scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly and annotation of two Asian bumble bees

Jixiang Cui[1], Ye Xu[2], Jiyao Liu[1], Xinyang Zhang[1], Zhaoyang Chen[1], Wanhu Yang[1], Jin Ge[3,8], Cheng Sun[4], Yifan Zhai[5], Zhimin Chang[6], Lixia Tian[7], Yuange Duan[1], Fan Song[1], Wanzhi Cai[1], Hu Li[1]✉ & Li Tian[1]✉

The bumblebee *Bombus patagiatus* Nylander, 1848 and *Bombus lantschouensis* Vogt, 1908 (Hymenoptera: Apidae) are ecologically important bumble bee species native to East Asia, with considerable value for agricultural pollination and domestication. Despite their ecological and economic relevance, the lack of high-quality genomic resources has hindered in-depth investigations into their genetic architecture and evolutionary adaptations. Here, we present chromosome-level genome assemblies for both species, generated using a combination of PacBio HiFi long-read sequencing, Illumina short-read resequencing, and Hi-C scaffolding. The assembled genomes span 240.28 Mb (*B. patagiatus*) and 241.30 Mb (*B. lantschouensis*), with 94.38% and 94.00% of sequences anchored to 18 chromosomes, respectively. Genome annotation identified 17,351 and 16,023 protein-coding genes in *B. patagiatus* and *B. lantschouensis*, along with comprehensive repetitive element characterization. Both assemblies exhibit exceptional completeness, with BUSCO scores exceeding 99%, confirming their high quality and reliability. These genomic resources provide a critical foundation for future research on bumble bee evolution, population genetics, and the molecular basis of domestication traits.

Background & Summary

Bumble bees (Hymenoptera: Apidae, *Bombus* Latreille, 1802) are vital pollinators in natural ecosystems[1–3], particularly in alpine and subalpine meadows[4], and are increasingly utilized for greenhouse crop pollination[5], highlighting their ecological and economic importance[3,6]. The genus *Bombus* comprises approximately 260 species across 15 subgenera[7], many of which exhibit remarkable adaptive diversity[8,9]. Among these, the subgenus *Bombus s. str*. includes 17–23 species, with about 12 native to Asia. Species in this subgenus display broad dietary breadth[10], pollinating over 37 plant families[10,11], and several—including the commercially domesticated *B. terrestris* (Europe)[5] and *B. ignitus* (Japan and Korea)[12] —are promising candidates for managed pollination.

Despite their significance, genomic resources for many *Bombus s. str*. species remain limited, particularly for East Asian lineages. Currently, only four species in this subgenus (*B. terrestris*, *B. ignitus*, *B. affinis*, and *B. lucorum*) have chromosome-level genome assemblies[13–15]. *B. patagiatus* and *B. lantschouensis*, two East Asian species with high domestication potential, exhibiting colony foundation rates exceeding 70% under artificial breeding and produce over 200 worker bees per colony[1,11]. While a draft genome of *B. lantschouensis* based on short-read sequencing has been used to study odorant receptor genes[16], the absence of high-quality, chromosome-scale references of the two species has hindered investigations into their genetic adaptations and domestication traits.

[1]Department of Entomology, State Key Laboratory of Agricultural and Forestry Biosecurity, MOA Key Lab of Pest Monitoring and Green Management, College of Plant Protection, China Agricultural University, Beijing, China. [2]International Research Center of Cross-Border Pest Management in Central Asia, Xinjiang Key Laboratory of Special Species Conservation and Regulatory Biology, College of Life Sciences, Xinjiang Normal University, Urumqi, China. [3]State Key Laboratory of Animal Biodiversity Conservation and Integrated Pest Management, Institute of Zoology, Chinese Academy of Sciences, Beijing, China. [4]College of Life Sciences, Capital Normal University, Beijing, China. [5]Present address: Institute of Plant Protection, Shandong Academy of Agricultural Sciences, Jinan, China. [6]Institute of Entomology, Provincial Special Key Laboratory for Developing and Utilizing of Insect Resources, Guizhou University, Guiyang, China. [7]Institute of Plant Protection, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China. [8]Present address: Institute of Plant Protection, Shandong Academy of Agricultural Sciences, Jinan, China. ✉e-mail: lih@cau.edu.cn; ltian@cau.edu.cn

Species	PacBio-HiFi (Gb)/ Sequencing coverage (x)	Illumina Hi-C (Gb)/ Sequencing coverage (x)	Illumina short reads (Gb)/ Sequencing coverage (x)	RNA (Gb)
B. patagiatus	13.88/56.42	21.08/89.85	10.80/42.85	21.76
B. lantschouensis	13.50/53.34	21.60/91.68	13.29/52.52	—

Table 1. Summary of sequencing data and genome coverage for *B. patagiatus* and *B. lantschouensis*.

Here, we present chromosome-level genome assemblies for *B. patagiatus* and *B. lantschouensis*, generated by integrating PacBio HiFi long-read sequencing, Illumina short-read resequencing, and Hi-C scaffolding. The *B. patagiatus* assembly spans 240.28 Mb, with 94.38% (226.75 Mb) anchored to 18 chromosomes and 17,351 protein-coding genes annotated, of which 67.35% (11,686 genes) received functional assignments. The *B. lantschouensis* assembly comprises 241.30 Mb, with 94.00% (226.82 Mb) anchored to 18 chromosomes and 16,023 protein-coding genes annotated, of which 70.21% (11,249 genes) received functional assignments. Both assemblies exhibit high completeness (BUSCO scores > 99%), ensuring their utility for downstream analyses. These high-quality genomes provide a critical foundation for studying genetic diversity, adaptive evolution, and the molecular mechanisms underlying domestication traits in bumble bees, with implications for conservation and sustainable agricultural pollination[13,14].

