Apidologie (2025) 56:10
© The Author(s), 2025
https://doi.org/10.1007/s13592-024-01140-1

Original article

De novo genome assembly and annotations of *Bombus lapidarius* and *Bombus niveatus* provide insights into the environmental adaptability

Vahap Eldem[1], Yusuf Ulaş Çınar[1], Selahattin Barış Çay[1], Onur Obut[1], Selim Can Kuralay[1], Mehmet Ali Balcı[1], Pınar Akbaba[1,7], Tuana Öğretici[1], Tunç Dabak[2], Burcu Daşer Özgışı[3], Çiğdem Özenirler[3], Gökmen Zararsız[4,5], Esma Gamze Aksel[6], and Fatih Dikmen[1]

[1] Department of Biology, Faculty of Science, Istanbul University, 34134 Istanbul, Turkey
[2] Department of Biology, The Pennsylvania State University State College, Pennsylvania, PA, USA
[3] Department of Biology, Faculty of Science, Hacettepe University, 06800 Ankara, Turkey
[4] Department of Biostatistics, Erciyes University, 38039 Kayseri, Turkey
[5] Drug Application and Research Center (ERFARMA), Erciyes University, 38280 Kayseri, Turkey
[6] Department of Genetics, Faculty of Veterinary Medicine, Erciyes University, 38280 Kayseri, Turkey
[7] Department of Molecular Biology and Genetics, Koc University, 34450 Istanbul, Turkey

Received 10 May 2024 – Revised 27 September 2024 – Accepted 10 December 2024

Abstract – Bumblebees are ubiquitous, cold-adapted, primitively eusocial bees and important pollinators for crops and vegetation. However, many species are declining worldwide due to multiple factors, including human-induced habitat loss, agricultural chemicals, global warming, and climate change. In particular, future climate scenarios predict a shift in the spatial distribution of bumblebees under global warming, with some species declining and others potentially expanding. Here, we report a de novo genome assembly and annotation for *Bombus lapidarius* and *Bombus niveatus* to decipher species-specific potential genomic capacity against such environmental stressors. With harboring more than 23,000 protein-coding genes, the assembled genomes of *B. lapidarius* and *B. niveatus* are 244.44 Mb (scaffold N50 of 9.45 Mb) and 259.84 Mb (scaffold N50 of 10.94 Mb), respectively, which exhibit similar trends in terms of genome size and composition with other bumblebees. Gene family analysis reveals differences in species-specific expanded gene families. *B. lapidarius* exhibits expanded genes related to pre/postsynaptic organization, while *B. niveatus* shows a distinct expansion in gene families regulating cellular growth, aging, and responses to abiotic and biotic stressors, such as those containing SCAN domains, WD-repeats, and Ras-related proteins. Our genome-wide screens revealed positive selection on environmental stress-responsive genes such as *dip2*, *yme1l*, and *spg7* in *B. lapidarius*, whereas positive selection signatures were found in genes such as *myd88*, *mybbp1A*, and *rhau*, which are involved in environmental stress resistance for *B. niveatus*. These high-quality genome assemblies and comparative genome analysis unveil potential drivers that underlie genome evolution in bumblebees, offering valuable insights into environmental adaptation and conservation efforts.

Bumblebees / Genome assembly / Adaptation / Positive selection signature / Gene family analysis

1. INTRODUCTION

More than 300,000 flowering plant species exist worldwide (Christenhusz and Byng 2016), and each species requires wind, water,

Corresponding author: F. Dikmen, fatih.dikmen@istanbul.edu.tr
Manuscript editor: Marina Meixner

and insects for pollination. Bees (Apoidea: Hymenoptera) are the leading insect pollinators and are essential for healthy ecosystems, playing a vital role in plant reproduction and the survival of countless plant species. There are approximately 17,500 bee species worldwide (Michener 2007). Bumblebees (*Bombus* spp.) are notable pollinators among these bee species, and this genus contains approximately 250 identified species globally (Cameron et al. 2007). They are easily recognized by their large body and striking color patterns, and thanks to their specialized physical and chemical thermoregulatory mechanisms, they manage to be active even at cold or inclement climatic conditions. Such features allow them to live in various habitats such as alpine, subalpine, and arctic belts of the Palearctic, Nearctic, Oriental, and Neotropic regions, from Tierra del Fuego to the Himalayas, which is above 5000 m (Heinrich 1979; Cameron and Williams 2003; Williams et al. 2009). In addition, their social life is also an important feature for colony persistence and reproductive success (Maebe et al. 2021).

However, it seems that even a group with such adaptive features and an essential part of the ecosystem has suffered from global biodiversity loss for decades. Factors such as diseases, parasites, agrochemicals, pollution, habitat loss, invasive species, and climate changes have caused population declines within bumblebees recently (Cameron and Sadd 2020; Rasmont et al. 2021), while there are only a few species that remain stable or are currently expanding their distribution range (Biella et al. 2021; Ghisbain et al. 2021). According to the European Red List of bees, 23.6% of *Bombus* species are threatened with extinction, and 4.4% are considered threatened (NT status). In addition, 45.6% of the species in general had a decreasing population trend (Nieto et al. 2014). This is worrisome because decreases in pollinators also induce a decrease in the reproduction of flowering plants (Biesmeijer et al. 2006; Pauw 2007; Potts et al. 2010; Carvalheiro et al. 2013), wherein 60–70% of the natural vegetation depends on animal pollinators—primarily insects—and are directly affected by the decrease in pollinator diversity and abundance (Potts et al. 2016).

The ability of natural populations to adapt to environmental changes depends on the extent and type of disturbances that the organisms encounter. Populations with high adaptive potential are expected to cope with changes in habitat suitability resulting from climate change. Understanding the genetic basis of climatic adaptation is crucial for predicting how species will respond to climate change, as numerous insect populations will be compelled to shift or adapt (Halsch et al. 2021; McCulloch and Waters 2022). With advances in genome sequencing technologies, increasing attention has been placed on the molecular basis of how insect species will adapt to changing environmental conditions and future climatic scenarios (You et al. 2020; Chen et al. 2021; Tong et al. 2022). *Bombus* species are no exception, and, as observed in other insects, it has been reported that they will need to adapt to changing environments in response to projected climate scenarios and trends (Rasmont et al. 2015; Ghisbain et al. 2024). Besides, recent ecological models on bumblebees indicated spatial variability in their response to climate change; patterns are heterogeneous, particularly among species (Kerr et al. 2015). Based on climate projection models suggested by Rasmont et al. (2015), *B. niveatus* Kriechbaumer 1870 is among the species poised to benefit from climate change, potentially expanding its current distribution in Europe. On the other hand, *B. lapidarius* (L. 1758), a generally abundant and ubiquitously distributed species with high dispersal capability, would experience a significant reduction in the area of its suitable habitats. *B. lapidarius* is also regarded as a wood-edge species (Rasmont et al. 2021) and displays an adaptation to urbanization, which means it has a high level of plasticity (Theodorou et al. 2018). However, *B. niveatus* is an open-space species and very common in all the steppe types (Rasmont et al. 2021). By comparing and mining the genomes, we might uncover key ecological traits and evolutionary

mechanisms of these essential pollinators to understand how they are adapting to changing environments.

Here, we generated the de novo genome assemblies and profiled comprehensive organ-specific transcriptomes of *B. lapidarius* and *B. niveatus*. The combination of genome and transcriptome analysis revealed a set of species-specific expanded gene families and potential genomic signatures in coding genes associated with adaptations to long-term environmental changes. Moreover, their chemosensory, immunity, and detoxification gene repertoires, crucial for environmental adaptation, were compared with those of other bumblebees. Species-specific expansions within gene families and genes exhibiting signatures of positive selection are particularly valuable for elucidating the molecular mechanisms underlying environmental adaptation and stress response in bumblebees.

2. MATERIALS AND METHODS

2.1. Sampling

Material for the study was obtained from field studies performed between 2021 and 2022. Eight female specimens each of *B. lapidarius* and *B. niveatus* were collected from distinct regions. The *B. lapidarius* specimens were collected from Eskişehir Seyitgazi region (39° 10′ 53.49″N, 30° 33′ 20.64″E, 1138 m), and *B. niveatus* specimens were collected from Ankara Beytepe region (39° 51′ 44″N, 32° 43′ 59″E, 1065 m). Bumblebees were sampled from flowers they visited by hand net and kept alive in separate falcon tubes (50 ml) until they arrived at the laboratory the same day. As soon as their identification was performed, the target species were kept at a − 80 °C freezer prior to DNA extraction.

2.2. Extraction of genomic DNA and sequencing

High molecular weight DNA was isolated from the thorax and head regions of one specimen using the NucleoBond HMW DNA kit (MACHEREY–NAGEL, Düren, Germany), following the manufacturer's guidelines. Briefly, the thorax and head tissues were homogenized using an IKA RW 20 (Staufen, Germany) stirrer homogenizer, while the abdomen tissues were excluded due to potential microbial contamination. Following homogenization, 200 µl of Proteinase K was added and incubated for an hour, followed by the addition of 100 µl RNAse A and a 30-min incubation to prevent RNA contamination. Then, the homogenate was placed onto a NucleoBond column filter, where DNA was collected in the silica gel alongside the buffers supplied in the kit. The DNA was subsequently washed with 70% ethanol and isopropanol, respectively, and eluted with a Tris–EDTA elution buffer, quantified using a Qubit fluorometer with a dsDNA HS assay kit. Finally, the fragmentation of the DNA was visualized via 2% agarose gel electrophoresis. The 3 µg of the unfragmented HMW DNA was used as the starting material for the Oxford Nanopore Technology (ONT) library preparation. DNA libraries were constructed using the Ligation Sequencing Kit (SQK-LSK109) in accordance with the manufacturer's guidelines (Oxford Nanopore Technologies, Cambridge, UK). DNA repair and end-prep were performed using the NEBNext® FFPE DNA Repair mix and NEBNext® Ultra™ II End-prep module, respectively. The mixture of 1 µl DNA CS, 3.5 µl FFPE Repair Buffer, 2 µl FFPE DNA Repair Mix, 3.5 µl Ultra II End-prep reaction buffer, and 3 µl Ultra II End-prep enzyme mix was added to 47 µl of the DNA sample. The final solution was incubated at 20 °C for 5 min and then at 65 °C for 5 min on a thermal cycler. The AMPure XP beads were resuspended via vortexing, and DNA samples were transferred into a 1.5-ml DNA LoBind tube and mixed with 60 µl of these beads and incubated for 5 min before being washed twice with 70% ethanol; this was then resuspended in 61 µl nuclease-free water and quantified with a Qubit v4 fluorometer (Thermo Scientific). Next, the DNA sample was mixed with 25 µl of Ligation buffer, 10 µl of NEBNext Quick T4 DNA Ligase, and 5 µl of Adapter Mix before adding

40 µl of resuspended AMPure beads for another 5 min incubation. The beads were then washed with 250 µl of long fragment buffer (LFB) and resuspended with 15 µl of elution buffer (EB) at 37 °C for 10 min. Finally, 37.5 µl of sequencing buffer (SQB), 25.5 µl of loading beads (LB), and 14 µl of DNA library were mixed into a LoBind tube, and 75 µl of the sample was introduced into the flow cell via the SpotON sample port. The ONT libraries were sequenced on the R9.4.1 flow cell using the MinION Mk1C sequencer and the ONT MinKNOW software (Oxford Nanopore Technologies, Cambridge, UK).

2.3. Transcriptome sequencing

Total RNA was extracted from four tissue types: antenna, head, leg, and thorax, using TRIzol reagent (Invitrogen, Carlsbad, CA) following the manufacturer's procedure. For each species, RNA samples from three biological replicates were pooled for each tissue before RNA-Seq analysis. Briefly, 1 ml of TRIzol reagent was added to each tube. Next, tissues were homogenized using MagNA Lyser (Roche Diagnostics GmbH, Germany) with zirconium beads. Subsequently, 200 µl chloroform was added into tubes and mixed by vortexing, and samples were centrifuged at 12,000 g for 15 min at 4 °C. Upper phase was transferred into new tubes and mixed with 1.5 volume of isopropanol and was centrifuged at 12,000 g for 15 min at 4 °C. Then, pellets were re-washed with 75% ethanol and centrifuged at 12,000 g for 15 min at 4 °C. Finally, the RNA was dissolved in 30 µl RNAse-free water. Total RNA was quantified using NanoDrop ND2000c (Thermo Scientific, USA). Total RNA samples with an OD260/280 value equal to or slightly higher than 2.0 and RIN (RNA integrity number) above 7.0 were selected for cDNA library construction. The rRNAs were depleted from the total RNA samples by using an Illumina Ribo-Zero rRNA removal kit (San Diego, CA, USA) in accordance with the Illumina manufacturer's guidelines. Immediately after the rRNA removal, cDNA libraries were generated using the TruSeq™ RNA sample preparation kit (San Diego, CA, USA) following the manufacturer's instructions. The obtained libraries were sequenced on the NovaSeq 6000 platform (Illumina) with paired-end 150 bp reads.

