fastest evolving functional categories in bumblebees, both in copy number variation and in sequence divergence (Figure 3; Supplementary figures Figure S20; Supplementary tables Table S18-19). These variable patterns of chitin-related gene evolution potentially underlie observed differences in exoskeleton morphological characters and insecticide resistance, but not color pattern variation, which is determined primarily by hair pigmentation (Tian et al. 2019). Across bumblebee genomes some of the fastest evolving genes are also related to processes including protein glycosylation, methylation, proteolysis, and tRNA aminoacylation for protein translation (Figure 3; Supplementary figures Figure S20; Supplementary tables Table S18-19). Protein glycosylation is involved in multiple physiological processes including growth, development, circadian rhythms, immunity, and fertility (Walski et al. 2017). tRNA aminoacylation for protein translation process are involved in response to the changing environment (Pan 2013).

Some genes that are not among the fastest evolving categories -for example, immune and detoxification genes, which are involved in the interaction of bumblebees with

external environments- show differential patterns of positive selection in subsets of species, which can lead to gene functional divergence. The positive diversifying selection found in some detoxification genes of a subset of species could also contribute to species differences in response to insecticide exposure (Baron et al. 2017). Likewise, evidence of positive selection among immune genes in some species but not others (Supplementary tables Table S30) suggests that interactions with pathogens and parasites have been important during the evolutionary history of these species, but the selective pressures imposed differ from species to species. These differences could affect susceptibility to emerging and re-emerging infectious diseases, and explain observed species-specific differences in contemporary pathogen prevalence (Cameron et al. 2016; Cameron and Sadd 2020). Taken together, identification of the fastest evolving genes and those showing patterns of differential positive selection reveals substantial genetic variation across bumblebee genera. Future experimental investigations will be required to determine how the identified genetic variation is linked to specific differences in traits such as food preference, morphogenesis, insecticide and pathogen resistance, and the response to changing environments.

In addition to our discoveries regarding protein-coding genes, we found that TE-related sequences likely contribute to the variation of coding and regulatory repertoires (Figure 1; Supplementary tables Table S10-12). Compared with non-*Mendacibombus* bumblebees, *Mendacibombus* species have smaller genomes (Figure 1) and relatively narrow geographical distributions (Williams et al. 2016). Considering TEs are the major determinant of genome size difference, with evidence that they were potentially domesticated in bumblebee genomes (Supplementary figures Figure S15; Supplementary tables Table S10-S12), TEs may be implicated in the dispersal of non-*Mendacibombus* species across the globe, as they have been in other taxa (Casacuberta and González 2013; Baduel et al. 2019; Schrader and Schmitz 2019).

More recent range expansions or contractions are driven, at least in part, by global climate change. In response to a warming climate, there is evidence that the ranges of some bumblebees are being pushed northward or to higher elevations (Ploquin et al. 2013; Kerr et al. 2015; Biella et al. 2017; Soroye et al. 2020). The sequenced genomes of species collected at high-elevation sites (> 4000 m) and others collected at low elevations (< 2000 m) (Figure 1) represent high quality genomic resources for investigating genes involved in high-elevation adaptation. For example, population-

genomics studies facilitated by the *A. cerana* reference assembly identified candidate high-altitude adaptation loci in that species (Monteromendieta et al. 2019) and the *B. impatiens* genome was used to identify climate adaptation across latitude and altitude in two montane North American bumblebee species (Jackson et al. 2020). We identified genes showing signs of positive selection in at least two subgenera of high-elevation species but not in any of the low-elevation species (Supplementary tables Table S31). These include genes putatively involved in eye development, flight muscle integrity, and metabolism, highlighting the importance of successful food-searching in high-elevation habitats where food is scarce. Signatures of positive selection in neuro-muscular genes mirror findings from the population genomic study on the two North American montane species (Jackson et al. 2020). Beyond specific genes, comparing high- and low-elevation species showed a consistent pattern of faster genome-wide evolutionary rates in those occupying lower elevations (Li et al. 2019). Exploring these trends and genes further and identifying additional genomic features linked to life at high altitudes will help to understand differential successes of bumblebee species in a changing world.

Conclusions

We have produced highly complete and accurate genome assemblies of 17 bumblebee species, including representatives from all 15 *Bombus* subgenera. Our genus-wide comparative analysis of bumblebee genomes revealed how genome structures, genome contents, and gene evolutionary dynamics vary across bumblebees, and identified genetic variations that may underlie species trait differences in foraging, diet and metabolism, morphology and insecticide resistance, immunity and detoxification, as well as adaptations for life at high altitudes. Our work provides genomic resources that capture genetic and phenotypic variation, which should advance our understanding of bumblebee success and help identify potential threats. These resources form a foundation for future research, including resequencing and population genomics studies for functional gene positioning and cloning, which will inform the use of bumblebees in agriculture, as well as the design of strategies to prevent the decline of these important pollinators.

Materials and Methods

For detailed methods, please see Supplementary Materials and Methods. Genomic DNA purified from one single haploid drone of each species was used to generate one “fragment” library to produce overlapping paired-end shotgun reads (2×250 bp). Genomic DNA purified from 3-5 individuals of each species was used to generate four “jumping” libraries (insert sizes: 4 kb, 6 kb, 8 kb, and 10 kb, respectively). For each species, the 250 bp overlapping paired-end shotgun reads from the fragment library were assembled using the software DISCOVAR *de novo* (Love et al. 2016) to produce contiguous sequences (contigs). Shotgun reads from jumping libraries were used to scaffold the contigs using the software BESST (Sahlin et al. 2014). Hi-C sequencing libraries were generated as described previously (Belton et al. 2012; Lin et al. 2018), and the 3D-DNA pipeline (Dudchenko et al. 2017) was applied to assemble the scaffold sequences to chromosome level. The contiguity of the genome assemblies was compared with other genomic features (Supplementary figures Figure S28). The completeness of the genome assemblies was evaluated using BUSCO v3 (Waterhouse et al. 2018). Transposable elements were *de novo* identified by LTRharvest (Ellinghaus et al. 2008), MGEScan-non-LTR (Rho and Tang 2009) and RepeatScout (Price et al. 2005). Protein-coding genes were annotated with the MAKER pipeline (Cantarel et al. 2008), based on *ab initio* gene predictions, transcript evidence (Supplementary tables Table S33), and homologous protein evidence, and assessed with BUSCO v3 (Waterhouse et al. 2018). Orthologous groups were delineated by using the OrthoDB software (Kriventseva et al. 2015), which was also used to compute evolutionary rates. IQ-TREE (Minh et al. 2020b) and ASTRAL (Zhang et al. 2018) were employed for phylogeny analysis. Testing for introgression employed the Δ statistic (Huson et al. 2005), which follows the same logic as the ABBA-BABA site patterns used to calculate D-statistics, but uses tree topologies instead of alignment sites. MCScanX (Wang et al. 2012) was used to perform gene synteny analysis. Global gene family evolution was estimated with CAFE v3.0 (Han et al. 2013).