Methods

Sample collection and preparation. Two male *B. patagiatus* and one male *B. lantschouensis* were collected in Yu County, Hebei Province, China (39.9°N, 115.0°E) using hand-held nets. Specimens were transported alive to the MOA Key Laboratory of Pest Monitoring and Green Management (China Agricultural University), where morphological identification was performed prior to euthanasia in liquid nitrogen. To minimize contamination, abdomens were removed, and each specimen was dissected into head, thorax, and right middle leg tissues. Genomic DNA was extracted from the right middle legs using the TIANamp Genomic DNA Kit (DP304, TIANGEN). Prior to genome sequencing, species identity was molecularly confirmed through DNA barcoding of the mitochondrial COI gene. Amplification was performed using universal primers LCO1490 (5′-GGTCAACAAATCATAAAGATATTGG-3′) and HCO2198 (5′-TAAACTTCAGGGTGACCAAAAAATCA-3′)[17], followed by Sanger sequencing. Resulting sequences were verified against the NCBI nucleotide database using BLAST analysis, confirming species identification as *B. patagiatus* and *B. lantschouensis* respectively.

Genome sequencing. All sequencing was performed by GrandOmics Biosciences (Wuhan, China). For *B. patagiatus*, we utilized two male individuals: one for PacBio HiFi sequencing and another for Hi-C scaffolding. Genomic DNA was extracted from head and thoracic tissues using the cetyltrimethylammonium bromide (CTAB) method, followed by purification with the Grandomics HMW-DNA Purification Kit (XJZZ-HMW-001-3). A 10–20 kb insert library was prepared using the SMRTbell Prep Kit 3.0 and sequenced on the PacBio Sequel II platform to generate high-fidelity (HiFi) long reads. For Hi-C scaffolding, chromatin from a second individual was cross-linked with 2% formaldehyde, digested with DpnII (NEB), and processed to generate a sequencing library with 350 bp fragments, which was sequenced on Illumina NovaSeq 6000 (150 bp paired-end). Additionally, we performed Illumina short-read resequencing using DNA extracted from the right mid-legs of both specimens, with 150 bp paired-end sequencing conducted on the Illumina NovaSeq 6000 platform. For *B. lantschouensis*, a single male provided thoracic tissue for PacBio HiFi sequencing and head tissue for Hi-C library preparation, following identical protocols. Based on the estimated ~250 Mb genome size, we obtained sufficient sequencing depth for high-quality assembly (see Table 1 for sequencing metrics).

Transcriptome sequencing for genome annotation. To support genome annotation, we collected one male and one female *B. patagiatus* specimen from Miyun District, Beijing, China (40.5°N, 117.5°E). After abdomen removal to minimize contamination, total RNA was extracted from remaining tissues using TRIzol reagent (Thermo Fisher Scientific) and processed by GrandOmics Biosciences. Equal quantities of RNA from both sexes were pooled for library preparation using the TruSeq RNA Library Prep Kit (Illumina), following manufacturer protocols. Libraries were size-selected (150–200 bp) with AMPure XP beads (Beckman Coulter), amplified with Phusion High-Fidelity DNA Polymerase, and purified. Cluster generation was performed on a cBot system (TruSeq PE Cluster Kit v3-cBot-HS), followed by 150 bp paired-end sequencing on an Illumina NovaSeq 6000 platform. Sequencing metrics are provided in Table 1.

Genome assembly. We assembled the genomes of *B. patagiatus* and *B. lantschouensis* using an integrated pipeline combining PacBio HiFi, Illumina, and Hi-C data. Initial processing involved extracting HiFi reads from PacBio BAM files using SAMtools 1.15[18], while Illumina short reads and Hi-C paired-end reads were quality filtered with fastp 0.23.2[19] to remove adapters and low-quality sequences. Genome size estimation via GenomeScope2.0[20] based on k-mer frequency analysis (k = 17) using Jellyfish v2.3.0[21] yielded estimates of 271.34 Mb for *B. patagiatus* (Fig. 1A) and 270.05 Mb for *B. lantschouensis*(Fig. 1B). The initial assembly was performed with wtdbg2 v2.5[22] in CCS mode using HiFi reads, followed by contig polishing with wtpoa-cns to generate a draft assembly. We then applied Purge_dups v1.2.5[23] to remove redundant sequences, mapping HiFi reads back to the assembly using minimap2 v2.24-r1122[24] and calculating coverage statistics with pbcstat and calcuts. The resulting haploid-purged assembly underwent base-level error correction using Pilon v1.23[25] with Illumina short reads aligned via BWA-MEM v0.7.17[26].

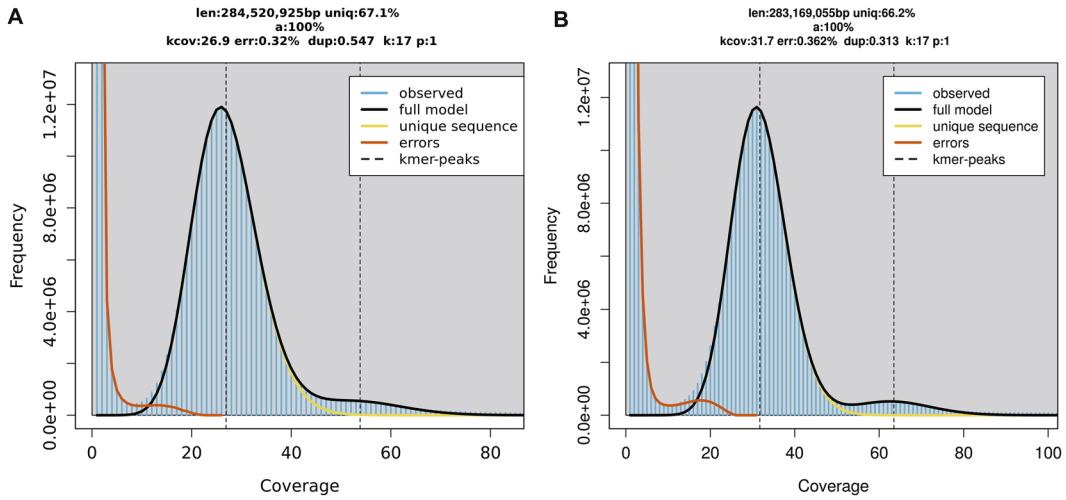

Fig. 1 Genome survey analysis of *B.patagiatus* and *B. lantschouensis* based on k-mer frequency distributions. (**A**) *B. patagiatus* and (**B**) *B. lantschouensis* 17-mer frequency profiles generated by GenomeScope, showing the relationship between coverage (x-axis) and k-mer frequency (y-axis).