2.4. Genome assembly

The Guppy v3.2.4 software (pre-installed in the MinION Mk1C, -c dna_r9.4.1_450bps_hac.cfg –num_callers 8 –gpu_runners_per_device 4 –chunks_per_runner 512 –device auto) was utilized to achieve high-quality base calling from the raw ONT data. Additionally, Porechop v0.2.4 (https://github.com/rrwick/Porechop) was implemented with default parameters to trim the ONT adapter sequences. De novo genome assembly was conducted separately for *B. lapidarius*, and *B. niveatus* reads using Flye v2.9.1 (Kolmogorov et al. 2019, –nano-hq), and draft assembly was polished with Racon v1.4.3 (https://github.com/isovic/racon), in four iterations. The polished Flye assemblies of two *Bombus* species were screened and cleaned of foreign contamination using NCBI's FCS-GX tool (Astashyn et al. 2024). For *B. niveatus*, the resulting polished assembly was further scaffolded and gap-filled using Samba script (from MaSuRCA v4.1.0) (Zimin and Salzberg 2022). Due to the availability of PacBio HiFi reads for the *B. lapidarius* in the NCBI SRA database (BioProject: PRJEB66044 and PRJEB51891), the polished *B. lapidarius* genome was scaffolded and gap-filled with PacBio HiFi reads similarly using the Samba script. The completeness of polished and scaffolded genomic assemblies was assessed with BUSCO v5.0 (Benchmarking Universal Single-Copy Orthologs, Manni, et al. (2021)) that was run on with the gVolante v1.2.1 online tool (Nishimura et al. 2017), using the Arthropoda database with default settings.

2.5. Genome characterization of repeats

To identify the repetitive elements in the de novo assembled genomes of *B. lapidarius* and *B. niveatus*, we utilized the RepeatModeler v2.0.4

(Flynn et al. 2020) and RepeatMasker v4.0.7 (Smit et al. 2017) softwares. Specifically, we employed the RepeatModeler software pipeline, which utilizes de novo repeat finder tools such as RepeatScout v1.0.6 (Price et al. 2005), RECON v1.08 (Bao and Eddy 2002), TRF v4.09.1 for tandem repeat finder (Benson 1999), and LTR_FINDER v2.9.0 (Xu and Wang 2007) for long terminal repeat elements, to construct a de novo repeat library. We then combined this library with the homolog-based repeat databases RepBase library (v20181026) (Bao et al. 2015) and Dfam release 3.6 (Storer et al. 2021) to generate a custom library. Finally, we used RepeatMasker v4.1.4 with RMBlast v2.13.0 as a search engine to mask the repeat regions with much greater sensitivity, using the custom library and the "*-s*" option.

2.6. Protein-coding gene annotation

The BRAKER pipeline v2.1.6 (Hoff et al. 2019) was utilized to predict protein-coding genes by integrating evidence data from homologous protein alignment, ab initio prediction, and RNA-Seq mapping. For homology-based gene prediction, the protein sequences of *Bombus terrestris* (GCA_910591885.2), *Bombus affinis* (GCA_024516045.2), *Bombus bifarius* (GCA_011952205.1), *Bombus huntii* (GCA_024542735.1), *Bombus impatiens* (GCA_000188095.4), *Bombus pyrosoma* (GCA_014825855.1), *Bombus vosnesenskii* (GCA_011952255.1), and *Bombus vancouverensis nearcticus* (GCA_011952275.1) were downloaded from related public databases as homologous species. The encoded protein sequences of these species were aligned with the genomic sequence of *B. lapidarius* and *B. niveatus* using "tblastn" v2.12 (https://blast.ncbi.nlm.nih.gov/Blast.cgi) an *e*-value cutoff of $1e-5$, and GeneWise (Birney et al. 2004) was then used to identify the gene structure. For RNA-Seq-based annotation, clean RNA-Seq reads were aligned to soft-masked genomes using Hisat2 v2.2.1 (Kim et al. 2019) to determine exons region and splice positions (stored in ".bam" alignment format). The alignment results were then input for StringTie v2.1.7 (Pertea et al. 2016) to assemble transcripts to the gene models. The BRAKER2 ab initio gene predictions were performed using both homologous protein and RNA-Seq evidence, utilizing the Augustus v3.5.0 (Hoff et al. 2016) and GeneMark-ET v4.71 (Brůna et al. 2020) software programs. Then, the completeness of the genome annotations was evaluated using the BUSCO v.5.0 "Arthropoda" database embedded in the gVolante v1.2.1 online tool (Nishimura et al. 2017). All predicted protein-coding genes were subjected to a search against various functional genome databases, including NCBI non-redundant hexapoda protein database ("nr"), UniRef90 (Suzek et al. 2007), Kyoto Encyclopedia of Genes and Genomes (KEGG) databases (Kanehisa et al. 2014), InterPro database (v93.0) (Paysan-Lafosse et al. 2023), and eggNOG-mapper v2.1.9 (Cantalapiedra et al. 2021) to annotate proteins by Gene Ontology (GO) terms.

2.7. Non-coding RNA annotation

Non-coding RNAs (ncRNAs) are a distinct class of RNA species that do not undergo peptide translation. In the genome of *B. lapidarius* and *B. niveatus*, four types of ncRNAs were identified, namely transfer RNAs (tRNAs), ribosomal RNA (rRNAs), microRNAs (miRNAs), and small nuclear RNAs (snRNAs). tRNAs were predicted with tRNAscan-SE v2.0.12 software with default parameters (Chan and Lowe 2019). The rRNA genes belonging to the Apideae family were retrieved from the NCBI nucleotide database and then searched against the de novo assembled genomes with "blastn" (an *e*-value cutoff of $1e-5$) to locate all the copies. Additional ncRNAs, such as miRNAs and snRNAs, were detected through using Infernal software v1.1.4 (Nawrocki and Eddy 2013) which performed a search against the Rfam v14.9 database (Kalvari et al. 2021).

2.8. Gene family identification, expansion, and contraction analysis

The proteomes of ten insect species were obtained from the NCBI Genome Dataset (https://www.ncbi.nlm.nih.gov/datasets/). The species included were *Aedes aegypti* (GCF_002204515), *Apis mellifera* (GCF_003254395), *Bombyx mori* (GCF_000151625), *Drosophila melanogaster* (GCF_000001215), *Manduca sexta* (GCF_014839805), *Nasonia vitripennis* (GCF_009193385), *Papilio machaon* (GCF_912999745), *Tribolium castaneum* (GCF_000002335), and *Acyrthosiphon pisum* (GCF_000142985, as an outgroup). To ensure consistency across species, the proteomes were filtered using AGAT v1.0.0 (https://github.com/NBISweden/AGAT, *agat_sp_keep_longest_isoform.pl*), retaining only the longest transcript for each gene. Gene family clusters and single-copy ortholog sequences within the proteome set were identified using OrthoFinder2 v2.5.4 with the default parameter (Emms and Kelly 2019). Orthologous gene alignments generated with OrthoFinder's "-m MSA" option were concatenated using SeqKit v2.3.1 for phylogenetic construction (Shen et al. 2016). Aligned orthologs were imported into IQ-TREE v2.1.1 to determine the best-fit model of evolution (Q.insect + R5) with 1000 bootstraps (Minh et al. 2020). To estimate changes in the gene repertoire of the two bumble bee species, we conducted an analysis following the methodologies proposed by Yang et al. (2021) and Ouyang et al. (2022). We conducted an expansion and contraction analysis using all orthologous families identified by OrthoFinder from the ten species mentioned earlier. To infer the significance of changes in gene family size in each branch, we utilized a random birth–death model of CAFE v4.2.1 (De Bie et al. 2006). The phylogenetic tree topology and branch lengths were considered in the analysis. Significance was determined by calculating *p*-values for each lineage, with values < 0.05 deemed significant. We also performed a genome-level synteny analysis and visualization of *B. lapidarius* and *B. niveatus* species compared to *B. terrestris* and *B. impatiens* using NGenomeSyn v1.41 software (He et al. 2023).

2.9. Genome-wide positive selection analysis

A genome-scale positive selection analysis was conducted on *Bombus* genomes using the paPAML tool (Steffen et al. 2022), which employs codeml from the PAML version 4 (Yang 2007). This tool was chosen due to its efficiency and ability to produce results in low computing time using multithreading and faster summarization of the results for each cluster by automatically calculating the likelihood ratio test (LRT) and Bayes empirical Bayes (BEB) (Yang et al. 2005) inference results for each site, making it more suitable for extensive datasets like in genome-scale analysis. The orthologous clusters are also obtained from OrthoVenn3 (Sun et al. 2023) for the positive selection analysis, using the same species in the gene expansion analysis, which are *A. pisum*, *A. aegypti*, *A. mellifera*, *B. mori*, *D. melanogaster*, *N. vitripennis*, and *T. castaneum*. Prior to the analysis, stop codons from the sequences of each orthologous cluster were removed. The clusters were subjected to codon-level alignment using the PRANK software (Loytynoja 2021) and subsequently curated using Gblocks (Castresana 2000) with the "$-t = c$" parameter to ensure codon-specific curation and removal of gaps. Afterwards, phylogenetic trees required to test branch-site models are constructed for each cluster using IQ-TREE 2 (Minh et al. 2020) with best-fitting models, and genes of the *Bombus* species are selected as foreground branches within the Newick files. Finally, control files required to run the paPAML v1.2 tool (Steffen et al. 2022) according to the branch-site model ($\omega = 1/\omega > 1$) were prepared and run with the test 2 "*–t 2*" option, which indices the branch-site model. GO enrichment analysis for positively selected genes (PSGs) was performed using ShinyGO online tool v0.741 (http://bioinformatics.sdstate.edu/go/) with an FDR cutoff of 0.05 for significantly enriched GO terms (Ge

et al. 2020). General protein–protein interactions (PPI) of PSGs were collected from the STRING online tool v12.0 (https://string-db.org/) (Szklarczyk et al. 2015). To ensure the quality of the incorporated PPIs, we selected a high confidence score of > 0.7 because a lower score leads to more interaction and gives more false positives.

2.10. Identification of genes involved in chemosensory, immunity, and detoxification

Two bumblebee genome sequences were manually analyzed to annotate detoxification, immune, and chemosensory genes. The OrthoVenn2 tool was utilized (Xu et al. 2019) to define orthologous groups from *B. lapidarius* and *B. niveatus* as well as from *B. affinis*, *B. bifarius*, *B. huntii*, *B. impatiens*, *B. terrestris*, *B. vancouverensis nearcticus*, and *B. vosnesenskii*. The official gene sets (NCBI RefSeqs) were utilized to conduct a blastp (an *e*-value cutoff of 1e − 5) search to identify homologs of known proteins from the two bumblebees. Furthermore, for the analysis of potential detoxification genes, Pfam was utilized to identify the deduced protein domains PF00067 (P450s), PF00135 (CCEs), PF00043, and PF02798 (GSTs) by conducting searches using HMMER v3.3.1 (Potter et al. 2018). A local library was constructed using previously identified chemosensory proteins in corbiculate bees (Forêt and Maleszka, 2006; Robertson and Wanner 2006; Brand et al. 2015; Sadd et al. 2015), and homology searches were performed against this library. Additionally, candidate chemosensory genes underwent a search for Pfam domains including PF02949 (ORs), PF08395 (GRs), PF00060 (IRs), PF10613 (IRs), PF01395 (OBPs), PF03392 (CSPs), and PF01130 (SNMPs) using HMMER v3.3.1. The OR subfamilies were determined based on previous publications (Forêt and Maleszka 2006; Zhou et al. 2012; Karpe et al. 2016; Obiero et al. 2021). The N-terminal signal peptide sequences of odorant binding proteins (OBPs) and chemosensory proteins (CSPs) were predicted using SignalP 6.0 (Teufel et al. 2022) with default parameters. TMHMM 2.0 (https://services.healthtech.dtu.dk/services/TMHMM-2.0/) was employed to predict transmembrane domains (TMDs) of odorant and gustatory receptors.