PAML 4 (Yang 2007) was used to detect positive selection and calculate dN/dS ratios across alignments of single-copy orthologs. This was followed by aBSREL (Smith et al. 2015) analysis for gene families of special interest (i.e. chemosensory genes, detoxification genes, immune genes, sex-determination genes, piRNA pathway genes) and for genes identified in the initial PAML scans as potentially showing signatures of adaptation to high elevation. Sex-determination genes were additionally analyzed using RELAX (Wertheim et al. 2015) and MEME (Murrell et al. 2012). Intron history analyses were performed using Malin (Csúros 2008). Whole genome alignments were produced using Cactus (Paten et al. 2011). PhyloCSF (Lin et al. 2011) was used to study stop codon readthrough. Codon usage bias was analyzed as described previously (Vicario et al. 2007).

Acknowledgements:

We thank Dr. Jiandong An (Institute of Apicultural Research, Chinese Academy of Agricultural Sciences) for providing excellent suggestions and assistance with bumblebee selection and collection. We gratefully acknowledge Olav Rueppell, members of the dwarf honeybee genome project, and the Baylor College of Medicine Human Genome Sequencing Center for pre-publication access to the *Apis florea* genome and annotations. We would like to thank two anonymous reviewers for their valuable feedback and suggestions on this paper. This work was supported by the National Natural Science Foundation of China (grant number 31971397), the Science and Technology Innovation Project of Chinese Academy of Agricultural Sciences (grant number CAAS-ASTIP-2019-IAR), the Fundamental Research Funds for Central Non-profit Scientific Institution of China (grant number Y2019XK13), the National Human Genome Research Institute of the National Institutes of Health (grant number U41HG007234), the National Institutes of Health (grant number R01HG004037), Novartis Foundation for medical-biological research (grant number 18B116), Swiss National Science Foundation (grant number PP00P3_170664), and

National Science Foundation (grant number DBI-1564611). No funding body participated in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Author contributions:

CS, RMW and RLM designed the study. JH, PHW, YL, CL, WX, HY and CS collected bumblebee samples; Xiaomeng Z and YC generated sequencing libraries; MZ and GC constructed Hi-C libraries; CS, QG, Ling M, Xiaomeng Z, QW, MZ and LD produced genome assemblies; Xingtian Z, LP-D, DM, Lina M and Zhang Z performed genome annotation. CS, Hongbo Z, Huahao Z, Ze Z and QW characterized transposable elements. LS and CS performed orthology delineation. GWCT, YW, MWH and Ze Z performed phylogeny analyses. CS, IVS, SMB, LS, and Xiaomeng Z investigated genome structure evolution. CNK and BP built whole genome alignments. GWCT, MWH, RMW, LR and CS investigated gene content and protein domain evolution. LS, LW, YW, LR and RMW analyzed the molecular evolution of gene functional categories. RLM and MWI performed piRNA analysis. IJ and MK performed stop codon readthrough analysis. RMW analyzed intron evolution. SV carried out codon bias analysis. CS and Xingtian Z performed gene synteny analysis. Xiaomeng Z, HMR and YZ characterized chemoreceptor genes. CS, Xiaomeng Z, Lauren M, and SMB characterized detoxification and immune genes. CS explored genes related to high-elevation adaptation. MH characterized sex-determination genes. CS, RMW, RLM, SS, SMB and BS wrote the manuscript, with input from all authors. All authors read and approved the final manuscript.

Competing interest Declaration:

The authors declare that they have no competing interests.

Data and Code Availability Statement:

All the data generated in this study, including raw sequence data and genome assemblies, have been submitted to GenBank at National Center for Biotechnology Information (NCBI), under BioProject number PRJNA659133. The dataset has also been submitted to the National Genomics Data Center, Beijing Institute of Genomics, Chinese Academy of Sciences, under BioProject number PRJCA002292. The code used in this study has been deposited at the GitHub repository

<https://github.com/mibiledna/Genus-wide-characterization-of-bumblebee-genomes>.

References

- Alem S, Perry CJ, Zhu X, Loukola OJ, Ingraham T, Søvik E, Chittka L. 2016. Associative mechanisms allow for social learning and cultural transmission of string pulling in an insect. *PLoS Biol.* 14:e1002564.
- Arbetman MP, Gleiser G, Morales CL, Williams P, Aizen MA. 2017. Global decline of bumblebees is phylogenetically structured and inversely related to species range size and pathogen incidence. *Proc Biol Sci.* 284:20170204.
- Ayala FJ, Coluzzi M. 2005. Chromosome speciation: humans, *Drosophila*, and mosquitoes. *Proc Natl Acad Sci USA.* 102:6535-6542. doi: 10.1073/pnas.0501847102
- Ayasse M, Jarau S. 2014. Chemical ecology of bumble bees. *Annu Rev Entomol.* 59:299-319.
- Ayasse M, Paxton RJ, Tengo J. 2001. Mating behavior and chemical communication in the order Hymenoptera. *Annu Rev Entomol.* 46:31-78.
- Bacon A, Harris A. 2004. Hypoxia-inducible factors and hypoxic cell death in tumour physiology. *Annals of medicine.* 36:530-539.
- Baduel P, Quadrana L, Hunter B, Bomblies K, Colot V. 2019. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. *Nat Commun.* 10:1-10.
- Barbehenn RV. 2001. Roles of peritrophic membranes in protecting herbivorous insects from ingested plant allelochemicals. *Arch Insect Biochem Physiol.* 47:86-99.
- Baron GL, Raine NE, Brown MJF. 2017. General and species-specific impacts of a neonicotinoid insecticide on the ovary development and feeding of wild bumblebee queens. *Proc Biol Sci.* 284:20170123.
- Barribeau SM, Sadd BM, du Plessis L, Brown MJ, Buechel SD, Cappelle K, Carolan JC, Christiaens O, Colgan TJ, Erler S, et al. 2015. A depauperate immune repertoire precedes evolution of sociality in bees. *Genome Biol.* 16:83. doi: 10.1186/s13059-015-0628-y
- Bartomeus I, Ascher JS, Gibbs J, Danforth BN, Wagner DL, Hedtke SM, Winfree R. 2013. Historical changes in northeastern US bee pollinators related to shared ecological traits. *Proc Natl Acad Sci USA.* 110:4656-4660.
- Behura SK, Severson DW. 2012. Comparative analysis of codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes. *PLoS One* 7:e43111. doi: 10.1371/journal.pone.0043111
- Belton J-M, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. 2012. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* 58:268-276.
- Biella P, Bogliani G, Cornalba M, Manino A, Neumayer J, Porporato M, Rasmont P, Milanese P. 2017. Distribution patterns of the cold adapted bumblebee *Bombus alpinus* in the Alps and hints of an uphill shift (Insecta: Hymenoptera: Apidae). *J Insect Conserv.* 21:357-366.