For chromosome-level scaffolding, cleaned Hi-C reads[27] were processed with Juicer v1.5[28] and aligned to the polished assembly, followed by scaffolding using 3D-DNA v180114[27]. Manual review and adjustment of the assembly was performed in Juicebox Assembly Tools (https://github.com/aidenlab/Juicebox) (Fig. 2A,B), with the final chromosome-level assembly generated using run-asm- pipeline-post-review.sh. Chromosomal identity and orientation were validated by aligning the final assemblies to the *B. terrestris* reference genome using MUMmer 4.0.0rc1[29] (nucmer, delta-filter, and show-coords), with manual confirmation of chromosomal correspondence. The final *B. patagiatus* assembly spanned 240.28 Mb (GC content 37.51%) comprising 18 pseudo-chromosomes anchored by Hi-C and 485 unanchored contigs, for a total of 503 sequences. The assembly achieved a scaffold N50 of 14.32 Mb and maximum contig length of 18.93 Mb, containing 163 gaps and 155,000 ambiguous bases (Ns). Similarly, the *B. lantschouensis* assembly measured 241.30 Mb (GC content 37.35%) and consisted of 18 pseudo-chromosomes and 569 unanchored contigs (587 sequences in total), with a scaffold N50 of 15.54 Mb and a maximum scaffold length of 19.31 Mb, containing 132 gaps and 66,000 Ns.

Repeat element composition analysis. Repetitive elements were identified in both genomes using RepeatModeler v2.0.1[30] and RepeatMasker v4.1.0 (https://github.com/Dfam-consortium/RepeatMasker). For each species, we first built a custom repeat library using RepeatModeler[30] with the NCBI engine, then masked the genome using RepeatMasker with rmblast (-e rmblast). The resulting alignments were analyzed to estimate sequence divergence (calcDivergenceFromAlign.pl) and visualize repeat age distributions (createRepeatLandscape.pl). Repetitive elements constituted 19.16% (48.28 Mb) of the *B. patagiatus* genome and 19.35% (48.95 Mb) of *B. lantschouensis* (Fig. 3A,B). Interspersed repeats dominated the repeat landscape, representing 16.90% and 14.11% of each genome respectively. Among these, LTR retroelements (particularly Gypsy/DIRS1 family members) and DNA transposons (including Tc1-IS630-Pogo and PiggyBac families) emerged as the most abundant transposable element classes. Beyond the characterized transposable elements, unclassified repeats represented significant portions of both genomes, comprising 9.52% of *B. patagiatus* and 8.94% of *B. lantschouensis*. Simple sequence repeats (SSRs) showed notable interspecies variation, accounting for 1.81% and 4.77% of each genome respectively. The detailed statistics of genomic repetitive elements of two species are provided in Table 2.

Structure and function annotation of protein-coding genes. For Protein-coding gene annotation, we employed MAKER v3.01.03[31] with protein homology evidence from eight insect species (*Apis mellifera* (IBG_00083), *Apis cerana* (IBG_00079), *Bombus terrestris* (IBG_00138), *Bombus impatiens* (IBG_00128), *Bombyx mori* (IBG_00145), *Drosophila melanogaster* (IBG_00296), *Tribolium castaneum* (IBG_00768), and *Anopheles gambiae* (IBG_00050)). All related genomes were downloaded from InsectBase 2.0[32] with InsectBase ID. For *B. patagiatus*, we incorporated transcriptome-derived evidence to enhance gene model accuracy. Cleaned RNA-seq reads were aligned to the soft-masked genome using Hisat2 v2.2.1[33] and assembled into transcript models using StringTie v2.2.1[34], which were then provided to MAKER in FASTA format. No transcriptomic data were included in the annotation of *B. lantschouensis*. Gene prediction in MAKER integrated multiple evidence types, including protein alignments and transcript data (where available), supplemented by *ab initio* predictions using Augustus trained on *B. terrestris* (augustus_species = bombus_terrestris2). Key parameters included: est2genome = 1 and protein2genome = 1 to allow inference of gene models from EST and protein alignment, respectively; always_complete = 1 to ensure models contain valid start and stop codons; min_protein = 30 to filter short ORFs; pred_flank = 200 for evidence cluster extension; split_hit = 10000 for maximum intron length; and min_intron = 20. We excluded contigs < 5 kb (min_contig = 5000) while permitting single-exon genes ≥ 250 bp