3. RESULTS

3.1. Genome assembly and annotations of *B.* lapidarius and *B.* niveatus

We have sequenced and assembled the *B. lapidarius* and *B. niveatus* genomes to decipher the genomic architecture of essential pollinators in natural ecosystems using long-read sequencing technology. We followed the sequencing recommendation proposed by Faulk (2023) to de novo assemble small-size diploid genomes (< 500 megabases, Mb) using long-read nanopore sequencing. After quality control and read length filtering, by using the MinION Mk1C sequencer, a total of 339,614 (~4.1 gigabases, Gb, 18X coverage, N50 of 22.55 Kb) and 579,883 (~10.1 Gb, 41X coverage, N50 of 29.04 Kb), clean reads were generated from R9.4.1 flow cells for *B. lapidarius* and *B. niveatus*, respectively (Table S1; Fig. S1A-B). After the assembling process, the final assembled genome sizes of *B. lapidarius* and *B. niveatus* were determined as 244.44 Mb and 259.84 Mb, with corresponding scaffold N50 values of 9.45 Mb and 10.94 Mb, respectively. The BUSCO completeness analysis showed that the genome completeness in Arthropoda ($n = 1013$) lineage datasets was 99.11% (1004) and 99.51% (1008) for *B. lapidarius* and *B. niveatus*. The genome GC content of both bumblebee species was approximately 37.6% (Table I). In the *B. niveatus* genome, 18 chromosome-level scaffolds were assembled, with a length range of approximately 3.7 to 20.1 Mb, which accounts for 87.4% of the complete genome. Moreover, 44 smaller scaffolds were assembled, which take up about 7.2% of the genome. The other sequences remained as contigs. Similarly, the *B. lapidarius* genome was assembled into 18 chromosome-level scaffolds, with a length range of around 3.7 to 18.3 Mb, accounting for approximately

Table I Summary of the *B. lapidarius* and *B. niveatus* genome assembly

Stats	*B. lapidarius*	*B. niveatus*
Assembly size (Mb)	244.48	259.84
Pseudochromosome number	18	18
Scaffolds (n)	228	356
Scaffold N50 (Mb)	9.45	10.94
Scaffold L50	10	10
Mean sequence length (nt)	1,072,302	729,899
Median sequence length (nt)	29,101	42,815
Largest scaffold length (nt)	14,753,331	14,878,647
Smallest scaffold length (nt)	505	213
# sequences > 1 M	45	36
# sequences > 10 M	8	11
Σlength of seq > 1 M	229,682,509 (93.9% of total length)	232,860,909 (89.6% of total length)
Σlength of seq > 10 M	105,749,925 (43.3% of total length)	143,417,695 (55.2% of total length)
GC content (%)	37.75	37.64
BUSCO Arthropoda		
Complete	1004 (99.11%)	1008 (99.51%)
Complete + partial	1012 (99.90%)	1013 (100.00%)
Missing	1 (0.09%)	0 (0.00%)
Complete + duplicated	5 (0.49%)	4 (0.39%)
BUSCO Hymenoptera		
Complete	5885 (98.23%)	5890 (98.31%)
Complete + partial	5931 (98.99%)	5933 (99.03%)
Missing	60 (1.00%)	58 (0.96%)
Complete + duplicated	23 (0.38%)	22 (0.36%)

91% of the complete genome. Additionally, 27 smaller scaffolds, which comprise about 6.9% of the genome, were assembled (Figure 1). Collinearity analysis revealed remarkably conserved syntenic relationships among *B. terrestris*, *B. niveatus*, and *B. lapidarius* chromosomes, with each chromosome pair showing close matches (Fig. S2). This finding also supports a high level of synteny conservation in *Bombus* species. To improve the accuracy of gene annotation and reveal a more complete transcriptome profile, RNA-Seq using the Illumina Total RNA-Seq kit was performed on the antenna, head, thorax, and leg tissues from bumblebee specimens. Approximately 56.2 Gb (416.4 million sequences, 150 bp read length) and 53.3 Gb (401.3 million sequences, 150 bp read length) of clean data were retained for *B. lapidarius* and *B. niveatus*, respectively (Table S3). After removing these low-quality genes, a total of 23,266 and 23,792 protein-coding genes were finally obtained for *B. lapidarius* and *B. niveatus*, respectively. We have compared the coding gene numbers and characters of these *Bombus* species with well-annotated *B. terrestris* (assembly iyBomTerr1) and *B. impatiens* (assembly BIMP_2) to validate the quality of our annotation. The average CDS length was 1098 bp and 1145 bp for *B. lapidarius* and *B. niveatus*, respectively. In addition, non-coding RNAs (ncRNAs) were identified across the genome. A total of 115 ribosomal RNAs (rRNAs) were identified for *B. lapidarius*, and 228 rRNAs were identified for *B. niveatus* using the Rfam databases. Similarly, 220 transfer

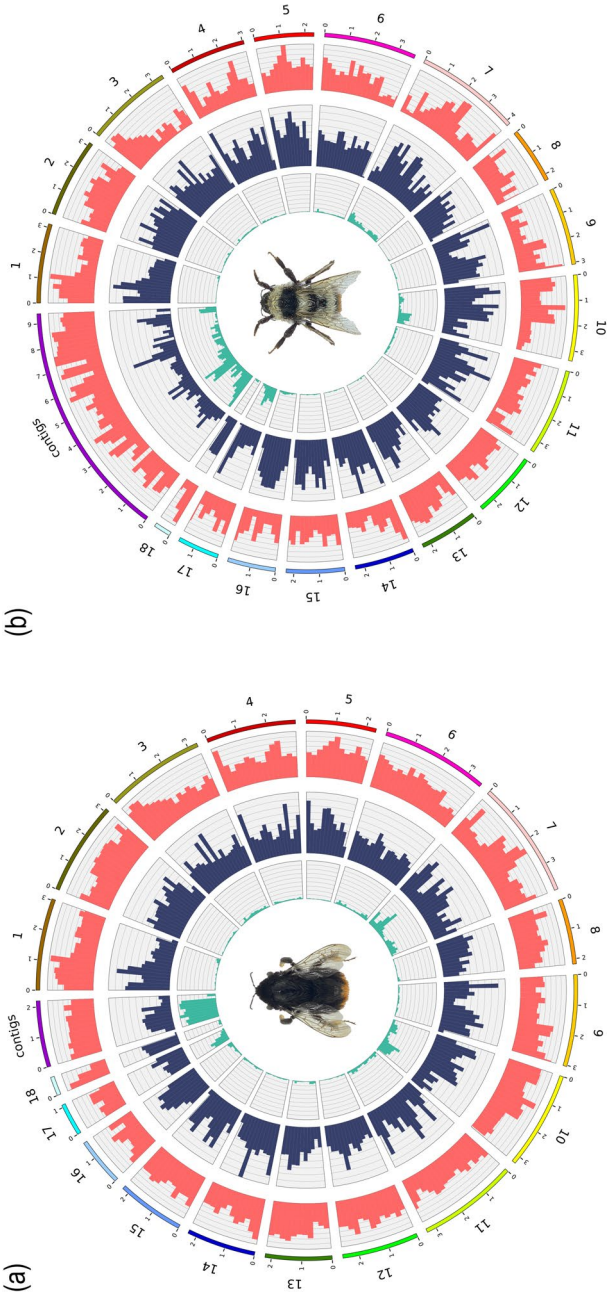


Figure 1. The circular diagram depicts the characteristics of the **a***B. lapidarius* and **b***B. niveatus* genomes. The diagram includes the eighteen pseudochromosomes and one merged scaffold, which consists of the shorter scaffolds and contigs. The innermost chart displays transposable element density, the middle chart shows gene density, and the outermost chart illustrates GC content distribution. The charts are plotted over 1 Mb base pair windows. The photos of the bee specimens were obtained from field samples.

RNAs (tRNAs) were determined for *B. lapidarius*, and 277 tRNAs were determined for *B. niveatus* using the tRNAscan-SE database (Table II).

In the genome of *B. lapidarius*, transposable elements (TEs) occupy approximately 11.05% of its total size, with long terminal repeats (LTRs) being the most prevalent class, accounting for 31% of all TEs. DNA transposons are the second most common TE class, accounting for 25% of the total TEs, followed by non-LTRs and Helitrons, representing 4% and 1%, respectively. About 39% of TEs in this genome were not classified. In the genome of *B. niveatus*, TEs occupy about 15.04% of its total size. Among all TE classes, LTRs are the most common, making up around 29% of the total TEs. LTRs, non-LTRs, and Helitrons represent 28%, 4%, and 2% of the total TEs, respectively. However, around 37% of TEs remain unclassified. Within the LTR class, gypsy elements are the most abundant in both genomes, representing 77% in *B. lapidarius* and 73% in *B. niveatus*. BEL elements accounted for 9% of *B. lapidarius* and 11% of *B. niveatus*, while 12% and 10% of LTRs remain unclassified in each genome. Copia elements comprised 6% of *B. lapidarius* and 9% of the *B. niveatus* LTRs. Regarding non-LTR elements, long interspersed nuclear elements (LINEs) dominate both genomes, representing about 50% in *B. niveatus* and around 95% in *B. lapidarius*. However, the percentage of PLEs is substantially higher in the *B. niveatus* genome, constituting about 29% of the total non-LTR elements, compared to only 4% in *B. lapidarius*. The remaining non-LTR elements could not be identified for further classification. DNA transposons mainly comprise TcMar elements, representing around 53% in *B. niveatus* and 56% in *B. lapidarius*. hAT is the second most prevalent DNA transposon in both species, accounting for 19% in *B. niveatus* and *B. lapidarius*. Mutator, Harbinger, and CACTA elements are relatively less frequent, constituting about 16%, 2%, and 10% in *B. niveatus* and 5%, 9%, and 11% in *B. lapidarius*, respectively (Table S4, Fig. S3).

3.2. Gene family evolution

The proteome set of *B. lapidarius* and *B. niveatus* was compared using CAFÉ (Mendes et al. 2020) with seven representative species, including *A. pisum*, *A. aegypti*, *A. mellifera*, *B. mori*, *D. melanogaster*, *N. vitripennis*, and *T. castaneum*, separately. The CAFÉ analysis identified 13,356 and 13,431 genes assigned to OrthoDB v11 gene families in *B. lapidarius* and *B. niveatus*, respectively. We identified 2810 single-copy ortholog groups in *B. lapidarius* and 2882 in *B. niveatus* for gene family expansion and construction analysis across insect genomes. The CAFÉ analysis of gene families unveiled significant expansions in 902 families and contractions in 331 families in *B. lapidarius*. Similarly, in *B. niveatus*, 682 families exhibited expansions, while 332 families underwent contractions (Supplementary Dataset 1). Furthermore, we identified 301 rapidly evolving gene families in *B. lapidarius*, comprising 217 expansions and 84 contractions, with a statistical significance level of $p < 0.05$.