- Biewer M, Schlesinger F, Hasselmann M. 2015. The evolutionary dynamics of major regulators for sexual development among Hymenoptera species. *Front Genet.* 6:124.
- Cameron SA, Hines HM, Williams PH. 2007. A comprehensive phylogeny of the bumble bees (*Bombus*). *Biol J Linn Soc.* 91:161-188.
- Cameron SA, Lim HC, Lozier JD, Duennes MA, Thorp RW. 2016. Test of the invasive pathogen hypothesis of bumble bee decline in North America. *Proc Natl Acad Sci USA.* 113:4386-4391.
- Cameron SA, Lozier JD, Strange JP, Koch JB, Cordes N, Solter LF, Griswold TL. 2011. Patterns of widespread decline in North American bumble bees. *Proc Natl Acad Sci USA.* 108:662-667.
- Cameron SA, Sadd BM. 2020. Global Trends in Bumble Bee Health. *Annu Rev Entomol.* 65:209-232.
- Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* 18:188-196.
- Casacuberta E, González J. 2013. The impact of transposable elements in environmental adaptation. *Mol Ecol.* 22:1503-1517.
- Cheong S-S, Hentschel L, Davidson AE, Gerrelli D, Davie R, Rizzo R, Pontikos N, Plagnol V, Moore AT, Sowden JC. 2016. Mutations in CPAMD8 cause a unique form of autosomal-recessive anterior segment dysgenesis. *Am J Hum Genet.* 99:1338-1352.
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450:203-218. doi: 10.1038/nature06341
- Colgan TJ, Carolan JC, Bridgett S, Sumner S, Blaxter M, Brown MJF. 2011. Polyphenism in social insects: insights from a transcriptome-wide analysis of gene expression in the life stages of the key pollinator, *Bombus terrestris*. *BMC Genomics* 12:623.
- Csűros M. 2008. Malin: maximum likelihood analysis of intron evolution in eukaryotes. *Bioinformatics* 24:1538-1539.
- Da Lage J-L, Thomas GW, Bonneau M, Courtier-Orgogozo V. 2019. Evolution of salivary glue genes in *Drosophila* species. *BMC Evol Biol.* 19:36.
- Denlinger DL. 2002. Regulation of diapause. *Annu Rev Entomol.* 47:93-122.
- Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP. 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356:92-95.
- Dunn JG, Foo CK, Belletier NG, Gavis ER, Weissman JS. 2013. Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *Elife* 2:e01179.

- Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC bioinformatics* 9:18.
- Erlandson MA, Toprak U, Hegedus DD. 2019. Role of the peritrophic matrix in insect-pathogen interactions. *J Insect Physiol.* 117:103894. doi: 10.1016/j.jinsphys.2019.103894
- Fontaine C, Dajoz I, Meriguet J, Loreau M. 2005. Functional diversity of plant–pollinator interaction webs enhances the persistence of plant communities. *PLoS Biol.* 4:e1.
- Garibaldi LA, Steffan-Dewenter I, Winfree R, Aizen MA, Bommarco R, et al. 2013. Wild Pollinators Enhance Fruit Set of Crops Regardless of Honey Bee Abundance. *Science* 339:1608-1611.
- Goulson D, Darvill B. 2004. Niche overlap and diet breadth in bumblebees; are rare species more specialized in their choice of flowers? *Apidologie* 35:55-63. doi: 10.1051/apido:2003062
- Goulson D, Lye GC, Darvill B. 2008. Decline and conservation of bumble bees. *Annu Rev Entomol.* 53:191-208. doi: 10.1146/annurev.ento.53.103106.093454
- Goulson D, Nicholls E, Botias C, Rotheray EL. 2015. Bee declines driven by combined stress from parasites, pesticides, and lack of flowers. *Science* 347:1255957. doi: 10.1126/science.1255957
- Grixti JC, Wong LT, Cameron SA, Favret C. 2009. Decline of bumble bees (*Bombus*) in the North American Midwest. *Biol Conserv.* 142:75-84. doi: 10.1016/j.biocon.2008.09.027
- Han MV, Thomas GW, Lugo-Martinez J, Hahn MW. 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol Biol Evol.* 30:1987-1997. doi: 10.1093/molbev/mst100
- Hasselmann M, Gempe T, Schiøtt M, Nunes-Silva CG, Otte M, Beye M. 2008. Evidence for the evolutionary nascence of a novel sex determination pathway in honeybees. *Nature* 454:519-522.
- Hines HM. 2008. Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: Apidae: *Bombus*). *Syst Biol.* 57:58-75. doi: 10.1080/10635150801898912
- Huson DH, Klöpper T, Lockhart PJ, Steel MA editors.; 2005 Berlin, Heidelberg.
- Jackson JM, Pimsler ML, Oyen KJ, Strange JP, Dillon ME, Lozier JD. 2020. Local adaptation across a complex bioclimatic landscape in two montane bumble bee species. *Mol Ecol.* 29:920-939.
- Jarvis ED, Mirarab S, Aberer AJ, Bo L, Houde P, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320-1331.
- Jungreis I, Chan CS, Waterhouse RM, Fields G, Lin MF, Kellis M. 2016. Evolutionary dynamics of abundant stop codon readthrough. *Mol Biol Evol.* 33:3108-3132.

