Chromosome-Level Assembly and Annotation of the Genome of the Endangered Giant Patagonian Bumble Bee *Bombus dahlbomii*

Lican Martínez[1], Eduardo E. Zattara[1], Marina P. Arbetman[1], Carolina L. Morales[1], Rick E. Masonbrink [2], Andrew J. Severin [2,3], Marcelo A. Aizen[1], Amy L. Toth [3,4,*]

[1]Instituto de Investigaciones en Biodiversidad y Medioambiente, CONICET-Universidad Nacional del Comahue, Bariloche, Argentina
[2]Genome Informatics Facility, Iowa State University, Ames, IA, USA
[3]Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA
[4]Department of Plant Pathology, Entomology, and Microbiology, Iowa State University, Ames, IA, USA

*Corresponding author: E-mail: amytoth@iastate.edu.

Accepted: June 25, 2024

Abstract

This article describes a genome assembly and annotation for *Bombus dahlbomii*, the giant Patagonian bumble bee. DNA from a single, haploid male collected in Argentina was used for PacBio (HiFi) sequencing, and Hi-C technology was then used to map chromatin contacts. Using Juicer and manual curation, the genome was scaffolded into 18 main pseudomolecules, representing a high-quality, near chromosome-level assembly. The sequenced genome size is estimated at 265 Mb. The genome was annotated based on RNA sequencing data of another male from Argentina, and BRAKER3 produced 15,767 annotated genes. The genome and annotation show high completeness, with >95% BUSCO scores for both the genome and annotated genes (based on conserved genes from Hymenoptera). This genome provides a valuable resource for studying the biology of this iconic and endangered species, as well as for understanding the impacts of its decline and designing strategies for its preservation.

Key words: pollinator conservation, genomic resource, PacBio, *Bombus*.

Significance

The giant Patagonian bumble bee *Bombus dahlbomii*, endemic to Chile and Argentina, has experienced a massive population decline in the past 20 years. Basic genomic information is needed to help inform conservation efforts, but currently available DNA- and RNA-level data on this species are extremely limited. We provide novel genomic resources for this species through a chromosome-level genome assembly, transcriptome, and gene annotation with high completeness. This information can be a useful resource for studying the basic organismic biology, population status, and conservation of this endangered South American species.

Introduction

The giant Patagonian bumble bee *Bombus dahlbomii* (Apidae: Apinae: Bombini), also known as "mangangá" in Argentina or "moscardón" in Chile, is endemic to the southern region of South America, including Andean Patagonia, and has the most southerly natural range of any bumble bee in the world. Like other bumble bees, this species is eusocial, with distinct queen and worker

castes, and forms annual colonies founded in spring by solitary queens in underground nests (Goulson 2010). This species possesses adaptations to the harsh Patagonian climate, including long and thick red setae (Peat et al. 2005) and queens with extremely large body sizes (supplementary fig. S1a, Supplementary Material online). As the only native bumble bee in its range, it has been a key species for the pollination of many native and endemic flowers of this region (Aizen et al. 2008) and several crops (Estay et al. 2001). Populations of this native bee were historically abundant throughout much of Andean Patagonia in Chile and Argentina and farther north to the Mediterranean region of central Chile (Aizen et al. 2008; Morales et al. 2016), easily recognizable as an emblem of Patagonian nature, and integrated into myths and cultural practices of the Mapuche people of this region (Montalva et al. 2020).

The species has declined precipitously in recent years, causing significant conservation concern; the major cause may be directly tied to the spread of invasive bumble bee species. Two European bumble bee species, *Bombus ruderatus* and *Bombus terrestris*, were introduced into Chile for pollination in 1982 and 1997, respectively; *B. terrestris* is still being imported into Chile today (Aizen et al. 2019). Both European bees rapidly spread across the region, and *B. ruderatus* and *B. terrestris* reached Argentina by 1993 and 2006, respectively (Roig-Alsina and Aizen 1996; Torretta et al. 2006). Of the two introduced species, *B. terrestris* has become more highly invasive, with extremely large populations now spreading all the way down to the southernmost tip of South America (Morales 2007; Morales et al. 2013; Geslin and Morales 2015; Rendoll-Carcamo et al. 2017). Concomitantly, *B. dahlbomii* experienced a dramatic decline in several areas of Patagonia, in some cases within only 5 years of *B. terrestris*' arrival, and has become locally extinct in multiple areas of its former range (Morales et al. 2013). In 2016, *B. dahlbomii* was listed as "endangered" on the "Red List" of Species by the International Union for the Conservation of Nature, with an estimated ∼54% reduction of the species' total population in the last 10 years (Morales et al. 2016). *Bombus dahlbomii* is also part of the subgenus *Thoracobombus* (Williams et al. 2008), one of the bumble bee groups most prone to decline (Arbetman et al. 2017).