Table II Genome annotation statistics of *B. lapidarius* and *B. niveatus*

Genome annotation	*B. lapidarius*			*B. niveatus*		
	Number	Size	Percentage of the genome	Number	Size	Percentage of the genome
Protein-coding genes	23,266	25,568,202	10.478	23,792	27,246,996	10.506
Pseudogenes	82	58,68	0.002	242	17,121	0.007
rRNAs	115	26,126	0.011	228	70,419	0.027
tRNAs	220	16,486	0.007	277	20,239	0.008
miRNA	62	4872	0.002	63	4972	0.002

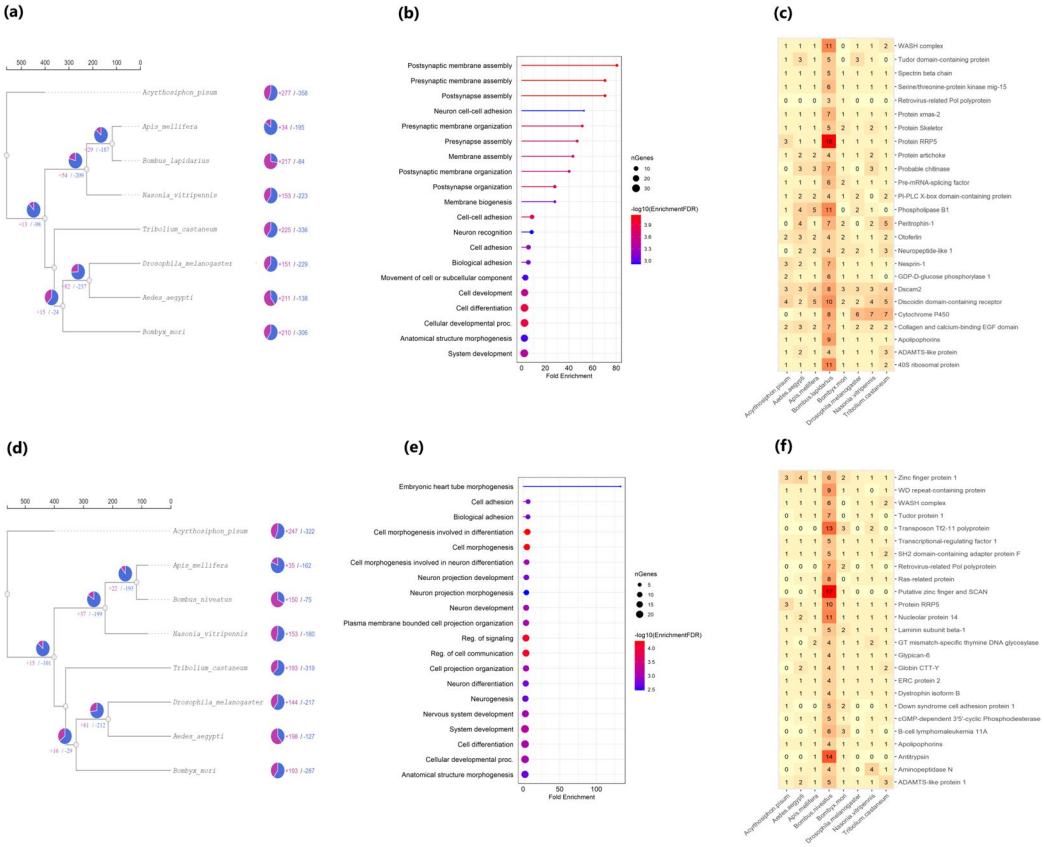


Figure 2. Phylogenetic trees illustrate the evolutionary relationships of seven insect species with *B. lapidarius* (**a**) and *B. niveatus* (**d**). Trees compare gene family expansions and contractions concerning their most recent common ancestor. Venn diagrams at the nodes depict the ratio of gene family expansions (blue) and contractions (magenta) across the insect species. These diagrams provide insights into shared and species-specific genomic changes. Enriched biological process pathways associated with expanded and contracted genes in *B. lapidarius* (**b**) and *B. niveatus* (**e**). Dot size indicates the number of genes associated with pathways, while colors denote the false discovery rate (FDR) significance level. The number of rapidly expanded genes per gene family is depicted for *B. lapidarius* (**c**) and *B. niveatus* (**f**). The *x*-axis represents species, and the *y*-axis shows gene families.

In *B. niveatus*, 225 rapidly evolving gene families were identified, encompassing 150 expansions and 75 contractions ($p < 0.05$) (Figure 2a, d). The GO analysis revealed that both *Bombus* species shared a significant joint enrichment (q-value < 0.05) in biological processes across several gene families. These processes included biological adhesion (GO:0007155), system development (GO:0048731), cell differentiation (GO:0030154), cellular developmental process (GO:0048869), and anatomical structure morphogenesis (GO:0009653). However, the expanded genes in *B. lapidarius* are uniquely associated with biological processes such as postsynapse assembly (GO:0099068), presynapse assembly (GO:0099054), membrane assembly (GO:0071709), postsynaptic membrane organization (GO:0001941), postsynapse organization (GO:0099173), and membrane biogenesis (GO:0044091) (Figure 2b, Table S5). In contrast, the expanded genes of *B. niveatus* are uniquely involved in biological processes like cell morphogenesis during differentiation (GO:0000902), cell morphogenesis during

neuron differentiation (GO:0048667), neuron projection development (GO:0031175), neuron projection morphogenesis (GO:0048812), neuron development (GO:0048666), regulation of signaling (GO:0023051), regulation of cell communication (GO:0010646), cell projection organization (GO:0030030), neuron differentiation (GO:0030182), and nervous system development (GO:0007399) (Figure 2e, Table S6). Based on these findings, it appears that *B. lapidarius* has expanded genes related to the pre- and post-synaptic organization in biological processes, while *B. niveatus* has expanded genes related to cell morphogenesis and neuron development. Across both species, apolipophorins, RRP5 protein, WASH complex, and Tudor proteins are gene families that have undergone expansion in gene number. Excluding housekeeping genes, gene families encoding phospholipase B1, discoidin domain-containing receptors, Dscam2, nesprin-1, protein xmas-2, and peritrophin-1 were found to be significantly expanded in *B. lapidarius*. Conversely, the SCAN domain, antitrypsin, WD-repeat-containing, and Ras-related protein gene families were found to be expanded in *B. niveatus* compared to other insect species (Figure 2c, f).

3.3. Identification of candidate coding genes under positive selection

A maximum likelihood phylogenetic analysis using PAML v4.0 embedded in PosiGene investigated whether gene orthogroups independently experience positive selection in *B. lapidarius* and *B. niveatus* lineages. Positive selection was identified by an elevated ratio of non-synonymous to synonymous mutations ($dN/dS > 1$) with a significance level of $p < 0.05$. The phylogenetic tests for selection in *B. lapidarius* and *B. niveatus* were performed by aligning CDS sequences from species used in the gene family analysis. A total of 180 and 182 genes, respectively, exhibited evidence of significant positive selection in *B. lapidarius* and *B. niveatus* (Supplementary Dataset 2). While approximately 25% of genes are common, most positively selected genes we identified are likely associated with species-specific adaptations. The functional analysis for common genes containing positive selection sites showed that those genes were significantly enriched for various biological processes, including rRNA methylation, modification, biogenesis, TOR signaling, brain, central nervous system and glial cell development, tissue morphogenesis, gland morphogenesis neurogenesis, open tracheal system development, tube development, metamorphosis, and locomotion (Supplementary Dataset 2). Notably, PSGs spanned a spectrum of biological and molecular functions in both *Bombus* species. The significant GO terms of PSGs in *B. lapidarius* were associated with biological processes such as ovarian follicle and epithelial cell migration and development, tissue migration, female gamete generation (oogenesis), response to growth factor stimulus, membrane organization, cell motility, and locomotion (Figure 3a). As for *B. niveatus*, PSGs were involved in the following biological processes: rRNA methylation and transport, TOR and TORC1 signaling, mitochondrial gene expression, gland development, imaginal disc and tube development, tissue morphogenesis, larval and pupal morphogenesis, ovarian follicle cell development, oogenesis, cell motility, and locomotion (Figure 3c). The most highly represented functional clusters in molecular function were ATP-dependent activities (hydrolysis, peptidase), nucleic acid phosphatase activity, ATP, RNA, purine and nucleoside phosphate binding, ribonucleotide binding, and small molecular binding function in *B. lapidarius* (Figure 3b). In contrast, the PSGs of *B. niveatus* exhibited enrichment in RNA methyltransferase, DNA helicase, and catalytic activities, as well as functions related to nucleotide, purine nucleotide, anion, and nucleoside phosphate binding (Figure 3d). A protein–protein interaction analysis was also performed on PSGs to gain insight into their enrichment and molecular functions. We filtered the interactions from the PPI network and considered only high-confidence protein–protein associations (combined score > 0.7). In *B. lapidarius*, a robust PPI is observed between Naxe and DCP1, as well as between DMAP and DIP2. Further analysis of

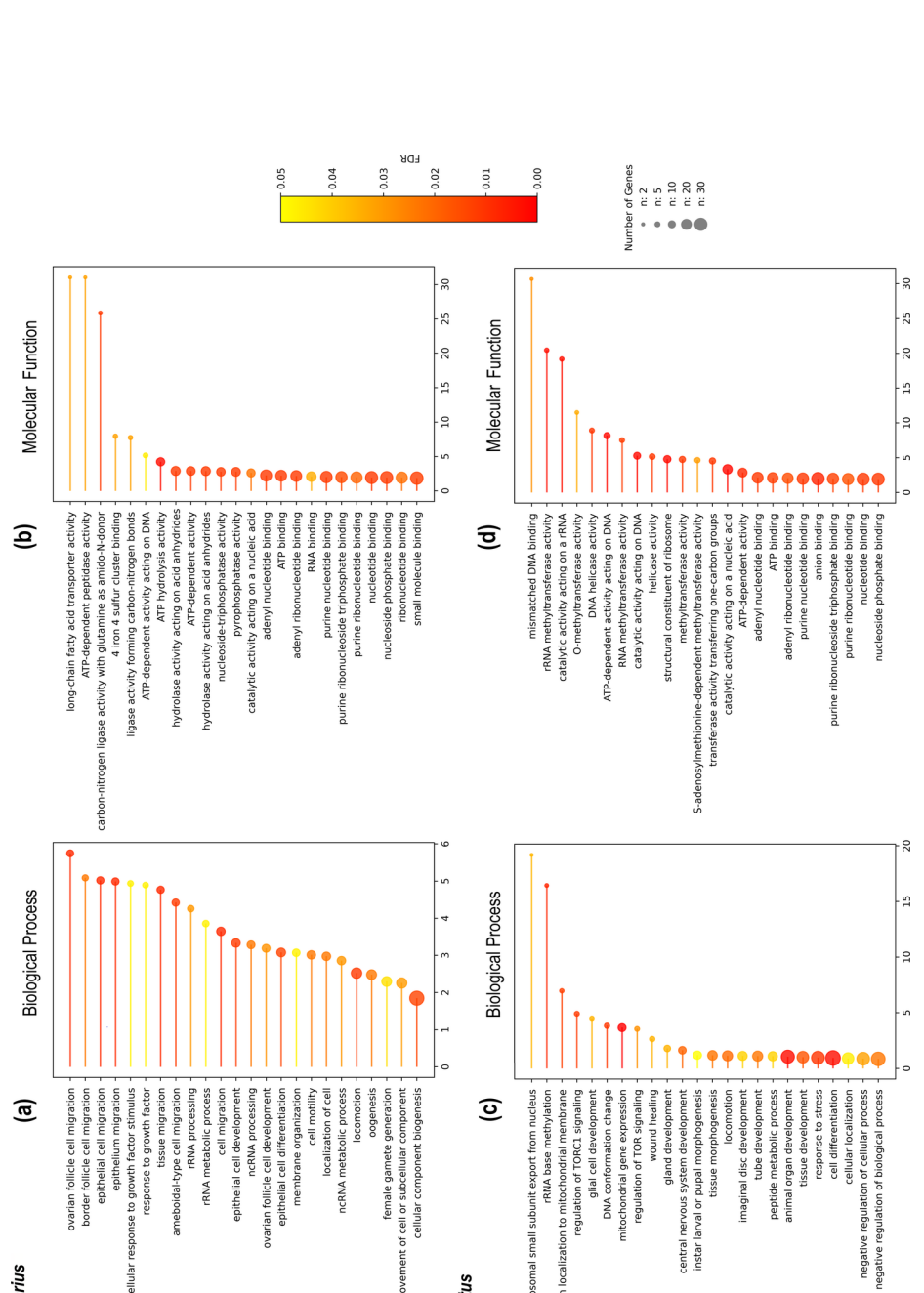


Figure 3. Gene ontology enrichment analysis for biological processes (left) and molecular functions (right) of positively selected genes of *B. lapidarius* (**a**, **b**) and *B. niveatus* (**c**, **d**). The *x*-axis of the graph represents the degree of enrichment fold changes, while the colors represent FDR values. The size of the dots corresponds to the number of genes enriched in a particular GO term.

the PPI network uncovered three distinct clusters: Smurf, dpp, Gli, and Su(fu); MCU, Spg7, and YME1L; and Ppcs, Got2, and AsnS (Figure 4a). Except for housekeeping genes involved in the assembly and function of ribosomes, such as mRpL (proteins L), mRpS (proteins L), mRRF (mitochondrial ribosomal release factor), and mRF (ribosomal release factor), potential strong interactions have been shown to exist between Rec and Mcm7, Spel1 and Pms2, and Rhau, Myd88, and Mybb1A proteins in *B. niveatus* (Figure 4b). Additionally, interactions have been found between Udt and Pck, Pex19 and Pex1, and Orco and GlyRS.