- Jungreis I, Lin MF, Spokony R, Chan CS, Negre N, Victorsen A, White KP, Kellis M. 2011. Evidence of abundant stop codon readthrough in *Drosophila* and other metazoa. *Genome Res.* 21:2096-2113.
- Kerr JT, Pindar A, Galpern P, Packer L, Potts SG, Roberts SM, Rasmont P, Schweiger O, Colla SR, Richardson LL. 2015. Climate change impacts on bumblebees converge across continents. *Science* 349:177-180.
- Koch JB, Lozier J, Strange JP, Ikerd H, Griswold T, Cordes N, Solter L, Stewart I, Cameron SA. 2015. USBombus, a database of contemporary survey data for North American Bumble Bees (Hymenoptera, Apidae, *Bombus*) distributed in the United States. *Biodivers Data J.* e6833. doi: 10.3897/BDJ.3.e6833
- Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simao FA, Pozdnyakov IA, Ioannidis P, Zdobnov EM. 2015. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. *Nucleic acids research* 43:D250-D256.
- Kubatko LS, Degnan JH. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst Biol.* 56:17-24.
- Kunze J, Gumbert A. 2001. The combined effect of color and odor on flower choice behavior of bumble bees in flower mimicry systems. *Behavioral Ecology* 12:447-456.
- Leonhardt SD, Blüthgen N. 2011. The same, but different: pollen foraging in honeybee and bumblebee colonies. *Apidologie* 43:449-464. doi: 10.1007/s13592-011-0112-y
- Lhomme P, Hines HM. 2019. Ecology and evolution of cuckoo bumble bees. *Annals of the Entomological Society of America* 112:122-140.
- Lin D, Hong P, Zhang S, Xu W, Jamal M, Yan K, Lei Y, Li L, Ruan Y, Fu ZF, et al. 2018. Digestion-ligation-only Hi-C is an efficient and cost-effective method for chromosome conformation capture. *Nat Genet.* 50:754-763. doi: 10.1038/s41588-018-0111-2
- Lin G, Huang Z, Wang L, Chen Z, Zhang T, Gillman LN, Zhao F. 2019. Evolutionary Rates of Bumblebee Genomes Are Faster at Lower Elevations. *Mol Biol Evol.* 36(6):1215-1219.
- Lin MF, Jungreis I, Kellis M. 2011. PhyloCSF: a comparative genomics method to distinguish protein coding and non-coding regions. *Bioinformatics* 27:i275-i282.
- Liu J, Beqaj S, Yang Y, Honoré B, Schuger L. 2001. Heterogeneous nuclear ribonucleoprotein-H plays a suppressive role in visceral myogenesis. *Mechanisms of development* 104:79-87.
- Love RR, Weisenfeld NI, Jaffe DB, Besansky NJ, Neafsey DE. 2016. Evaluation of DISCOVAR de novo using a mosquito sample for cost-effective short-read genome assembly. *BMC Genomics.* 17:187. doi: 10.1186/s12864-016-2531-7
- Maddison WP. 1997. Gene trees in species trees. *Syst Biol.* 46:523-536.

- Manjila SB, Kuruvilla M, Ferveur J-F, Sane SP, Hasan G. 2019. Extended flight bouts require disinhibition from GABAergic mushroom body neurons. *Curr Biol.* 29:283-293. e285.
- Martin CD, Fountain MT, Brown MJF. 2019. Varietal and seasonal differences in the effects of commercial bumblebees on fruit quality in strawberry crops. *Agric Ecosyst Environ.* 281:124-133. doi: 10.1016/j.agee.2019.04.007
- Mendes FK, Hahn MW. 2016. Gene tree discordance causes apparent substitution rate variation. *Syst Biol.* 65:711-721.
- Mendes FK, Hahn MW. 2018. Why concatenation fails near the anomaly zone. *Syst Biol.* 67:158-169.
- Merzendorfer H, Zimoch L. 2003. Chitin metabolism in insects: structure, function and regulation of chitin synthases and chitinases. *J Exp Biol.* 206:4393-4412. doi: 10.1242/jeb.00709
- Meyer-Rochow V. 2019. Eyes and Vision of the Bumblebee: a Brief Review on how Bumblebees Detect and Perceive Flowers. *Journal of Apiculture* 34:107-115.
- Minh BQ, Hahn MW, Lanfear R. 2020a. New methods to calculate concordance factors for phylogenomic datasets. *Mol Biol Evol.* msaa106. doi: 10.1093/molbev/msaa106
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020b. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 37:1530-1534.
- Monnin T. 2006. Chemical recognition of reproductive status in social insects. *Ann Zool Fennici.* 43:515-530.
- Monteromendieta S, Tan K, Christmas MJ, Olsson A, Vila C, Wallberg A, Webster MT. 2019. The genomic basis of adaptation to high-altitude habitats in the eastern honey bee (*Apis cerana*). *Mol Ecol.* 28:746-760.
- Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK. 2012. Detecting Individual Sites Subject to Episodic Diversifying Selection. *PLoS Genet.* 8:e1002764.
- Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, Amon J, Arca B, Arensburger P, Artemov G, et al. 2015. Mosquito genomics. Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science* 347:1258522. doi: 10.1126/science.1258522
- Oppenheim S, Cao X, Rueppel O, Krongdang S, Phokasem P, DeSalle R, Goodwin S, Xing J, Chantawannakul P, Rosenfeld JA. 2020. Whole Genome Sequencing and Assembly of the Asian Honey Bee *Apis dorsata*. *Genome Biol Evol.* 12:3677-3683.
- Owen RE. 1983. Chromosome numbers of 15 North American bumble bee species (Hymenoptera, Apidae, Bombini). *Canadian Journal of Genetics and Cytology* 25:26-29.

- Owen RE, Richards KW, Wilkes A. 1995. Chromosome Numbers and Karyotypic Variation in Bumble Bees (Hymenoptera: Apidae; Bombini). *J Kansas Entomol Soc.* 68:290-302.
- Page RE, Breed MD. 1987. Kin recognition in social bees. *Trends Ecol Evol.* 2:272-275.
- Pan T. 2013. Adaptive translation as a mechanism of stress response and adaptation. *Annu Rev Genet.* 47:121-137. doi: 10.1146/annurev-genet-111212-133522
- Park D, Jung JW, Choi B-S, Jayakodi M, Lee J, Lim J, Yu Y, Choi Y-S, Lee M-L, Park Y, et al. 2015. Uncovering the novel characteristics of Asian honey bee, *Apis cerana*, by whole genome sequencing. *BMC Genomics* 16:1.
- Pashalidou FG, Lambert H, Peybernes T, Mescher MC, De Moraes CM. 2020. Bumble bees damage plant leaves and accelerate flower production when pollen is scarce. *Science* 368:881-884.
- Paten B, Earl D, Nguyen N, Diekhans M, Zerbino DR, Haussler D. 2011. Cactus: Algorithms for genome multiple sequence alignment. *Genome Res.* 21:1512-1528.
- Pease JB, Haak DC, Hahn MW, Moyle LC. 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. *PLoS Biol.* 14:e1002379.
- Persson AS, Rundlöf M, Clough Y, Smith HG. 2015. Bumble bees show trait-dependent vulnerability to landscape simplification. *Biodivers Conserv.* 24:3469-3489. doi: 10.1007/s10531-015-1008-3
- Pigna E, Simonazzi E, Sanna K, Bernadzki KM, Proszynski T, Heil C, Palacios D, Adamo S, Moresi V. 2019. Histone deacetylase 4 protects from denervation and skeletal muscle atrophy in a murine model of amyotrophic lateral sclerosis. *EBioMedicine* 40:717-732.
- Ploquin EF, Herrera JM, Obeso JR. 2013. Bumblebee community homogenization after uphill shifts in montane areas of northern Spain. *Oecologia* 173:1649-1660.
- Potts SG, Biesmeijer JC, Kremen C, Neumann P, Schweiger O, Kunin WE. 2010. Global pollinator declines: trends, impacts and drivers. *Trends Ecol Evol.* 25:345-353. doi: 10.1016/j.tree.2010.01.007
- Rajput B, Pruitt KD, Murphy TD. 2019. RefSeq curation and annotation of stop codon recoding in vertebrates. *Nucleic acids research* 47:594-606.
- Rho M, Tang H. 2009. MGEScan-non-LTR: computational identification and classification of autonomous non-LTR retrotransposons in eukaryotic genomes. *Nucleic acids research* 37:e143.
- Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics* 21:351-358.
- Ruedenauer FA, Spaethe J, Leonhardt SD. 2015. How to know which food is good for you: bumblebees use taste to discriminate between different concentrations of food differing in nutrient content. *J Exp Biol.* 218:2233-2240. doi: 10.1242/jeb.118554