Conservation efforts are urgently needed to decelerate the decline of *B. dahlbomii* and the shrinkage of its geographical distribution. Generating a genome reference for this species can facilitate the development of modern genetic resources and enable a wide variety of applications to assist in the conservation of this species. Future population genetic studies can help elucidate whether this precipitous population decline is associated with reduced genetic diversity, especially in small or shrinking populations. This knowledge can be used to determine whether inbreeding or genetic bottlenecks are obstacles to species recovery, and if so, whether infusions of genetic diversity from different populations could help through breeding or reintroduction programs. Also, a complete and annotated genomic reference can significantly improve the quality of functional studies such as transcriptomics and genome wide association studies, allowing researchers to identify population-associated genetic variants and local adaptations that could assist species reintroduction programs. In addition, there are only four available genomes out of the over 50 species in the subgenus *Thoracobombus*; thus, this genome can inform comparative studies on whether species vulnerability is related to genic or genomic features.

Results and Discussion

Genome Assembly

Using 60,659,827 read pairs of Hi-C data aligned to the 57 contigs generated from PacBio reads using WTDBG2, automatic scaffolding was performed with HiRise by Cantata Bio (with 17 joins made) resulting in an initial assembly with 40 scaffolds, with a total assembled size of 265.32 Mb and an N50 scaffold length of 15.34 Mb (supplementary fig. S2 and table S1, Supplementary Material online). Further manual curation of the HiRise assembly using Juicer improved its contiguity, reducing the number of scaffolds from 40 to 20 scaffolds at a genome size of 265.3 Mb. Due to the lower-than-expected BUSCO scores, we investigated PacBio HiFi reads that did not map to this assembly. We reassembled these 18,018 reads into 197 scaffolds with FLYE and found some of the missing BUSCO genes. These new contigs were scaffolded into the genome using the Juicer/3D-DNA/Juicebox pipeline, resulting in 18 pseudomolecules (putative chromosomes) and two smaller scaffolds (Fig. 1a). The identified 18 chromosomes are consistent with other bumble bee species' chromosome counts (Owen et al. 1995; Sun et al. 2021; Sang et al. 2024). The final assembly has a total size of 274.05 Mb and a N50 scaffold length of 17.3 Mb. This improved assembly was subjected to BUSCO analysis, revealing a high percentage of complete and single copy conserved genes (97.4% of BUSCOs based on hymenoptera_odb10; additional assembly statistics are shown in Table 1). The improved, final assembly has been deposited as Bdahl_1.0 in NCBI at accession #JAZEWW010000000.

Annotation

RNA sequencing resulted in 165.16 Gb of read data, for an estimated 5,864× coverage of the transcriptome. The total transcriptome size was estimated at 28.17 Mb, which is approximately 10% of the total genome size (supplementary table S2, Supplementary Material online). The annotation using BRAKER 3 resulted in 15,028 annotated genes with 19,197 mRNAs. Average gene length was 6,732 bp, with a range of 200 to 330,148 bp (supplementary table S3,