3.4. Annotation of immunity, chemosensory, and detoxification genes in *B.* lapidarius and *B.* niveatus

Insects are constantly exposed to pathogenic microorganisms and eukaryotic parasites, and their immune genes play a crucial role in recognizing and defending against these pathogens. A total of 360 genes have been identified in the immune gene repertoire in two bumblebees (Figure 3a, Table S7). The immune repertoires of *B. lapidarius* and *B. niveatus* show similarities, reflecting a wider trend seen among various *Bombus* species. The genes encoding SRRPs (small RNA regulatory pathway), CLIPs (CLIP domain serine proteases), APHAGs (autophagy-related genes), PRDXs (peroxiredoxins, pluralize), SCRs (scavenger receptors, pluralize), and CTLs (C-type lectins, pluralize) are highly represented in all *Bombus* species. While there are minor differences (e.g., orthologs of the cytokine receptor domeless and the antimicrobial effector TEPB are absent in *B. lapidarius* and *B. niveatus*, respectively), the overall number of immunogens within the Apidae family remains largely consistent across species. In contrast, the model Diptera, *D. melanogaster*, exhibits a considerably more extensive immunogen repertoire than the *Bombus* species (Table S7).

Chemical sensing plays a vital role in the life cycle of insects, enabling them to locate food sources, choose mating partners, establish nests, recognize nest mates, facilitate communication among nest mates, and respond to various environmental signals. Chemosensory receptor proteins comprise odorant receptors (ORs), gustatory receptors (GRs), and ionotropic receptors (IRs). On the other hand, non-receptor proteins play a crucial role in signal formation for detecting odor molecules. These proteins encompass odorant-binding proteins (OBPs), chemosensory proteins (CSPs), and sensory neuron membrane proteins (SNMPs). Insects possess six distinct gene families in their chemosensory repertoire (Sánchez-Gracia et al. 2009; Hansson and Stensmyr 2011; Fleischer et al. 2018; Breer et al. 2019; Cassau and Krieger 2021). A total of 384 genes across six chemoreceptor gene families were identified in the genomes of *B. lapidarius* and *B. niveatus*. While the overall number of ORs remained similar across the studied bumblebee species, the L clade emerged as the most populous clade within each species, followed by the 9-exon clade. Notably, apart from the well-conserved clades A, B, C, F, G, I, M, Q, W, and Z, which contained single sequence copies shared across all bumblebees, the remaining clades harbored multiple distinct OR sequences (Figure 3b, c; Table S8-9). A total of 183 genes have been identified in the repertoire of carboxyl/cholinesterases (CCE), cytochrome P450 monooxygenases (P450), and glutathione S-transferases (GST) in the two bumblebee species. In general, *B. lapidarius* and *B. niveatus* exhibit patterns similar to those of other *Bombus* species in terms of the number of genes known to be related to detoxification. P450 monooxygenase genes, well-known for their involvement in insect detoxification, were more frequently detected in the genomes of bumblebee species than GST and CCE genes. Genes encoding enzymes within the GST superfamily were detected similarly in other *Bombus* species, except for *B. bifarius*, *B. vancouverensis*, and *B. vosnesenskii* (Figure 5d, Table S10).

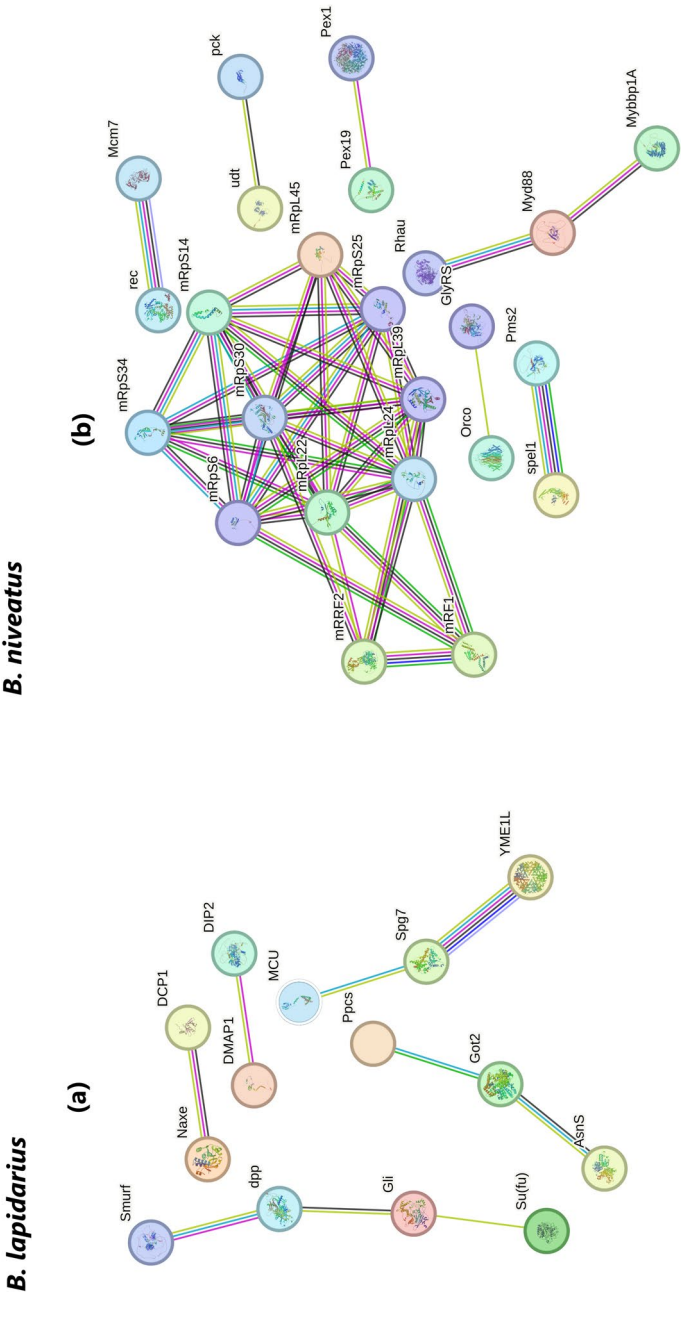


Figure 4. The high-confidence protein–protein association network of positively selected genes in **a** *B. lapidarius* and **b** *B. niveatus* (combined score > 0.7).

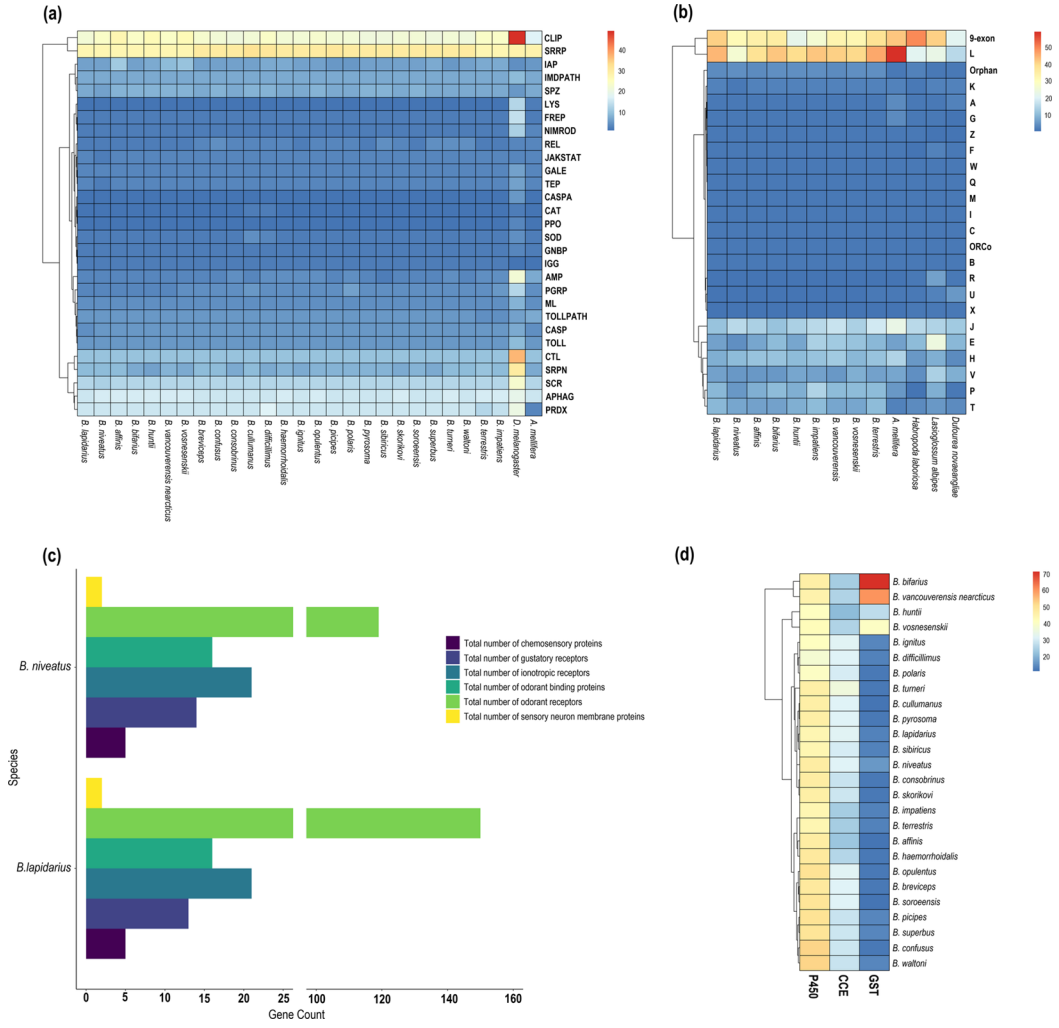


Figure 5. **a** Comparison of immune gene expression in *B. lapidarius* and *B. niveatus* across *Bombus* species, along with a model derived from *D. melanogaster*. The heatmap illustrates the differential expression of immune-related genes, highlighting variations between the bumblebee species, *A. mellifera*, and *D. melanogaster*. **b** This heatmap illustrates the distribution of OR clades within various bee genera. The color intensity represents the prevalence of specific OR clades across the different species. **c** Bar chart illustrating the chemosensory repertoires of *B. lapidarius* and *B. niveatus*, with 384 genes across six gene families. **d** Heatmap visualizing the distribution of 183 detoxification genes across three main families (carboxyl/cholinesterases, cytochrome P450 monooxygenases, and glutathione S-transferases) in *B. lapidarius*, *B. niveatus*, and other bumblebees.

4. DISCUSSION

Recent advances in long-read sequencing technologies and dropping costs have enabled the contiguous and complete sequencing of small-sized genomes (< 500 Mb), such as bumblebees, with unprecedented accuracy (Faulk 2023). In particular, given the synergistic impact of third-generation sequencing technologies and de novo assembly methods, insect genome sequencing and annotation studies have experienced rapid development, mirroring advancements observed in other eukaryotic organisms. In the Arthropoda assembly assessment catalog database (A^3Cat),

reportedly, there were 5792 assemblies representing 3412 species across 68 orders (Feron and Waterhouse 2022). Owing to its ecological significance, particularly in pollination, 43 whole-genome studies on *Bombus* species are reported in the NCBI Genome database, but only 8 of these species have genome annotations, which are indispensable for obtaining biological inference. This study is the first to report both genomic and comprehensive transcriptome data for *B. niveatus*. While genomic sequence information for *B. lapidarius* exists in NCBI, we generated its genome annotation for the first time. When examining *Bombus* genomes in the NCBI Genome database, it is observed that over 85% of *Bombus* species have an approximate genome length ranging from 250 to 280 Mb. Our study found that the genome lengths of *B. lapidarius* and *B. niveatus* fall within the expected range and are consistent with the typical length of the *Bombus* genomes previously reported (Sun et al. 2021). Synteny analysis also revealed a similarity in terms of chromosomal structure between *B. lapidarius*, *B. niveatus*, and *B. terrestris* (see Fig. S2).