- Sadd BM, Barribeau SM, Bloch G, Graaf DCd, Dearden P, Elsik CG, Gadau J, Grimmelikhuijzen CJ, Hasselmann M, Lozier JD, et al. 2015. The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol.* 16:76.
- Sahlin K, Vezzi F, Nystedt B, Lundeberg J, Arvestad L. 2014. BESST-efficient scaffolding of large fragmented assemblies. *BMC bioinformatics* 15:281.
- Schrader L, Schmitz J. 2019. The impact of transposable elements in adaptive evolution. *Mol Ecol.* 28:1537-1549.
- Shirato K, Nakajima K, Korekane H, Takamatsu S, Gao C, Angata T, Ohtsubo K, Taniguchi N. 2010. Hypoxic regulation of glycosylation via the N-acetylglucosamine cycle. *Journal of clinical biochemistry and nutrition* 48:20-25.
- Sikora A, Kelm M. 2012. Flower Preferences of the Wrocław Botanical Garden Bumblebees (*Bombus* spp.). *Journal of Apicultural Science* 56:27-36. doi: 10.2478/v10289-012-0021-y
- Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Pond SLK. 2015. Less Is More: An Adaptive Branch-Site Random Effects Model for Efficient Detection of Episodic Diversifying Selection. *Mol Biol Evol.* 32:1342-1353.
- Somme L, Vanderplanck M, Michez D, Lombaerde I, Moerman R, Wathélet B, Wattiez R, Lognay G, Jacquemart AL. 2015. Pollen and nectar quality drive the major and minor floral choices of bumble bees. *Apidologie* 46:92-106.
- Soroye P, Newbold T, Kerr J. 2020. Climate change contributes to widespread declines among bumble bees across continents. *Science* 367:685-688.
- Thomas GWC, Dohmen E, Hughes DST, Murali SC, Poelchau M, Glastad K, Anstead CA, Ayoub NA, Batterham P, Bellair M, et al. 2020. Gene content evolution in the arthropods. *Genome Biol.* 21:15. doi: 10.1186/s13059-019-1925-7
- Tian L, Rahman SR, Ezray BD, Franzini L, Strange JP, Lhomme P, Hines HM. 2019. A homeotic shift late in development drives mimetic color variation in a bumble bee. *Proc Natl Acad Sci USA.* 116(24):11857-11865.
- Vanderpool D, Minh BQ, Lanfear R, Hughes D, Murali S, Harris RA, Raveendran M, Muzny DM, Gibbs RA, Worley KC, et al. 2020. Primate phylogenomics uncovers multiple rapid radiations and ancient interspecific introgression. *bioRxiv* 2020.2004.2015.043786.
- Véga C, R. Sachleben Jr L, Gozal D, Gozal E. 2006. Differential metabolic adaptation to acute and long-term hypoxia in rat primary cortical astrocytes. *J Neurochem.* 97:872-883.
- Velthuis HHW, Van Doorn A. 2006. A century of advances in bumblebee domestication and the economic and environmental aspects of its commercialization for pollination. *Apidologie* 37:421-451. doi: 10.1051/apido:2006019
- Vicario S, Moriyama EN, Powell JR. 2007. Codon usage in twelve species of *Drosophila*. *BMC Evol Biol.* 7:226. doi: 10.1186/1471-2148-7-226

- von Grotthuss M, Ashburner M, Ranz JM. 2010. Fragile regions and not functional constraints predominate in shaping gene organization in the genus *Drosophila*. *Genome Res.* 20:1084-1096.
- Walski T, De Schutter K, Van Damme EJM, Smagghe G. 2017. Diversity and functions of protein glycosylation in insects. *Insect Biochem Mol Biol.* 83:21-34. doi: 10.1016/j.ibmb.2017.02.005
- Wang Y, Tang H, D DJ, Xu T, Li J, Wang X, Tae-ho L, Jin H, Barry M, Hui G. 2012. MCSanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic acids research* 40:e49.
- Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol.* 35:543-548.
- Weinstock GM, Robinson GE, Gibbs RA, Weinstock GM, Weinstock GM, Robinson GE, Worley KC, Evans JD, Maleszka R, Robertson HM. 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443:931-949.
- Wertheim JO, Murrell B, Smith MD, Pond SLK, Scheffler K. 2015. RELAX: Detecting Relaxed Selection in a Phylogenetic Framework. *Mol Biol Evol.* 32:820-832.
- Wilfert L, Gadau J, Schmid-Hempel P. 2007. Variation in genomic recombination rates among animal taxa and the case of social insects. *Heredity (Edinb)* 98:189-197. doi: 10.1038/sj.hdy.6800950
- Williams PH. 1985. A preliminary cladistic investigation of relationships among the bumble bees (Hymenoptera, Apidae). *Systematic Entomology* 10:239-255.
- Williams P, Colla S, Xie Z. 2009. Bumblebee vulnerability: common correlates of winners and losers across three continents. *Conserv Biol.* 23:931-940. doi: 10.1111/j.1523-1739.2009.01176.x
- Williams PH. 1998. An Annotated Checklist of Bumble Bees with an Analysis of Patterns of Description (Hymenoptera: Apidae, Bombini). *Bulletin of the Natural History Museum Entomology* 67:79-152.
- Williams PH. 1994. Phylogenetic relationships among bumble bees (*Bombus* Latr.): a reappraisal of morphological evidence. *Systematic Entomology* 19:327-344.
- Williams PH, Berezin MV, Cannings SG, Cederberg B, Ødegaard F, Rasmussen C, Richardson LL, Rykken J, Sheffield CS, Thanosing C. 2019. The arctic and alpine bumblebees of the subgenus *Alpinobombus* revised from integrative assessment of species' gene coalescents and morphology (Hymenoptera, Apidae, *Bombus*). *Zootaxa* 4625:1-68.
- Williams PH, Cameron SA, Hines HM, Cederberg B, Rasmont P. 2008. A simplified subgeneric classification of the bumblebees (genus *Bombus*). *Apidologie* 39:46-74. doi: 10.1051/apido:2007052
- Williams PH, Huang J, Rasmont P, An J. 2016. Early-diverging bumblebees from across the roof of the world: the high-mountain subgenus *Mendacibombus* revised

- from species' gene coalescents and morphology (Hymenoptera, Apidae). *Zootaxa* 4204:1-72. doi: 10.11646/zootaxa.4204.1.1
- Williams PH, Huang J, An J. 2017. Bear wasps of the Middle Kingdom: a decade of discovering China's bumblebees. *Antenna* 41 (1): 21-24.
- Williams PH, Lobo JM, Meseguer AS. 2018. Bumblebees take the high road: climatically integrative biogeography shows that escape from Tibet, not Tibetan uplift, is associated with divergences of present-day *Mendacibombus*. *Ecography* 41:461-477. doi: 10.1111/ecog.03074
- Williams PH, Osborne JL. 2009. Bumblebee vulnerability and conservation worldwide. *Apidologie* 40:367-387. doi: 10.1051/apido/2009025
- Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol Biol Evol.* 24:1586-1591.
- Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC bioinformatics* 19:15-30.
- Zhu KY, Merzendorfer H, Zhang W, Zhang J, Muthukrishnan S. 2016. Biosynthesis, turnover, and functions of chitin in insects. *Annu Rev Entomol.* 61:177-196.