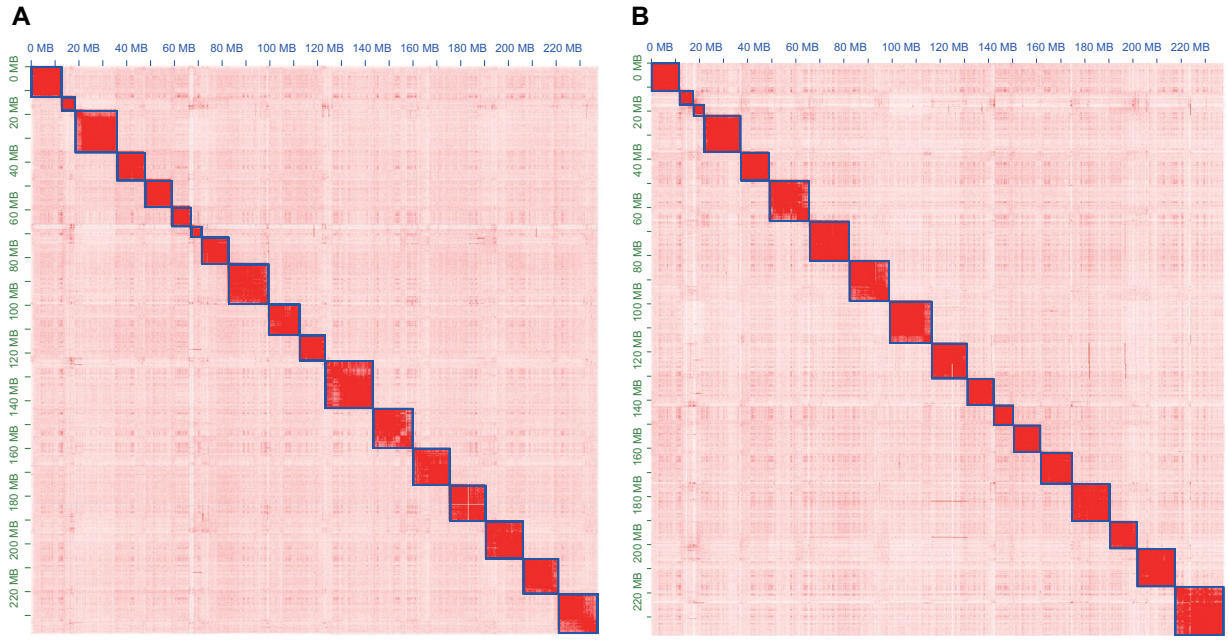

Fig. 2 Chromosome-scale Hi-C interaction maps of *B. patagiatus* and *B. lantschouensis*. (**A**) *B. patagiatus* and (**B**) *B. lantschouensis* Hi-C contact matrices demonstrating intra- and inter-chromosomal interaction frequencies. Heatmap intensity (red gradient) reflects normalized contact frequencies between genomic loci, with stronger signals indicating higher interaction probabilities. All 18 chromosomes (demarcated by dark blue borders) are shown, revealing clear spatial partitioning of chromosomal territories. The prominent diagonal patterns indicate strong intra-chromosomal interactions, confirming successful chromosome-scale scaffolding of both genomes.

(single_exon = 1, single_length = 250). After independent annotation of chromosomal and scaffold subsets, all GFF3, transcript, and protein files were merged to form the final gene sets for each species. The final annotation identified 17,351 protein-coding genes in *B. patagiatus*, spanning 65.79 Mb (27.38% of genome) with average gene length of 3,975.79 bp and 4.75 exons per gene (Fig. 4A). For *B. lantschouensis*, we annotated 16,023 genes covering 50.96 Mb (21.12% of genome) (Fig. 4B), averaging 3,335.21 bp with 4.72 exons per gene. The genome was divided into 200 Kb windows, and features such as GC content and gene density were calculated (Fig. 4 A,B). Genomic features were visualized using shinyCircos-V2.0[35].

For comprehensive functional characterization, we annotated protein-coding genes from both species using eggNOG-mapper v2.1.12[36] with the eggNOG 5.0 database. The analysis was performed in DIAMOND-based ultra-sensitive mode (--sensmode ultra-sensitive) using stringent parameters: an e-value threshold of 0.01 (--evalue 0.01), experimental evidence-only GO terms (--go_evidence experimental), and Pfam domain realignment (--pfam_realign realign). Functional assignments were obtained for 67.35% (11,686 genes) of *B. patagiatus* and 70.21% (11,249 genes) of *B. lantschouensis* protein-coding genes. Gene Ontology (GO) terms annotation covered 42.65% (7,400 genes, *B. patagiatus*) and 45.30% (7,258 genes, *B. lantschouensis*) of the respective gene sets, while KEGG pathway mapping identified 27.44% (4,761 genes, *B. patagiatus*) and 28.24% (4,525 genes, *B. lantschouensis*) of genes with metabolic pathway associations.

Data Records

The genomes of the two bumblebee species assembled in this study have been deposited in NCBI GenBank, with the accession JBSRNJ000000000.1[37] for *B. patagiatus* and JBSRNK000000000.1[38] for *B. lantschouensis*. All raw data generated in this study have been deposited in the National Genomics Data Center (NGDC), China National Center for Bioinformation / Beijing Institute of Genomics, Chinese Academy of Sciences[39] under the BioProject accession number PRJCA043262[40]. Raw sequencing data have been submitted to the Genome Sequence Archive (GSA)[41] with the accession number CRA027999[42], including PacBio HiFi, Hi-C, Illumina short reads, and transcriptome data. The corresponding identification numbers for *B. patagiatus* are CRR1999938[43] (PacBio HiFi), CRR1999939[44] (Hi-C), CRR1999940[45] (Illumina), and CRR1999941[46] (transcriptome), and for *B. lantschouensis* are CRR1999942[47] (PacBio HiFi), CRR1999943[48] (Hi-C), CRR1999944[49] (Illumina). The assembled genomes of *B. patagiatus* and *B. lantschouensis* have also deposited in the Genome Warehouse (GWH)[50] under the accession numbers GWHGEUT00000000.1[51] and GWHGEUU00000000.1[52], respectively. The annotation files for the two assembled genomes, including repetitive element annotation, gene structure annotation, and gene function annotation, are available on Figshare[53].