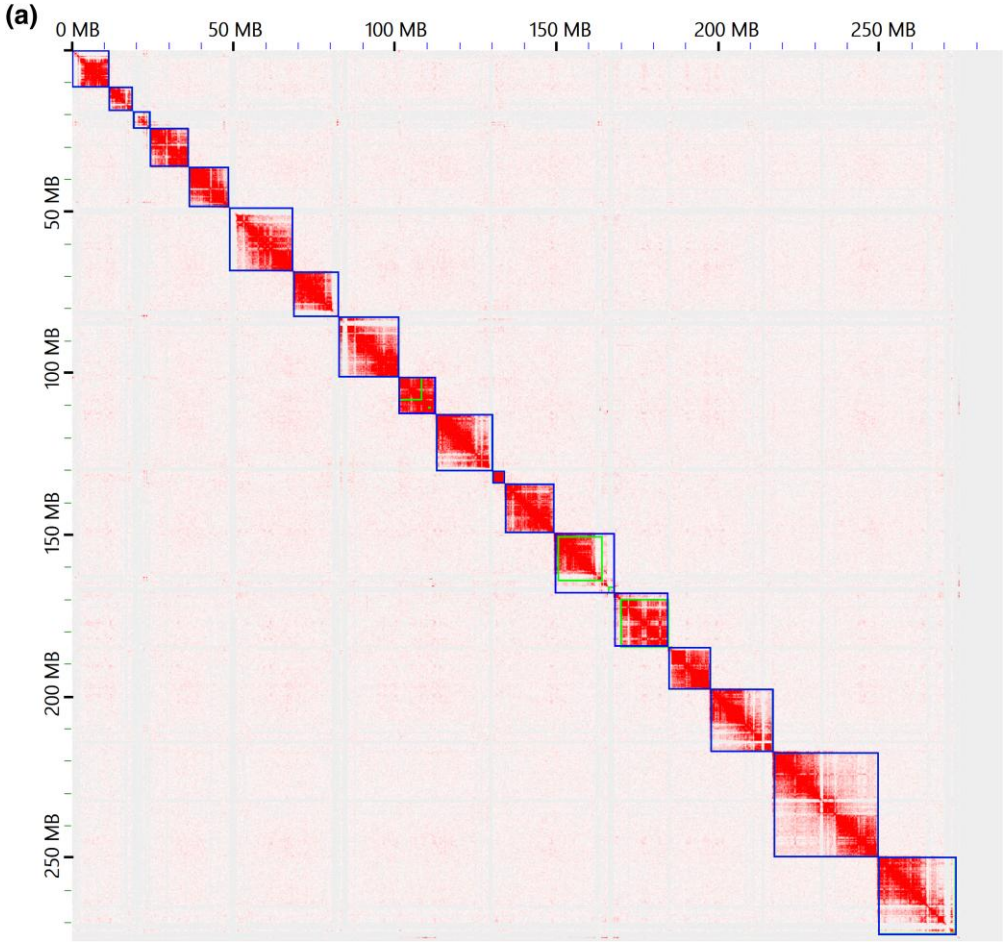

Fig. 1. a) Hi-C contact map from Hi-C read pairs in the genome assembly resulting from the final manual curation with Juicer. The *x* axis represents the read position of the first read of the read pair, and the *y* axis represents the read position of its mate, and boxes represent distinct scaffolds, with blue squares added to delimit individual chromosomes. The genome includes 20 scaffolds in the final assembly (18 chromosomes and 2 small scaffolds, not visible here), plus 12 additional scaffolds representing contaminants. b) BlobTools snail plot of the rescaffolded assembly showing length distribution of scaffolds, BUSCO scores of the final assembly, and GC/AT/N composition.

Supplementary Material online). Annotated genes had an average of 1.28 transcripts per gene, with a range from 1 to 9 transcripts (supplementary fig. S3, Supplementary Material online). We assessed the completeness of the annotation using BUSCO hymenoptera_odb10, revealing 95.1% complete BUSCO genes, of which 67.5% correspond to single hits and 27.3% to duplicated hits (supplementary table S4, Supplementary Material online). This discrepancy in duplicated BUSCO percentages between hits against the assembled genomic sequence (0.3%, Table 1) and hits against the annotated genes is expected due to the presence of alternative isoforms in the annotated gene set (supplementary fig. S3, Supplementary Material online). We functionally annotated protein coding genes via Diamond queries to NCBI NR and Swiss-Prot databases resulting in a functional annotation for 79.1% of mRNAs and 73% of genes. Among these, 33 actively transcribed transposon genes were identified in the genome.

Quality and Completeness Comparison to Other *Bombus* Genomes

The quality and completeness of the genome follow the high standards of other published chromosome-level assemblies of bumble bees (genus *Bombus*). A comparison to other chromosome-level publicly available *Bombus* genomes is presented in Table 1 and demonstrates a highly feasible number of chromosomes and genes, high completeness (BUSCOs from assembly), and very low (near chromosome level) number of scaffolds. The *B. dahlbomii* genome possesses a similar overall genome size and structure, based on numerous genome quality statistics, to other bumble bees' genomes. Overall, our genome currently