Table 1. Genome assembly results of the 17 newly sequenced bumblebees.

	Contig Size (Mb)	Contig N50 (Kb)	Scaffold Size (Mb)	Scaffold N50 (Mb)	Chromosome Size (Mb)	Chromosome N50 (Mb)
<i>B. superbus</i>	229.84	441.61	230.16	6.90	NA	NA
<i>B. waltoni</i>	230.89	430.54	231.17	4.66	NA	NA
<i>B. confusus</i>	238.52	227.26	239.12	3.26	NA	NA
<i>B. haemorrhoidalis</i>	239.34	572.47	239.59	4.74	240.54	15.09
<i>B. ignitus</i>	240.60	374.12	241.36	3.02	242.57	15.19
<i>B. skorikovi</i>	241.25	225.53	242.05	4.34	NA	NA
<i>B. opulentus</i>	241.99	267.78	242.38	2.42	NA	NA
<i>B. turneri</i>	242.39	212.53	243.01	4.34	243.11	9.70
<i>B. soroensis</i>	243.19	244.99	243.68	2.12	NA	NA
<i>B. polaris</i>	245.17	152.35	245.82	2.25	NA	NA
<i>B. breviceps</i>	246.03	578.55	246.41	4.04	248.12	14.71
<i>B. cullumanus</i>	246.56	422.80	247.01	4.58	NA	NA
<i>B. difficillimus</i>	247.45	177.31	248.33	2.07	NA	NA
<i>B. consobrinus</i>	248.56	284.90	249.09	4.77	NA	NA
<i>B. pyrosoma</i>	251.86	472.32	252.70	6.07	254.80	15.22
<i>B. picipes</i>	253.31	185.91	254.01	5.88	NA	NA
<i>B. sibiricus</i>	261.72	253.94	262.49	3.14	NA	NA

Mb, megabase; Kb, kilobase; NA, not applicable.

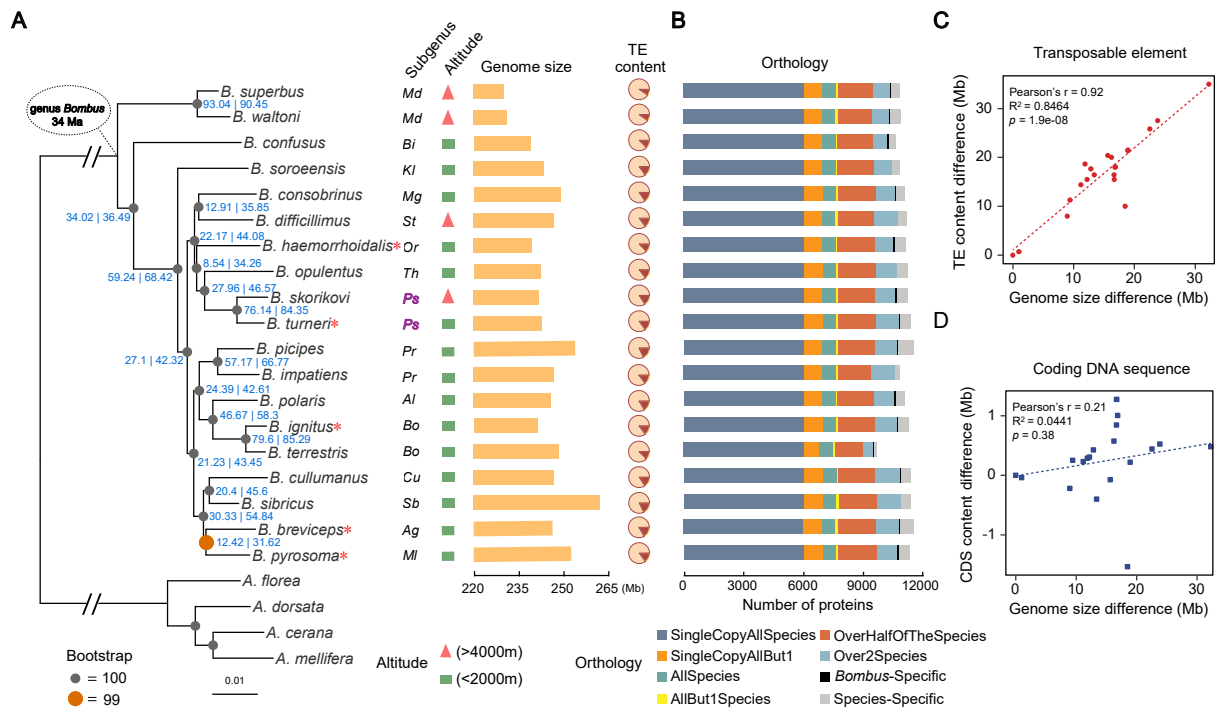


Figure 1. Phylogenetic, genomic and proteomic comparisons of 19 bumblebee species representing all 15 *Bombus* subgenera. **(A)** From left to right: the maximum likelihood molecular species phylogeny built from 2,918 concatenated single-copy orthologous groups from all sequenced bumblebees and honeybee outgroups using IQ-TREE. Node labels in blue are of the following format: gene concordance factors | site concordance factors. Branches scaled by relative number of substitutions; Red asterisks after species names indicate the five species with chromosomal-level assemblies; the subgenus that each bumblebee species belongs to (Md, *Mendacibombus*; Bi, *Bombias*; Kl, *Kallobombus*; Mg, *Megabombus*; St, *Subterraneobombus*; Or, *Orientalibombus*; Th, *Thoracobombus*; Ps, *Psithyrus*; Cu, *Cullumanobombus*; Sb, *Sibiricobombus*; Ag, *Alpigenobombus*; Ml, *Melanobombus*; Pr, *Pyrobombus*; Al, *Alpinobombus*; Bo, *Bombus*); altitude of species collection site (red triangle: extreme high-altitude; green rectangle: low-altitude); genome assembly size of each sequenced species; fraction of transposable elements (TE) (brown) in each genome. **(B)** Bar plots show total gene counts for each bumblebee partitioned according to their orthology profiles, from ancient genes found across bumblebees to lineage-restricted and species-specific genes. **(C)** and **(D)** represent the contribution of transposable element and coding DNA sequence to genome size variation across bumblebees, respectively. Differences in the total content of transposable elements **(C)** and coding DNA sequences **(D)** of the 19 genomes relative to that of *B. superbus* (which has the smallest genome assembly size) are plotted against their genome size differences (relative to that of *B. superbus*).