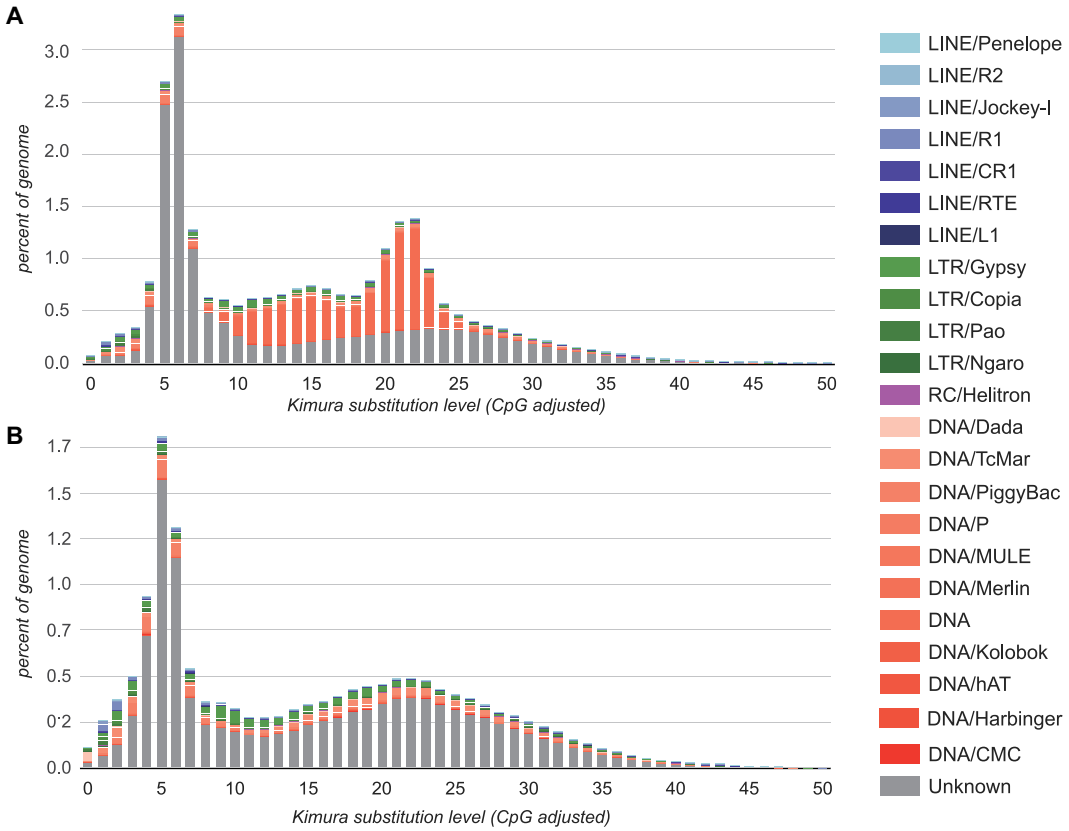

Fig. 3 Repeat element landscape analyses of *B. patagiatus* and *B. lantschouensis* genomes. (**A**) *B.patagiatus* and (**B**) *B. lantschouensis*. The plots display the relative abundance and evolutionary age distribution of transposable elements (TEs), with genomic proportions (y-axis) plotted against Kimura substitution levels (x-axis, CpG-adjusted). The distinct peak patterns reveal species-specific differences in historical transposable element activity bursts.

Technical Validation

Hi-C scaffolding refinement was performed in Juicebox by optimizing intra-chromosomal interaction signals while minimizing inter-chromosomal noise in the 3D-DNA generated.assembly files. The resulting Hi-C contact maps (Fig. 2A,B) demonstrate clear chromosomal compartmentalization, validating the accuracy of our chromosome-level assemblies.

The raw PacBio HiFi reads and Illumina short reads were re-mapped to the assembled reference genomes using minimap2 v2.24-r1122[24] and SAMtools 1.15[18], and mapping rates were calculated with SAMtools flagstat. For *B. patagiatus*, the mapping rates were 99.98% for PacBio HiFi reads and 98.65% for Illumina short reads; for *B. lantschouensis*, the rates were 99.53% and 98.28%, respectively. We also employed Merqury[54] to evaluate the assemblies using k-mer-based analysis (k = 19) derived from Illumina short reads and the assembled genomes. The results showed that *B. patagiatus* achieved a QV value of 38.42 with 98.07% completeness, while *B. lantschouensis* achieved a QV value of 36.75 with 98.31% completeness.

Assembly and annotation completeness were rigorously assessed using BUSCO v5.4.7[55] with the insecta_odb10 dataset (n = 1,367). The -m geno mode was used to assess the assembled genome, while the -m prot mode was applied to the predicted protein sets. For *B. patagiatus*, genome assembly completeness reached 99.1% (1,355 complete BUSCOs: 1,353 single-copy, 2 duplicated), while protein annotation achieved 97.4% completeness (1,332 complete BUSCOs). *B. lantschouensis* showed comparable quality, with 99.0% genome completeness (1,353 complete BUSCOs) and exceptional 99.4% protein annotation completeness (1,360 complete BUSCOs). Complete BUSCO results, involving percentage of complete, fragmented and missing BUSCOs were provided in (Fig. 5). The results confirm the high quality and completeness of both genome assemblies and their corresponding gene annotations.

Data availability

All data generated in this study have been deposited in public repositories and are publicly available. The chromosome-level genome assemblies of *Bombus patagiatus* and *Bombus lantschouensis* are available in NCBI GenBank under the accession numbers JBSRNJ000000000[37] and JBSRNK000000000[38], respectively. Raw sequencing data, including PacBio HiFi, Hi-C, Illumina short reads, and transcriptome data, have been deposited in the Genome Sequence Archive (GSA) under the accession number CRA027999[42]. The corresponding genome