High-quality genome sequences, comparative genomic analysis, and population genetics have enabled the search for genes involved in environmental adaptation and stress tolerance in bees (Grozinger and Zayed 2020; Heraghty et al. 2022; Everitt et al. 2023). Furthermore, recent studies have shown that it is necessary to elucidate the evolutionary dynamics within bumblebee genomes to understand bumblebees' adaptability to human-altered environments or climate changes (Jackson et al. 2018; Maebe et al. 2021; Liu et al. 2024). Considering habitat preferences, a global climate scenario suggests that the current distribution area of *B. lapidarius*, a wood-edge species, is projected to reduce, while that of *B. niveatus*, which prefers steppe habitats, is expected to expand (Rasmont et al. 2021). Therefore, highlighting the expansion and contraction of gene family across taxa is crucial for understanding the genomic plasticity of the organism and environmental adaptation (Ramasamy et al. 2016; Wang et al. 2019; Almudi et al. 2020; Xiao et al. 2020; Ylla et al. 2021). Analysis of gene family size variations revealed that expanded genes in *B. lapidarius* and *B. niveatus* involve biological processes such as neural system development, neural differentiation and morphogenesis, and presynaptic and postsynaptic membrane organization. However, *B. niveatus* exhibits a distinct additional expansion in the gene family, particularly regulating cell communication. Compared to other insect species, *B. niveatus* also shows expansion in gene families encoding SCAN domains, antitrypsin, WD-repeats, and Ras-related proteins. The SCAN domain is an 84-residue motif highly conserved and found near the N-terminus of a subfamily of C2H2 zinc finger proteins (ZFPs) (Edelstein and Collins 2005). ZFPs are abundant in eukaryotic genomes and constitute a class of transcription factors containing zinc finger domains. They play important regulatory roles in cellular growth, aging, and responses to abiotic and biotic stressors (Seetharam and Stuart 2013; Guo et al. 2021; Tao et al. 2023). A study with *Apis cerana* found that ZFP41 expression was significantly induced in response to various environmental stressors that cause oxidative damage in vivo, including heat, cold, UV light, and pesticides. When the ZFP41 gene was silenced using the RNA interference method, the transcript levels of antioxidant genes and the activities of antioxidant enzymes showed an increase. These results collectively showed us that ZFP41 protects honeybees against oxidative stress (Guo et al. 2021). A study on fall armyworm (*Spodoptera frugiperda*) revealed that heat stress in females' upregulated zinc finger proteins and genes associated with reproduction (Tao et al. 2023). A study using RNAi and dual-luciferase assays in the cotton aphid (*Aphis gossypii*) showed that CF2-II, a transcription factor, as a negative regulator of ABC transporters (ABCB5, ABCF2, MRP12) linked to multi-insecticide resistance and this finding suggests how CF2-II regulates the aphid's resistance development to multi-insecticide stress (Lv et al. 2023). In addition to responding to abiotic stress, C2H2-ZFPs have also been shown to be responsive to viral infection. Specifically, infection by nucleopolyhedrovirus in silkworm (*B. mori*) induces the

upregulation of six Zf-AD-containing C2H2 genes (Li et al. 2020). The genome analysis of *Nebria riversi* (Carabidae), adapted to high-elevation alpine habitats, has also revealed expansions in zinc finger proteins, potentially linked to cold adaptation (Weng et al. 2021).

In *B. niveatus*, another gene family showing significant expansion is the WD sequence repeat family. This family plays essential roles in diverse cellular processes, including signal transduction, RNA processing, vesicular trafficking, cytoskeleton assembly, cell cycle regulation, and transcription initiation complex assembly (Smith 2008). For instance, a recent study demonstrated that RACK1, a member of the WD-repeat protein family, also known as AccRACK1, plays a significant role in the development of *A. cerana*'s response to abiotic stress (Wang et al. 2023). Collectively, these results suggest that genes with positive expansion in *B. niveatus* potentially enhance the organism's response to various biotic and abiotic stressors, including environmental factors (heat, cold, and pesticides) and viruses. Furthermore, the differential expansion of gene families observed in *B. lapidarius* compared to *B. niveatus* may indicate distinct adaptation capacities between these species.

Besides, recent research highlights the dynamic evolution of neurological genes in bumblebees, potentially driven by environmental pressures (Colgan et al. 2022; Hart et al. 2022; Huml et al. 2023). This evolution manifests in both positive selection of genes related to complex behaviors and distinct patterns of gene family expansions associated with neuronal functions. Notably, *B. lapidarius* exhibits expansions in genes involved in synaptic organization, potentially refining neuronal communication, whereas *B. niveatus* shows expansions in genes linked to neuronal development and complexity. These findings suggest a crucial role for neurological evolution in shaping species-specific adaptations and behaviors in bumblebees.

Genome-wide positive selection analysis provides new insights into various aspects of evolutionary biology, encompassing the adaptive capacities of organisms to respond to their natural environments, as well as evolutionary changes in reaction to anthropogenic environmental gradients and shifts over ecologically relevant timeframes (McCulloch and Waters 2022; Larragy et al. 2023). Our study identified positive selection acting on the DISCO Interacting Protein 2 (DIP2) gene in *B. lapidarius*. This gene is known to be involved in fatty acid metabolism (Nitta et al. 2017; Mondal et al. 2022). Disparities in dietary preferences or access to food sources may exist between *B. lapidarius* and *B. niveatus*. Fatty acids are essential components of lipids, and changes in fatty acid metabolism may be associated with the utilization of specific dietary lipids. Furthermore, differences in the habitat types utilized by *B. lapidarius* and *B. niveatus* may influence their access to distinct floral resources. The composition of floral nectar and pollen varies considerably, and adaptations in fatty acid metabolism might be intricately linked to the processing of these lipid-rich floral resources. In *B. lapidarius*, both the *yme1l* gene, encoding the i-AAA protease of the inner mitochondrial membrane, and the *spg7* gene, belonging to the conserved metazoan mitochondrial AAA + protease family, harbor positive selection sites and exhibit strong protein–protein interactions. A recent study used the CRISPR/Cas9 method to generate *Drosophila* mutants lacking the *spg7* gene homolog. These mutants displayed a shortened lifespan, progressive locomotor defects, increased sensitivity to chemical and environmental stressors, and exhibited both muscular and neuronal degeneration (Pareek et al. 2018). Another study in *Drosophila* involved the deletion of the dYME1L and generated mutant flies (dYME1LΔ) exhibiting premature aging, progressive locomotor dysfunction, and neurodegeneration. Furthermore, aged dYME1LΔ flies displayed reduced complex I (NADH/ubiquinone oxidoreductase) activity, elevated reactive oxygen species levels, severely disorganized mitochondrial membranes, and increased apoptosis (Qi et al. 2016). The positive selection observed in the strongly interacting proteins SPG7 and YME1L might contribute to enhanced environmental stress resistance in *B. lapidarius*, which exhibits a broader geographical distribution but restricts its presence to specific habitat types. A

strong interaction was observed among proteins exhibiting positive selection in *B. nivetus*, including *Myd88*, *Mybbp1A*, and *Rhau* (Figure 4b). Studies suggest MyD88 plays a central role in stress response. In *B. mori*, MyD88 expression increases upon pesticide exposure, potentially acting as a biomarker (Hao et al. 2023). Conversely, in *D. melanogaster* infected with *Staphylococcus xylosus*, MyD88 mutants show increased resistance to starvation, suggesting its negative role in starvation response (Ayyaz et al. 2013). These findings suggest that MyD88 may regulate metabolism in response to stress. MYBBP1A, another key regulator, plays a crucial role in diverse cellular processes, including mitosis, metabolism, cellular signaling, and stress responses (Owen et al. 2007; Mori et al. 2012). This multifaceted role contributes to cellular homeostasis and adaptation. In response to environmental stress, eukaryotic cells reprogram their translational machinery. The DEAH-box RNA helicase RHAU, a cytoplasmic protein, localizes to stress granules (SGs) within the cell. A study revealed that RHAU physically interacts with RNA through a newly identified N-terminal RNA-binding domain, both in vitro and in vivo. This domain is essential and sufficient for RHAU's localization to SGs. Furthermore, the study demonstrated that RHAU's ATPase activity plays a role in RNA interaction and regulates protein retention within SGs (Chalupníková et al. 2008). In *B. niveatus*, *Myd88*, *Mybbp1A*, and *Rhau* genes exhibit positive selection and may be associated with adaptations for environmental stress responses within its steppe habitat. Regarding immune, detoxification, and chemoreceptor genes, both *B. lapidarius* and *B. niveatus* generally exhibit similar trends to other *Bombus* species in the number of immunity, chemosensory, and detoxification genes. *B. niveatus* exhibits a reduced repertoire of odorant receptors compared to *B. lapidarius*, as evidenced by the lower number of genes in the 9-exon and L clade families (Table S8-9). This disparity may be associated with the global distribution patterns and habitat preferences of these species. *B. lapidarius*, with its wider geographic distribution but more specific habitat selection, might have evolved a more elaborate sense of smell, reflected by its expanded odorant receptor repertoire, to navigate its particular ecological niche.

In conclusion, the genome assemblies and annotations of *B. lapidarius* and *B. niveatus* presented here constitute a valuable resource for bumblebee genomic research. We found that the expansion and contraction of gene families and genome-wide positive selection analysis reveal potentially comprising adaptive gene repertoires for each *Bombus* species. We anticipate that this data will significantly enhance studies to determine the population-level genomic plasticity of bumblebees against changing climate conditions.

SUPPLEMENTARY INFORMATION

The online version contains supplementary material available at https://doi.org/10.1007/s13592-024-01140-1.

ACKNOWLEDGEMENTS

We are grateful to all the members of the Molecular Biodiversity and Systematic Entomology laboratory for fruitful discussions.

AUTHOR CONTRIBUTION

VE, FD, and ÇÖ conceived the manuscript. FD, ÇÖ, BDÖ, MAB, SBÇ, and YUÇ collected bumblebee samples. VE, GZ, YB, OO, SCK, TD, YUÇ, SBÇ, PA, TÖ, and MAB performed the experiments and analysis. SCK, OO, and YUÇ prepared the figures and tables. VE, FD, TD, ÇÖ, and BDÖ wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

Open access funding provided by the Scientific and Technological Research Council of Türkiye (TÜBİTAK). This study was funded by the Scientific Research Projects Coordination Unit of Istanbul University under Project number 37223 and 36621. Computing resources used in this work were provided by the National Center for High Performance Computing of Turkey (UHeM) under grant number 5004732017 and by the Research Fund of Erciyes University [TSG-2021–10912]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DATA AVAILABILITY

All genomic oxford nanopore raw sequencing data used in this study have been deposited in the NCBI Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) database with the accession numbers PRJNA875259 and PRJNA876206. Raw RNA-sequencing data for *B. lapidarius* and *B. niveatus* were deposited in the NCBI SRA database under BioProject accession numbers PRJNA1045976 and PRJNA1045977, respectively. The genomic data, including coding and peptide sequences, as well as annotation files for both *Bombus* species, are accessible on our GitHub repository: https://github.com/MolecularBioDiversityLab/Bombus_genome_data. All the rest of the data are available in the main text or the supplementary materials.

CODE AVAILABILITY

Not applicable.

DECLARATIONS

Ethics approval No approval of the Research Ethics Committee was required to achieve the goals of this study, as the experimental work involved unregulated invertebrate species (bumblebee species).

Consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

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

REFERENCES

Almudi I, Vizueta J, Wyatt CDR, de Mendoza A, Marlétaz F, Firbas PN, Feuda R, Masiero G, Medina P, Alcaina-Caro A, Cruz F, Gómez-Garrido J, Gut M, Alioto TS, Vargas-Chavez C, Davie K, Misof B, González J, Aerts S, Lister R, Paps J, Rozas J, Sánchez-Gracia A, Irimia M, Maeso I, Casares F (2020) Genomic adaptations to aquatic and aerial life in mayflies and the origin of insect wings. Nat Commun 11(1):2631

Astashyn A, Tvedte ES, Sweeney D, Sapojnikov V, Bouk N, Joukov V, Murphy TD (2024) Rapid and sensitive detection of genome contamination at scale with FCS-GX. Genome Biol 25(1):60

Ayyaz A, Giammarinaro P, Liégeois S, Lestradet M, Ferrandon D (2013) A negative role for MyD88 in the resistance to starvation as revealed in an intestinal infection of Drosophila melanogaster with the Gram-positive bacterium Staphylococcus xylosus. Immunobiology 218(4):635–644

Bao W, Kojima KK, Kohany O (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA 6:1–6

Bao Z, Eddy SR (2002) Automated de novo identification of repeat sequence families in sequenced genomes. Genome Res 12:1269–1276

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580

Biella P, Ćetković A, Gogala A, Neumayer J, Sárospataki M, Šima P, Smetana V (2021) Northwestward range expansion of the bumblebee *Bombus haematurus* into Central Europe is associated with warmer winters and niche conservatism. Insect Sci 28(3):861–872

Biesmeijer JC, Roberts SP, Reemer M, Ohlemuller R, Edwards M, Peeters T, Schaffers AP, Potts SG, Kleukers R, Thomas CD, Settele J, Kunin WE (2006) Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. Science 313(5785):351–354

Birney E, Clamp M, Durbin R (2004) GeneWise and Genomewise. Genome Res 14:988–995