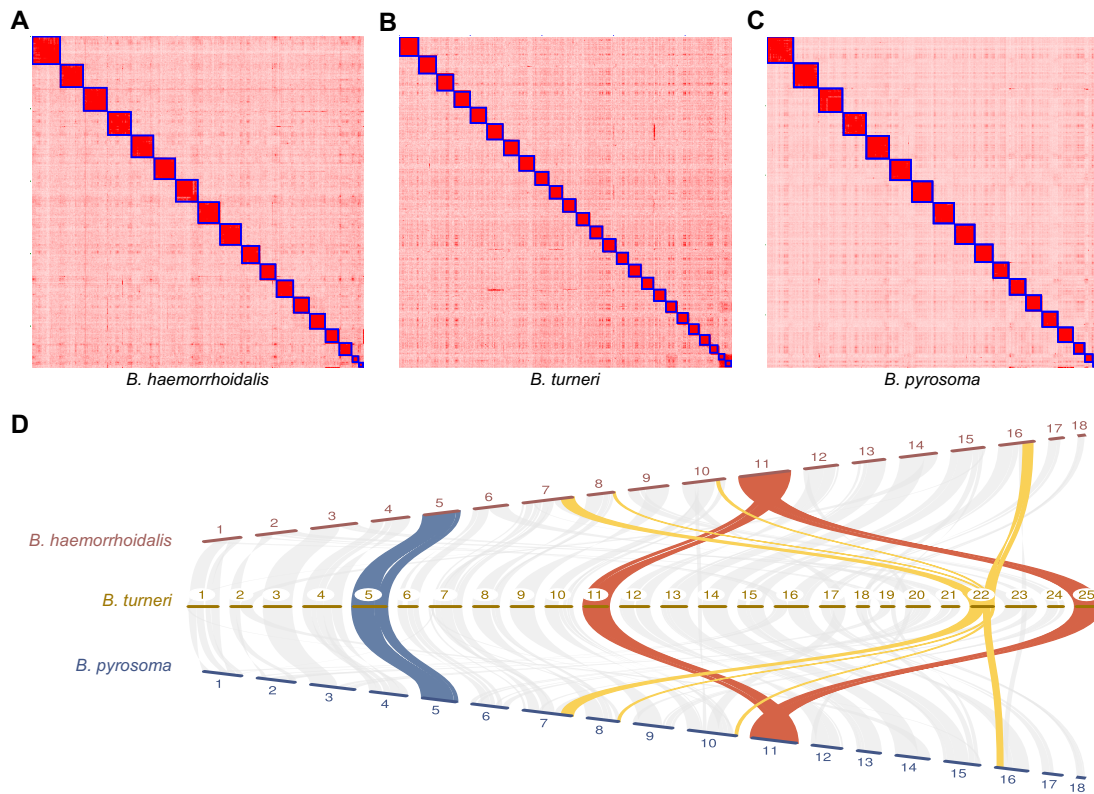


Figure 2. Chromosome number evolution in representative bumblebee species from three different subgenera. Hi-C contact heatmaps for *B. haemorrhoidalis* (A), *B. turneri* (B), and *B. pyrosoma* (C) show that the three species have 18, 25, and 18 chromosomes, respectively. The 18-chromosome karyotype is the inferred ancestral genome structure, with 25 chromosomes found in social parasite bumblebees of the subgenus *Psithyrus*. (D) Macrosynteny comparisons across *B. haemorrhoidalis*, *B. turneri* and *B. pyrosoma* shows how the 25 *B. turneri* chromosomes result from a combination of fission (red), fusion (yellow), and retention (blue) of ancestral chromosomes.