Species	*B. patagiatus*			*B. lantschouensis*		
Repeat class	**Number**	**Length (bp)**	**% of Genome**	**Number**	**Length (bp)**	**% of Genome**
Retroelements	12,770	7,670,733	3.04%	11,552	7,651,549	3.02%
SINEs	96	12,937	0.01%	0	0	0.00%
Penelope	114	8,333	0.00%	205	19,372	0.01%
LINEs	6,573	3,925,586	1.56%	5,281	3,953,886	1.56%
L2/CR1/Rex	130	40,076	0.02%	126	34,577	0.01%
R1/LOA/Jockey	2,622	1,023,589	0.41%	1,665	1,064,339	0.42%
R2/R4/NeSL	95	119,795	0.05%	74	91,577	0.04%
RTE/Bov-B	278	105,473	0.04%	308	106,853	0.04%
L1/CIN4	0	0	0.00%	43	11,137	0.00%
LTR elements	6,101	3,732,210	1.48%	6,271	3,697,663	1.46%
BEL/Pao	697	435,739	0.17%	1,072	456,467	0.18%
Ty1/Copia	298	313,266	0.12%	215	322,946	0.13%
Gypsy/DIRS1	5,106	2,983,205	1.18%	4,234	2,786,346	1.10%
DNA transposons	27,680	10,921,208	4.33%	26,492	5,432,039	2.15%
hobo-Activator	3,063	396,542	0.16%	2,684	290,622	0.11%
Tc1-IS630-Pogo	16,687	2,960,904	1.18%	15,120	2,739,919	1.08%
PiggyBac	4,137	1,366,264	0.54%	4,469	1,435,551	0.57%
Tourist/Harbinger	194	33,210	0.01%	0	0	0.00%
Other	0	0	0.00%	47	9,717	0.00%
Rolling-circles	772	193,646	0.08%	619	195,533	0.08%
Unclassified	111,789	23,997,109	9.52%	109,599	22,628,277	8.94%
Total interspersed	—	42,589,050	16.90%	—	35,711,865	14.11%
Simple repeats	84,245	4,561,608	1.81%	86,508	12,074,834	4.77%
Low complexity	17,517	916,323	0.36%	17,592	923,278	0.36%
Small RNA	143	23,799	0.01%	309	41,876	0.02%

Table 2. Summary of repetitive elements in the genomes of *B. patagiatus* and *B. lantschouensis*.

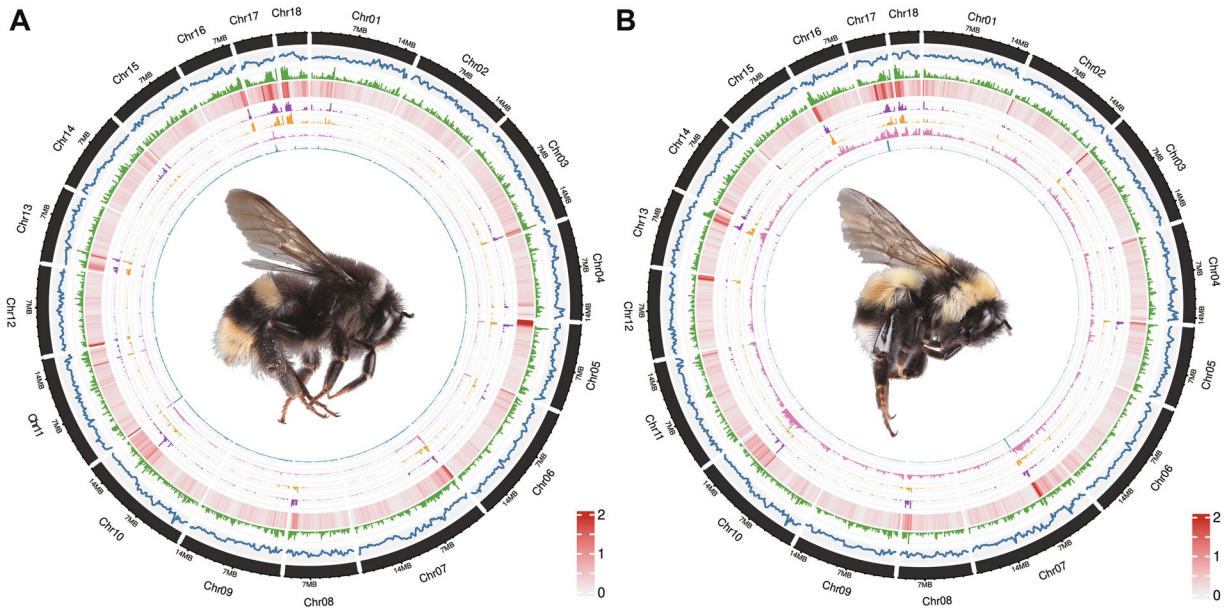

Fig. 4 Circos plots showing genomic features of *B. patagiatus* (**A**) and *B. lantschouensis* (**B**). From outer to inner tracks, each plot displays chromosome scaffolds, GC content (line plot), protein-coding gene density (bar plot), total repetitive elements (heatmap), LINE density (bar plot), LTR density (bar plot), DNA transposon density (bar plot), and simple repeat density (bar plot). Feature densities are shown as color gradients (heatmap) or bar heights, with corresponding labels positioned at the bottom right of each image. The visualization reveals distinct patterns of genome organization and repeat element distribution between the two species.

annotation files are publicly available in the Figshare repository[53]. The assembled genomes are also archived in the Genome Warehouse (GWH)[51,52].

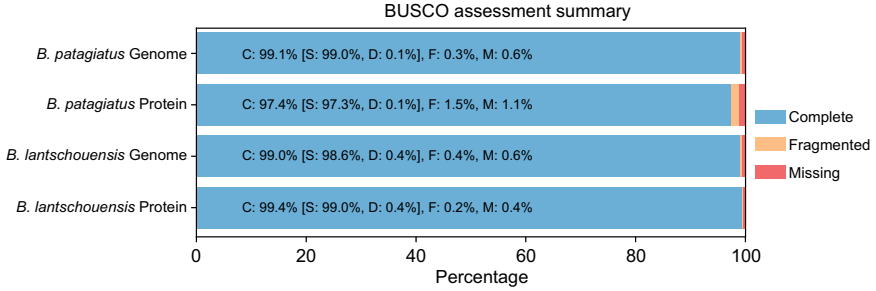

Fig. 5 BUSCO completeness assessment of *B. patagiatus* and *B. lantschouensis* genome assemblies and annotated protein-coding genes (PCGs). Evaluations were performed against the insecta_odb10 database Results demonstrate high-quality assemblies for both species, with >99% completeness for genome sequences and >97% for predicted PCGs. The stacked bar plots distinguish complete(C) [single-copy (S) and duplicated(D)], fragmented (F), and missing (M) BUSCOs.