Brand P, Ramirez SR, Leese F, Quezada-Euan JJ, Tollrian R, Eltz T (2015) Rapid evolution of chemosensory receptor genes in a pair of sibling species of orchid bees (Apidae: Euglossini). BMC Evol Biol 15:176

Breer H, Fleischer J, Pregitzer P, Krieger J (2019) Molecular mechanism of insect olfaction: olfactory receptors. In: Picimbon JF (ed) Olfactory concepts of insect control - alternative to insecticides. Springer, Cham, pp 93–114

Brůna T, Lomsadze A, Borodovsky M (2020) GeneMark-EP+: eukaryotic gene prediction with self-training in the space of genes and proteins. NAR genom bioinform 2(2):lqaa026

Cameron SA, Hines HM, Williams PH (2007) A comprehensive phylogeny of the bumble bees (Bombus). Biol J Lin Soc 91(1):161–188

Cameron SA, Sadd BM (2020) Global trends in bumble bee health. Annu Rev Entomol 65:209–232

Cameron SA, Williams PH (2003) Phylogeny of bumble bees in the New World subgenus Fervidobombus (Hymenoptera: Apidae): congruence of molecular and morphological data. Mol Phylogenet Evol 28(3):552–563

Cantalapiedra CP, Hernandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J (2021) eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol 38:5825–5829

Carvalheiro LG, Kunin WE, Keil P, Aguirre-Gutiérrez J, Ellis WN, Fox R, Groom Q, Hennekens S, Van Landuyt W, Maes D, Van de Meutter F, Michez D, Rasmont P, Ode B, Potts SG, Reemer M, Roberts SP, Schaminée J, WallisDeVries MF, Biesmeijer JC (2013) Species richness declines and biotic homogenisation have slowed down for NW-European pollinators and plants. Ecol Lett 16(7):870–878

Cassau S, Krieger J (2021) The role of SNMPs in insect olfaction. Cell Tissue Res 383(1):21–33

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17(4):540–552

Chalupníková K, Lattmann S, Selak N, Iwamoto F, Fujiki Y, Nagamine Y (2008) Recruitment of the RNA helicase RHAU to stress granules via a unique RNA-binding domain. J Biol Chem 283(50):35186–35198

Chan PP, Lowe TM (2019) tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol 1962:1–14

Chen YT, Liu ZX, Régnière J, Vasseur L, Lin J, Huang SG, Ke FS, Chen SP, Li JY, Huang JL, Gurr GM, You MS, You SJ (2021) Large-scale genome-wide study reveals climate adaptive variability in a cosmopolitan pest. Nat Commun 12(1):7206

Christenhusz MJ, Byng JW (2016) The number of known plants species in the world and its annual increase. Phytotaxa 261(3):201–217

Colgan TJ, Arce AN, Gill RJ, Ramos Rodrigues A, Kanteh A, Duncan EJ, … Li L, Chittka L, Wurm, Y (2022) Genomic signatures of recent adaptation in a wild bumblebee. Mol Biol Evol 39(2):msab366

De Bie T, Cristianini N, Demuth JP, Hahn MW (2006) CAFE: a computational tool for the study of gene family evolution. Bioinformatics 22:1269–1271

Edelstein LC, Collins T (2005) The SCAN domain family of zinc finger transcription factors. Gene 359:1–17

Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol 20:238

Everitt T, Wallberg A, Christmas MJ, Olsson A, Hoffmann W, Neumann P, Webster MT (2023) The genomic basis of adaptation to high elevations in Africanized honey bees. Genome Biol Evol 15(9):evad157

Faulk C (2023) *De novo* sequencing, diploid assembly, and annotation of the black carpenter ant, Camponotus pennsylvanicus, and its symbionts by one person for $1000, using nanopore sequencing. Nucleic Acids Res 51(1):17–28

Feron R, Waterhouse RM (2022) Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes. Gigascience 11

Fleischer J, Pregitzer P, Breer H, Krieger J (2018) Access to the odor world: olfactory receptors and their role for signal transduction in insects. Cell Mol Life Sci 75:485–508

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF (2020) RepeatModeler2 for automated genomic discovery of transposable element families. PNAS 117(17):9451–9457

Forêt S, Maleszka R (2006) Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). Genome Res 16(11):1404–1413

Ge SX, Jung D, Yao R (2020) ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics 36(8):2628–2629

Ghisbain G, Gérard M, Wood TJ, Hines HM, Michez D (2021) Expanding insect pollinators in the Anthropocene. Biol Rev 96(6):2755–2770

Ghisbain G, Thiery W, Massonnet F, Erazo D, Rasmont P, Michez D, Dellicour S (2024) Projected decline in European bumblebee populations in the twenty-first century. Nature 628(8007):337–341

Grozinger CM, Zayed A (2020) Improving bee health through genomics. Nat Rev Genet 21(5):277–291

Guo HJ, Wang LJ, Wang C, Guo DZ, Xu BH, Guo XQ, Li H (2021) Identification of an *Apis cerana* zinc finger protein 41 gene and its involvement in the oxidative stress response. Arch Insect Biochem Physiol 108(1)

Halsch CA, Shapiro AM, Fordyce JA, Nice CC, Thorne JH, Waetjen DP, Forister ML (2021) Insects and recent climate change. P Natl Acad Sci USA 118(2)

Hansson BS, Stensmyr MC (2011) Evolution of insect olfaction. Neuron 72(5):698–711

Hao ZH, Lu QY, Zhou YY, Liang YT, Gao Y, Ma HY, Xu YS, Wang HB (2023) Molecular characterization of MyD88 as a potential biomarker for

pesticide-induced stress in *Bombyx mori*. Pestic Biochem Phys 196

Hart AF, Verbeeck J, Ariza D et al (2022) Signals of adaptation to agricultural stress in the genomes of two European bumblebees. Front Genet 13

He W, Yang J, Jing Y, Xu L, Yu K, Fang X (2023) NGenomeSyn: an easy-to-use and flexible tool for publication-ready visualization of syntenic relationships across multiple genomes. Bioinformatics 39(3):btad121

Heinrich B (1979) Resource heterogeneity and patterns of movement in foraging bumblebees. Oecologia 40:235–245

Heraghty SD, Rahman SR, Jackson JM, Lozier JD, Ware J (2022) Whole genome sequencing reveals the structure of environment-associated divergence in a broadly distributed montane bumble bee, Bombus vancouverensis. Insect Syst Diver 6(5):5

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M (2016) BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics 32(5):767–769

Hoff KJ, Lomsadze A, Borodovsky M, Stanke M (2019) Whole-genome annotation with BRAKER. Gene prediction: methods and protocols, pp 65–95

Huml JV, Ellis JS, Rustage S, Brown MJ, Billington R, Knight ME (2023) The tragedy of the common? A comparative population genomic study of two bumblebee species. Insect Conserv Divers 16(3):335–354

Jackson JM, Pimsler ML, Oyen KJ, Koch-Uhuad JB, Herndon JD, Strange JP, Dillon ME, Lozier JD (2018) Distance, elevation and environment as drivers of diversity and divergence in bumble bees across latitude and altitude. Mol Ecol 27(14):2926–2942

Kalvari I, Nawrocki EP, Ontiveros-Palacios N, Argasinska J, Lamkiewicz K, Marz M, Griffiths-Jones S, Toffano-Nioche C, Gautheret D, Weinberg Z, Rivas E, Eddy SR, Finn RD, Bateman A, Petrov AI (2021) Rfam 14: expanded coverage of metagenomic, viral and microRNA families. Nucleic Acids Res 49(D1):D192–D200

Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2014) Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic Acids Res 42(D1):D199–D205

Karpe SD, Jain R, Brockmann A, Sowdhamini R (2016) Identification of complete repertoire of Apis florea odorant receptors reveals complex orthologous relationships with *Apis mellifera*. Genome Biol Evol 8(9):2879–2895

Kerr JT, Pindar A, Galpern P, Packer L, Potts SG, Roberts SM, Rasmont P, Schweiger O, Colla SR, Richardson LL, Wagner DL, Gall LF, Sikes DS, Pantoja A (2015) Climate change impacts on bumblebees converge across continents. Science 349(6244):177–180

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature Biotechnol 37(8):907–915

Kolmogorov M, Yuan J, Lin Y, Pevzner PA (2019) Assembly of long, error-prone reads using repeat graphs. Nature Biotechnol 37(5):540–546

Larragy SJ, Möllmann JS, Stout JC, Carolan JC, Colgan TJ (2023) Signatures of adaptation, constraints, and potential redundancy in the canonical immune genes of a key pollinator. Genome Biol Evol 15(4):evad039

Li C, Zhang H, Gao R, Zuo W, Liu Y, Hu H, Luan Y, Lu C, Tong X, Dai F (2020) Identification and effect of Zf-AD-containing C2H2 zinc finger genes on BmNPV replication in the silkworm (*Bombyx mori*). Pestic Biochem Physiol 170

Liu YZ, Olsson A, Larva T, Cantwell-Jones A, Gill RJ, Cederberg B, Webster MT (2024) Genomic variation in montane bumblebees in Scandinavia: high levels of intraspecific diversity despite population vulnerability. Mol Ecol 33(4)

Loytynoja A (2021) Phylogeny-aware alignment with PRANK and PAGAN. Multiple Sequence Alignment 2231:17–37

Lv Y, Pan Y, Li J, Ding Y, Yu Z, Yan K, Shang Q (2023) The C2H2 zinc finger transcription factor CF2-II regulates multi-insecticide resistance-related gut-predominant ABC transporters in Aphis gossypii Glover. Int J Biol Macromol 253(Pt 2)

Maebe K, Hart AF, Marshall L, Vandamme P, Vereecken NJ, Michez D, Smagghe G (2021) Bumblebee resilience to climate change, through plastic and adaptive responses. Global Change Biol 27(18):4223–4237

Manni M, Berkeley MR, Seppey M, Zdobnov EM (2021) BUSCO: assessing genomic data quality and beyond. Current Protocols 1(12)

McCulloch GA, Waters JM (2022) Rapid adaptation in a fast-changing world: emerging insights from insect genomics. Global Change Biol 29(4):943–954

Mendes FK, Vanderpool D, Fulton B, Hahn MW (2020) CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics 36(22–23):5516–5518

Michener CD (2007) The bees of the world. John Hopkins University Press, p 913

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R (2020) IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534

Mondal S, Kinatukara P, Singh S, Shambhavi S, Patil GS, Dubey N, Singh SH, Pal B, Shekar PC, Kamat SS, Sankaranarayanan R (2022) DIP2 is a unique regulator of diacylglycerol lipid homeostasis in eukaryotes. Elife 11

Mori S, Bernardi R, Laurent A, Resnati M, Crippa A, Gabrieli A, Keough R, Gonda TJ, Blasi F (2012) Myb-binding protein 1A (MYBBP1A) is essential for early embryonic development, controls cell cycle and mitosis, and acts as a tumor suppressor. PLoS ONE 7(10)

Nawrocki EP, Eddy SR (2013) Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29(22):2933–2935

Nieto A, Roberts SPM, Kemp J, Rasmont P, Kuhlmann M, García Criado M, Biesmeijer JC, Bogusch P, Dathe HH, De la Rúa P, De Meulemeester T, Dehon M, Dewulf A, Ortiz-Sánchez FJ, Lhomme P, Pauly A, Potts SG, Praz C, Quaranta M, Radchenko VG, Scheuchl E, Smit J, Straka J, Terzo M, Tomozii B, Window J, Michez D (2014) European Red List of bees. Publication Office of the European Union, Luxembourg

Nishimura O, Hara Y, Kuraku S (2017) gVolante for standardizing completeness assessment of genome and transcriptome assemblies. Bioinformatics 33(22):3635–3637

Nitta Y, Yamazaki D, Sugie A, Hiroi M, Tabata T (2017) DISCO Interacting Protein 2 regulates axonal bifurcation and guidance of Drosophila mushroom body neurons. Dev Biol 421(2):233–244

Obiero GF, Pauli T, Geuverink E, Veenendaal R, Niehuis O, Grosse-Wilde E (2021) Chemoreceptor diversity in apoid wasps and its reduction during the evolution of the pollen-collecting lifestyle of bees (Hymenoptera: Apoidea). Genome Biol Evol 13(3):evaa269

Ouyang J, Zheng S, Huang M et al (2022) Chromosome-level genome and population genomics reveal evolutionary characteristics and conservation status of Chinese indigenous geese. Commun Biol 5(1):1191