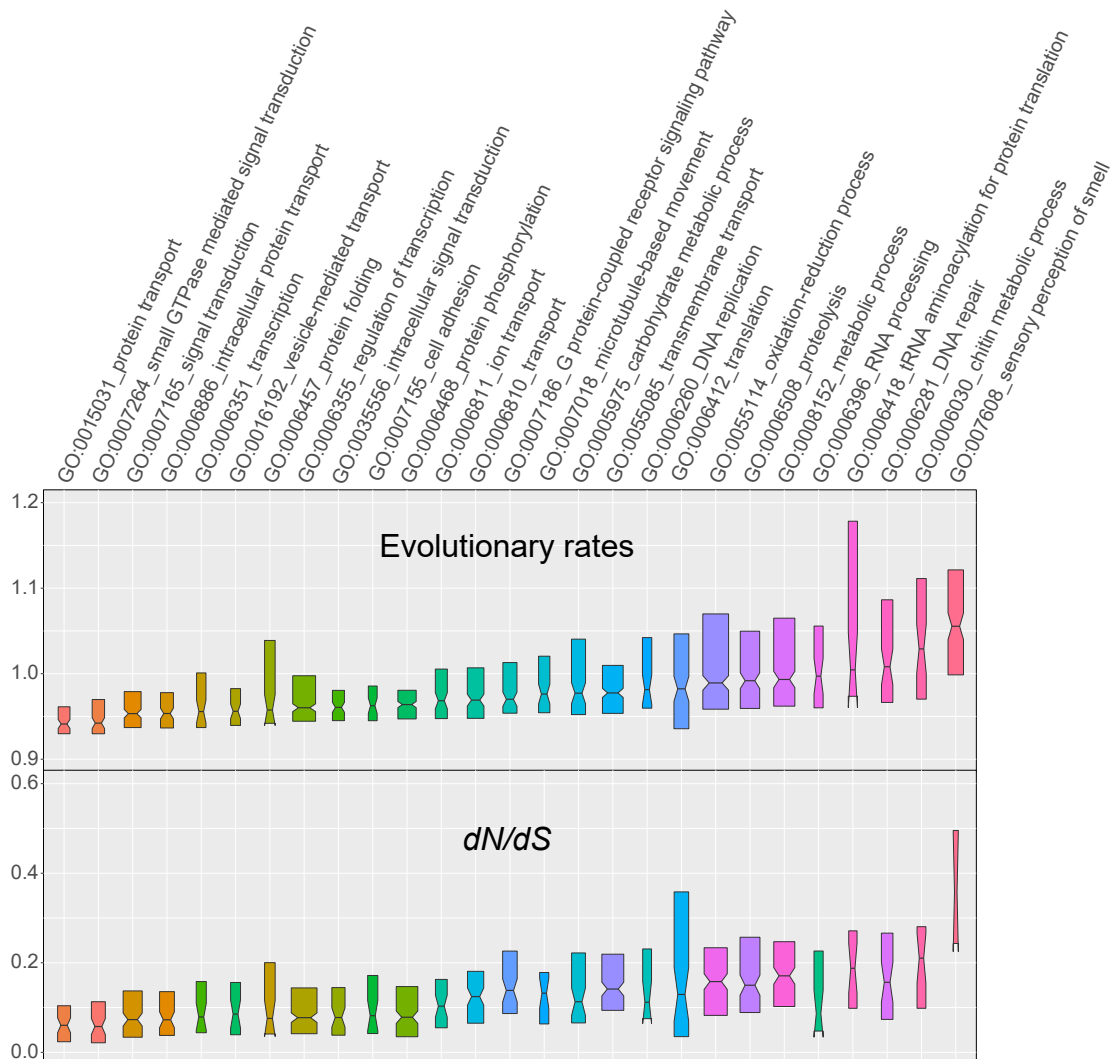


Figure 3. Molecular evolution of protein-coding genes in terms of evolutionary rate and dN/dS ratio. Gene Ontology (GO) Biological Process terms are sorted by evolutionary rate from the most conservative (left) to the most dynamic (right) and colored from the highest values (red) to the median value (blue) to the lowest values (orange). Evolutionary rate refers to amino acid sequence divergence measured as the mean of normalized inter-species ortholog protein sequence identities. dN/dS refers to the ratio of the number of nonsynonymous substitutions per non-synonymous site to the number of synonymous substitutions per synonymous site. Notched boxes show medians of orthologous group values with the limits of the upper and lower quartiles, and box widths are proportional to the number of orthologous groups in each category.

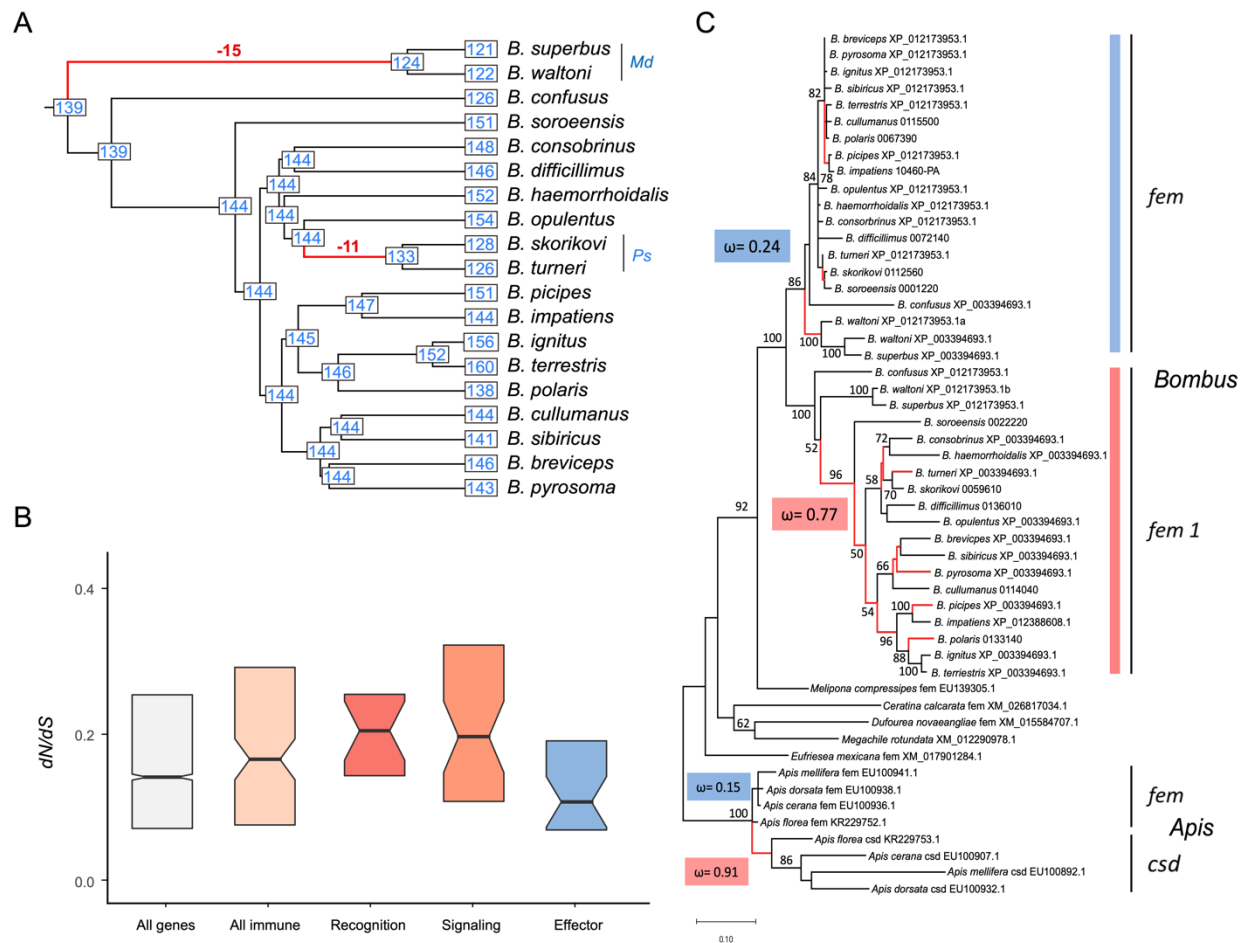


Figure 4. Evolution of genes associated with ecology and reproduction. **(A).** Observed gene counts and inferred ancestral gene counts of bumblebee odorant receptors (ORs) on an ultrametric phylogeny, highlighting two major gene loss events (the complete result is available in Supplementary figures Figure S25). **(B).** Boxplots showing dN/dS ratios for different categories of immune genes and all single-copy genes in bumblebee (All genes). Elevated dN/dS ratios among immune-related genes is driven by higher ratios for genes involved in recognition and signaling processes. Notched boxes show medians of orthologous group values with the limits of the upper and lower quartiles. **(C).** The evolutionary history of *fem* genes of bees including their paralogs *fem1* in bumblebees (*Bombus*) and *csd* in honeybees (*Apis*). Global non-synonymous to synonymous rate ratio (ω) were calculated for *fem*_{Bombus} (reference, blue) and *fem1*_{Bombus} (test, red), including a branch-site testing framework with model fitting and Likelihood Ratio Tests, showing evidence for relaxation of selection in *fem1*_{Bombus} ($P < 0.001$, LR = 36.34). Spurious actions of diversifying selection on branches predominantly found in *fem1*_{Bombus} are marked in red. For comparison, ω for *fem* and *csd* in *Apis* is given, known as striking example of neo-functionalization.