Code availability

All analyses in this study were conducted using established bioinformatics tools available in the public domain. We exclusively employed published software packages, with each program's specific version, parameters, and implementation details comprehensively described in the respective Methods sections. No custom scripts or in-house developed software were utilized during any phase of the genome assembly, annotation, or analysis pipeline.

Received: 29 July 2025; Accepted: 6 January 2026;
Published online: 09 January 2026

References

1. Huang, J. & An, J. Species diversity, pollination application and strategy for conservation of the bumblebees of China. *Biodivers. Sci.* **26**, 486–497, https://doi.org/10.17520/biods.2018068 (2018).
2. Yuan, Y., Smagghe, G., Chen, X., Long, J. & Chang, Z. The research hotspots and frontiers of bumblebees during 1999–2024: a bibliometric analysis. *J. Appl. Entomol.* **149**, 536–548, https://doi.org/10.1111/jen.13400 (2025).
3. Cameron, S. A. & Sadd, B. M. Global trends in bumble bee health. *Annu. Rev. Entomol.* **65**, 209–232, https://doi.org/10.1146/annurev-ento-011118-111847 (2020).
4. Yu, W. B., Li, D. Z. & Wang, H. Highly efficient pollination by bumblebees ensures seed production in *Pedicularis lachnoglossa* (Orobanchaceae), an early-flowering Himalayan plant. *J. Syst. Evol.* **50**, 218–226, https://doi.org/10.1111/j.1759-6831.2012.00180.x (2012).
5. Velthuis, H. H. W. & van Doorn, A. A century of advances in bumblebee domestication and the economic and environmental aspects of its commercialization for pollination. *Apidologie* **37**, 421–451, https://doi.org/10.1051/apido:2006019 (2006).
6. Potts, S. G. *et al.* Safeguarding pollinators and their values to human well-being. *Nature* **540**, 220–229, https://doi.org/10.1038/nature20588 (2016).
7. Williams, P. H., Cameron, S. A., Hines, H. M., Cederberg, B. & Rasmont, P. A simplified subgeneric classification of the bumblebees (genus *Bombus*). *Apidologie* **39**, 46–74, https://doi.org/10.1051/apido:2007052 (2008).
8. Cameron, S. A., Hines, H. M. & Williams, P. H. A comprehensive phylogeny of the bumble bees (*Bombus*): bumble bee phylogeny. *Biol. J. Linnean Soc.* **91**, 161–188, https://doi.org/10.1111/j.1095-8312.2007.00784.x (2007).
9. Hines, H. M. Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: Apidae: *Bombus*). *Syst. Biol.* **57**, 58–75, https://doi.org/10.1080/10635150801898912 (2008).
10. Williams, P. H. Not just cryptic, but a barcode bush: PTP re-analysis of global data for the bumblebee subgenus *Bombus s. str.* Supports additional species (Apidae, genus *Bombus*). *J. Nat. Hist.* **55**, 271–282, https://doi.org/10.1080/00222933.2021.1900444 (2021).
11. An, J. *et al.* The bumblebees of North China (Apidae, *Bombus* Latreille). *Zootaxa* **3830**, 001–089, https://doi.org/10.11646/zootaxa.3830.1.1 (2014).
12. Tokoro, S., Yoneda, M., Kawate Kunitake, Y. & Goka, K. Geographic variation in mitochondrial DNA of *Bombus ignitus* (Hymenoptera: Apidae). *Appl. Entomol. Zoolog.* **45**, 77–87, https://doi.org/10.1303/aez.2010.77 (2010).
13. Sadd, B. M. *et al.* The genomes of two key bumblebee species with primitive eusocial organization. *Genome Biol.* **16**, 76, https://doi.org/10.1186/s13059-015-0623-3 (2015).
14. Sun, C. *et al.* Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits. *Mol. Biol. Evol.* **38**, 486–501, https://doi.org/10.1093/molbev/msaa240 (2021).
15. Koch, J. B. U., Sim, S. B., Scheffler, B., Geib, S. M. & Smith, T. A. Chromosome-scale genome assembly of the rusty patched bumble bee,*bombus affinis* (cresson) (hymenoptera: apidae), an endangered north american pollinator. *G3. Genes Genm. Genet.* **13**, jkad119, https://doi.org/10.1093/g3journal/jkad119 (2023).
16. Wang, Y. Identification and characteristics of odorant receptors in bumblebee. *Bombus lantschouensis. Scientia Agric. Sin.* **50**, 1904–1913, https://doi.org/10.3864/j.issn.0578-1752.2017.10.015 (2017).
17. Hebert, P. D. N., Cywinska, A., Ball, S. L. & DeWaard, J. R. Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B.* **270**, 313–321, https://doi.org/10.1098/rspb.2002.2218 (2003).
18. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079, https://doi.org/10.1093/bioinformatics/btp352 (2009).
19. Chen, S., Zhou, Y., Chen, Y. & Gu, J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890, https://doi.org/10.1093/bioinformatics/bty560 (2018).
20. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. Genomescope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432, https://doi.org/10.1038/s41467-020-14998-3 (2020).
21. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770, https://doi.org/10.1093/bioinformatics/btr011 (2011).

22. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods* **17**, 155–158, https://doi.org/10.1038/s41592-019-0669-3 (2020).
23. Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898, https://doi.org/10.1093/bioinformatics/btaa025 (2020).
24. Li, H. New strategies to improve minimap2 alignment accuracy. *Bioinformatics* **37**, 4572–4574, https://doi.org/10.1093/bioinformatics/btab705 (2021).
25. Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**, e112963, https://doi.org/10.1371/journal.pone.0112963 (2014).
26. Li, H. & Durbin, R. Fast and accurate short read alignment with burrows–wheeler transform. *Bioinformatics* **25**, 1754–1760, https://doi.org/10.1093/bioinformatics/btp324 (2009).
27. Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95, https://doi.org/10.1126/science.aal3327 (2017).
28. Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98, https://doi.org/10.1016/j.cels.2016.07.002 (2016).
29. Marçais, G. *et al.* Mummer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944, https://doi.org/10.1371/journal.pcbi.1005944 (2018).
30. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA.* **117**, 9451–9457, https://doi.org/10.1073/pnas.1921046117 (2020).
31. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491, https://doi.org/10.1186/1471-2105-12-491 (2011).
32. Mei, Y. *et al.* InsectBase 2.0: a comprehensive gene resource for insects. *Nucleic. Acids. Res.* **50**, D1040–D1045, https://doi.org/10.1093/nar/gkab1090 (2022).
33. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with hisat2 and hisat-genotype. *Nat. Biotechnol.* **37**, 907–915, https://doi.org/10.1038/s41587-019-0201-4 (2019).
34. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome. Biol.* **20**, 278, https://doi.org/10.1186/s13059-019-1910-1 (2019).
35. Wang, Y. *et al.* shinyCircos-v2.0: leveraging the creation of Circos plot with enhanced usability and advanced features. *iMeta* **2**, e109, https://doi.org/10.1002/imt2.109 (2023).
36. Cantalapiedra, C. P., Hernández-Plaza, A., Letunic, I., Bork, P. & Huerta-Cepas, J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829, https://doi.org/10.1093/molbev/msab293 (2021).
37. Cui, J. *et al. Bombus patagiatus isolate Bpat_Field-collected male 1, whole genome shotgun sequencing project. GenBank* https://identifiers.org/ncbi/insdc:JBSRNJ000000000.1 (2025).
38. Cui, J. *et al. Bombus lantschouensis isolate Blan_Field-collected male 1, whole genome shotgun sequencing project. GenBank* https://identifiers.org/ncbi/insdc:JBSRNK000000000.1 (2025).
39. CNCB-NGDC Members and Partners. Database resources of the national genomics data center, China national center for bioinformation in 2025. *Nucleic. Acids. Res.* **53**, D30–D44, https://doi.org/10.1093/nar/gkae978 (2025).
40. *NGDC/CNCB* https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA043262 (2025).
41. Chen, T. *et al.* The genome sequence archive family: toward explosive data growth and diverse data types. *Genom. Proteom. Bioinform.* **19**, 578–583, https://doi.org/10.1016/j.gpb.2021.08.001 (2021).
42. *National Genomics Data Center Genome Sequence Archive* https://ngdc.cncb.ac.cn/gsa/browse/CRA027999 (2025).
43. *National Genomics Data Center Genome Sequence Archive* https://ngdc.cncb.ac.cn/gsa/browse/CRA027999/CRR1999938 (2025).
44. *National Genomics Data Center Genome Sequence Archive* https://ngdc.cncb.ac.cn/gsa/browse/CRA027999/CRR1999939 (2025).
45. *National Genomics Data Center Genome Sequence Archive* https://ngdc.cncb.ac.cn/gsa/browse/CRA027999/CRR1999940 (2025).
46. *National Genomics Data Center Genome Sequence Archive* https://ngdc.cncb.ac.cn/gsa/browse/CRA027999/CRR1999941 (2025).
47. *National Genomics Data Center Genome Sequence Archive* https://ngdc.cncb.ac.cn/gsa/browse/CRA027999/CRR1999942 (2025).
48. *National Genomics Data Center Genome Sequence Archive* https://ngdc.cncb.ac.cn/gsa/browse/CRA027999/CRR1999943 (2025).
49. *National Genomics Data Center Genome Sequence Archive* https://ngdc.cncb.ac.cn/gsa/browse/CRA027999/CRR1999944 (2025).
50. Ma, Y. *et al.* The updated genome warehouse: enhancing data value, security, and usability to address data expansion. *Genom. Proteom. Bioinform.* **23**, qzaf10, https://doi.org/10.1093/gpbjnl/qzaf010 (2025).
51. *NGDC/CNCB* https://ngdc.cncb.ac.cn/gwh/Assembly/98333/show (2025).
52. *NGDC/CNCB* https://ngdc.cncb.ac.cn/gwh/Assembly/98334/show (2025).
53. Cui, J. *et al.* Chromosome-level genome assembly and annotation of two Asian bumble bees, *figshare. Dataset* https://doi.org/10.6084/m9.figshare.29608145 (2025).
54. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome. Biol.* **21**, 245, https://doi.org/10.1186/s13059-020-02134-9 (2020).
55. Simão, F. A., Waterhouse, R. M., Loannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212, https://doi.org/10.1093/bioinformatics/btv351 (2015).

Acknowledgements

We thank Dr. Ling Ma from China Agricultural University for assisting data analyses; This work was supported by National Natural Science Foundation of China (32370487), China National Tobacco Corporation of Science and Technology Major Projects (110202201018 [LS-02]), the project of the Northeast Asia Biodiversity Research Center (2572022DS09) and the 2115 Talent Development Program of China Agricultural University.

Author contributions

W.C., H.L. and L.T. contributed to the research design. J.C., W.Y., J.G. and L.T. collected the samples. J.C. and X.Z. performed the experiments. J.C., Z.C. and L.T. took photos of specimens. J.C., Y.X. and C.S. analyzed the data. J.C. and L.T. wrote the draft manuscript. J.C., J.L., L.T., Y.Z., Z.C., Y.D., F.S. and L.T. revised the manuscript. All authors contributed to this manuscript and agreed to the current version.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.L. or L.T.