Owen HR, Elser M, Cheung E, Gersbach M, Kraus WL, Hottiger MO (2007) MYBBP1a is a novel repressor of NF-κB. J Mol Biol 366(3):725–736

Pareek G, Thomas RE, Pallanck LJ (2018) Loss of the Drosophila m-AAA mitochondrial protease paraplegin results in mitochondrial dysfunction, shortened lifespan, and neuronal and muscular degeneration. Cell Death Dis 9(3):304

Pauw A (2007) Collapse of a pollination web in small conservation areas. Ecology 88(7):1759–1769

Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, Bileschi ML, Bork P, Bridge A, Colwell L, Gough J, Haft DH, Letunić I, Marchler-Bauer A, Mi H, Natale DA, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A (2023) InterPro in 2022. Nucleic Acids Res 51(D1):D418–D427

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat Protoc 11(9):1650–1667

Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD (2018) HMMER web server: 2018 update. Nucleic Acids Res 46(W1):W200–W204

Potts SG, Biesmeijer JC, Kremen C, Neumann P, Schweiger O, Kunin WE (2010) Global pollinator declines: trends, impacts and drivers. Trends Ecol Evol 25(6):345–353

Potts SG, Ngo HT, Biesmeijer JC, Breeze TD, Dicks LV, Garibaldi LA, Hill R, Settele J, Vanbergen A (2016) The assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination and food production. Secretariat of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, Bonn, Germany, p 556

Price AL, Jones NC, Pevzner PA (2005) *De novo* identification of repeat families in large genomes. Bioinformatics 21(Suppl1):i351–i358

Qi Y, Liu H, Daniels MP, Zhang G, Xu H (2016) Loss of Drosophila i-AAA protease, dYME1L, causes abnormal mitochondria and apoptotic degeneration. Cell Death Differ 23(2):291–302

Ramasamy S, Ometto L, Crava CM, Revadi S, Kaur R, Horner DS, Pisani D, Dekker T, Anfora G, Rota-Stabelli O (2016) The evolution of olfactory gene families in drosophila and the genomic basis of chemical-ecological adaptation in. Genome Biol Evol 8(8):2297–2311

Rasmont P, Franzén M, Lecocq T, Harpke A, Roberts S, Biesmeijer JC, Castro L, Cederberg B, Dvorak L, Fitzpatrick Ú, Gonseth Y, Haubruge E, Mahé G, Manino A, Michez D, Neumayer J, Ødegaard F, Paukkunen J, Pawlikowski T, Potts S, Reemer M, Settele J, Straka J, Schweiger O (2015) Climatic risk and distribution atlas of European bumblebees. BioRisk 10:1–236

Rasmont P, Ghisbain G, Terzo M (2021) Bumblebees of Europe and neighbouring regions. NAP Editions, p 631

Robertson HM, Wanner KW (2006) The chemoreceptor superfamily in the honey bee, *Apis mellifera*: expansion of the odorant, but not gustatory, receptor family. Genome Res 16(11):1395–1403

Sadd BM, Barribeau SM, Bloch G, de Graaf DC, Dearden P, Elsik CG, Gadau J, Grimmelikhuijzen CJ, Hasselmann M, Lozier JD, Robertson HM, Smagghe G, Stolle E, Van Vaerenbergh M, Waterhouse RM, Bornberg-Bauer E, Klasberg S, Bennett AK, Camara F, Guigo R, Hoff K, Mariotti M, Munoz-Torres M, Murphy T, Santesmasses D, Amdam GV, Beckers M, Beye M, Biewer M, Bitondi MM, Blaxter ML, Bourke AF, Brown MJ, Buechel SD, Cameron R, Cappelle K, Carolan JC, Christiaens O, Ciborowski KL, Clarke DF, Colgan TJ, Collins DH, Cridge AG, Dalmay T, Dreier S, du Plessis L, Duncan E, Erler S, Evans J, Falcon T, Flores K, Freitas FC, Fuchikawa T, Gempe T, Hartfelder K, Hauser F, Helbing S, Humann FC, Irvine F, Jermiin LS, Johnson CE, Johnson RM, Jones AK, Kadowaki T,

Kidner JH, Koch V, Kohler A, Kraus FB, Lattorff HM, Leask M, Lockett GA, Mallon EB, Antonio DS, Marxer M, Meeus I, Moritz RF, Nair A, Napflin K, Nissen I, Niu J, Nunes FM, Oakeshott JG, Osborne A, Otte M, Pinheiro DG, Rossie N, Rueppell O, Santos CG, Schmid-Hempel R, Schmitt BD, Schulte C, Simoes ZL, Soares MP, Swevers L, Winnebeck EC, Wolschin F, Yu N, Zdobnov EM, Aqrawi PK, Blankenburg KP, Coyle M, Francisco L, Hernandez AG, Holder M, Hudson ME, Jackson L, Jayaseelan J, Joshi V, Kovar C, Lee SL, Mata R, Mathew T, Newsham IF, Ngo R, Okwuonu G, Pham C, Pu LL, Saada N, Santibanez J, Simmons D, Thornton R, Venkat A, Walden KK, Wu YQ, Debyser G, Devreese B, Asher C, Blommaert J, Chipman AD, Chittka L, Fouks B, Liu J, O'Neill MP, Sumner S, Puiu D, Qu J, Salzberg SL, Scherer SE, Muzny DM, Richards S, Robinson GE, Gibbs RA, Schmid-Hempel P, Worley KC (2015) The genomes of two key bumblebee species with primitive eusocial organization. Genome Biol 16(1):76

Sánchez-Gracia A, Vieira FG, Rozas J (2009) Molecular evolution of the major chemosensory gene families in insects. Heredity 103(3):208–216

Seetharam A, Stuart GW (2013) A study on the distribution of 37 well conserved families of C2H2 zinc finger genes in eukaryotes. BMC Genomics 14

Shen W, Le S, Li Y, Hu F (2016) SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS ONE 11

Smit AFA, Hubley R, Green P (2017) RepeatMasker Open-4.0 [cited 2018 Oct 1]. Available from: http://www.repeatmasker.org/

Smith TF (2008) Diversity of WD-Repeat proteins. In: Clemen CS, Eichinger L, Rybakin V (eds) The Coronin Family of Proteins. Subcellular Biochemistry, vol 48. Springer, pp 20-30

Steffen R, Ogoniak L, Grundmann N, Pawluchin A, Soehnlein O, Schmitz J (2022) paPAML: an improved computational tool to explore selection pressure on protein-coding sequences. Genes 13(6)

Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF (2021) The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob DNA 12(1):1–14

Sun C, Huang JX, Wang Y, Zhao XM, Su L, Thomas GWC, Zhao MY, Zhang XT, Jungreis I, Kellis M, Vicario S, Sharakhov IV, Bondarenko SM, Hasselmann M, Kim CN, Paten B, Penso-Dolfin L, Wang L, Chang YX, Gao Q, Ma L, Ma LN, Zhang Z, Zhang HB, Zhang HH, Ruzzante L, Robertson HM, Zhu YH, Liu YJ, Yang HP, Ding LL, Wang QG, Ma DN, Xu WL, Liang C, Itgen MW, Mee L, Cao G, Zhang Z, Sadd BM, Hahn MW, Schaack S, Barribeau SM, Williams PH, Waterhouse RM, Mueller RL (2021) Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits. Mol Biol Evol 38(7):3031–3031

Sun JH, Lu F, Luo YJ, Bie LZ, Xu L, Wang Y (2023) OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. Nucleic Acids Res 51(W1):W397–W403

Suzek BE, Huang H, McGarvey P, Mazumder R, Wu CH (2007) UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics 23(10):1282–1288

Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, Roth A, Santos A, Tsafou KP, Kuhn M, Bork P, Jensen LJ, von Mering C (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43 (Database issue):D447–452

Tao YD, Liu Y, Wan XS, Xu J, Fu DY, Zhang JZ (2023) High and low temperatures differentially affect survival, reproduction, and gene transcription in male and female moths of *Spodoptera frugiperda*. Insects 14(12)

Teufel F, Almagro Armenteros JJ, Johansen AR, Gislason MH, Pihl SI, Tsirigos KD, Winther O, Brunak S, von Heijne G, Nielsen H (2022) SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 40(7):1023–1025

Theodorou P, Radzevičiūtė R, Kahnt B, Soro A, Grosse I, Paxton RJ (2018) Genome-wide single nucleotide polymorphism scan suggests adaptation to urbanization in an important pollinator, the red-tailed bumblebee (Bombus lapidarius L.). Proc R Soc Ser B Biol Sc Lond 285(1877):20172806

Tong XL, Han MJ, Lu KP, Tai SS, Liang SB, Liu YC, Hu H, Shen JH, Long AX, Zhan CY, Ding X, Liu S, Gao Q, Zhang BL, Zhou LL, Tan D, Yuan YJ, Guo NK, Li YH, Wu ZY, Liu LL, Li CL, Lu YR, Gai TT, Zhang YH, Yang RK, Qian HY, Liu YQ, Luo JW, Zheng L, Lou JH, Peng YW, Zuo WD, Song JB, He SZ, Wu SY, Zou YL, Zhou L, Cheng L, Tang YX, Cheng GT, Yuan LW, He WM, Xu JB, Fu T, Xiao Y, Lei T, Xu AY, Yin Y, Wang J, Monteiro A, Westhof E, Lu C, Tian ZX, Wang W, Xiang ZH, Dai FY (2022) High-resolution silkworm pan-genome provides genetic insights into artificial selection and ecological adaptation. Nat Commun 13(1):5619

Wang HF, Xie YC, Wang XQ, Geng XS, Gao LJ (2023) Characterization of the gene of *Aips cerana cerana* and its role in adverse environmental stresses. Comp Biochem Physiol B Biochem Mol Biol 263

Wang N, Yang Y, Moore MJ, Brockington SF, Walker JF, Brown JW, Liang B, Feng T, Edwards C, Mikenas J, Olivieri J, Hutchison V, Timoneda A, Stoughton T, Puente R, Majure LC, Eggli U, Smith SA (2019) Evolution of Portulacineae marked by gene tree conflict and gene family expansion associated with adaptation to harsh environments. Mol Biol Evol 36(1):112–126

Weng YM, Francoeur CB, Currie CR, Kavanaugh DH, Schoville SD (2021) A high-quality carabid genome assembly provides insights into beetle genome evolution and cold adaptation. Mol Ecol Resour 21(6):2145–2165

Williams P, Colla S, Xie Z (2009) Bumblebee vulnerability: common correlates of winners and losers across three continents. Conserv Biol 23(4):931–940

Xiao HM, Ye XH, Xu HX, Mei Y, Yang Y, Chen X, Yang YJ, Liu T, Yu YY, Yang WF, Lu ZX, Li F (2020) The genetic adaptations of fall armyworm facilitated its rapid global dispersal and invasion. Mol Ecol Resour 20(4):1050–1068

Xu L, Dong Z, Fang L, Luo Y, Wei Z, Guo H, Zhang G, Gu YQ, Coleman-Derr D, Xia Q, Wang Y (2019) OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. Nucleic Acids Res 47(W1):W52–W58

Xu Z, Wang H (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res 35(Suppl2):W265–W268

Yang Y, Huang L, Xu C, Qi L, Wu Z, Li J, Chen H, Wu Y, Fu T, Zhu H, Saand MA, Li J, Liu L, Fan H, Zhou H, Qin W (2021) Chromosome-scale genome assembly of areca palm (*Areca catechu*). Mol Ecol Resour 21(7):2504–2519

Yang ZH (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yang ZH, Wong WSW, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. Mol Biol Evol 22(4):1107–1118

Ylla G, Nakamura T, Itoh T, Kajitani R, Toyoda A, Tomonari S, Bando T, Ishimaru Y, Watanabe T, Fuketa M, Matsuoka Y, Barnett AA, Noji S, Mito T, Extavour CG (2021) Insights into the genomic evolution of insects from cricket genomes. Commun Biol 4(1):733

You M, Ke F, You S, Wu Z, Liu Q, He W, Baxter SW, Yuchi Z, Vasseur L, Gurr GM (2020) Variation among 532 genomes unveils the origin and evolutionary history of a global insect herbivore. Nature communications 11(1):2321

Zhou X, Slone JD, Rokas A, Berger SL, Liebig J, Ray A, Reinberg D, Zwiebel LJ (2012) Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding. PLoS Genet 8(8)

Zimin AV, Salzberg SL (2022) The SAMBA tool uses long reads to improve the contiguity of genome assemblies. PLoS Comput Biol 18(2)

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