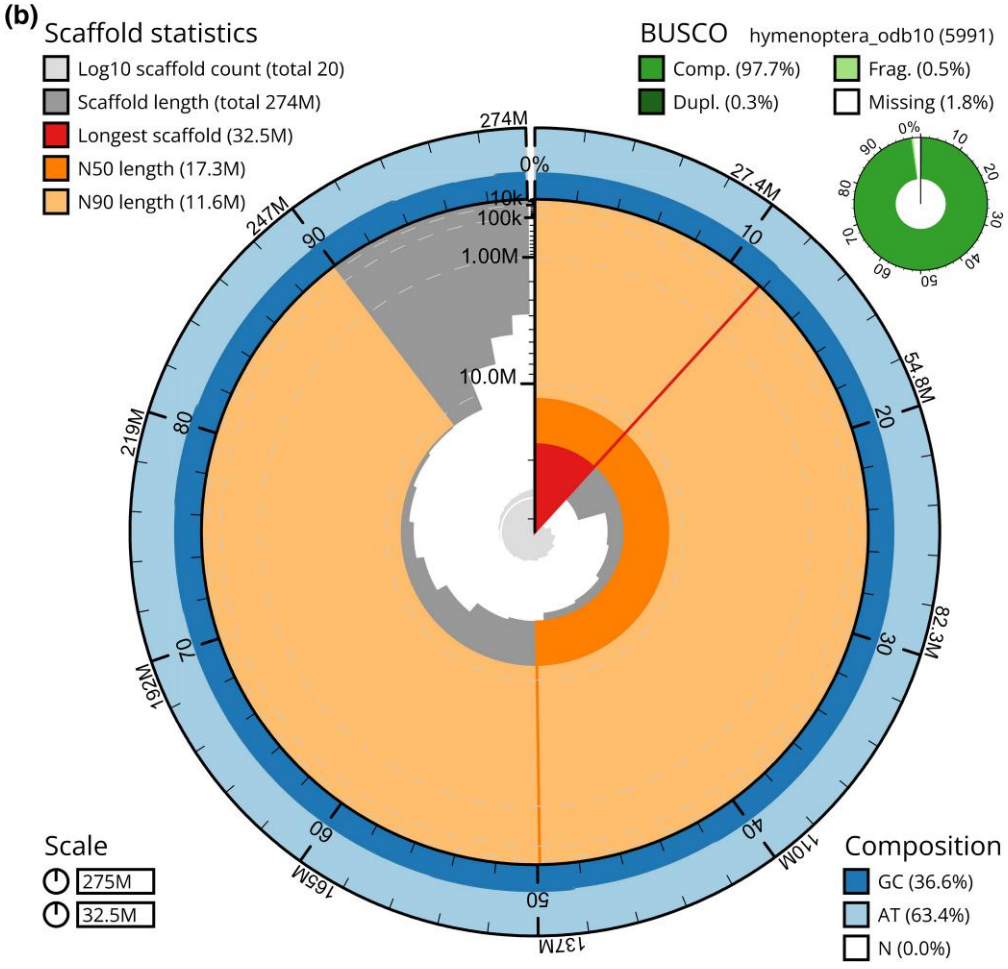

Fig. 1. Continued

stands among the best of the 36 available genomes (out of 265 species) in the genus *Bombus*, in terms of assembly quality and completeness.

Microbial Content

While invasive bumble bees have been pinpointed as a driver of decline of *B. dahlbomii*, spillover of pathogens and parasites potentially cointroduced with invasive bumble bees has been proposed as one of the causative agents (Arbetman et al. 2013; Schmid-Hempel et al. 2014; Morales et al. 2016 and references therein). Although our sequencing data are derived from only two individuals from a single population, they represent the first genomic data for this endangered species, and thus, we used the data as an opportunity to document the presence of possibly pathogenic microbes. When assembling the previously mentioned unmappable PacBio reads, we found nine bacterial scaffolds (matching the known gut symbiont *Bifidobacterium* and the nonspecific pathogen *Serratia*) and three viral scaffolds (matching a *Pantoea* phage and a *Caudoviricetes*) among our contigs (supplementary table S5, Supplementary Material online). We also used kraken2 to search the raw reads from PacBio, Hi-C, and RNA sequencing and found a wealth of microbial sequences (full results on GitHub, summary in supplementary table S6, Supplementary Material online). In particular, we scanned the reads for evidence of several common bumble bee and honey bee pathogens, parasites (Macfarlane et al. 1995), and certain gut bacteria (Hammer et al. 2021), listed in supplementary table S6, Supplementary Material online. We did not find any evidence of known bumble bee parasites; however, we did detect low read counts from some known honey bee brood pathogens, including the bacteria that cause American and European foulbrood (*Paenibacillus larvae* and *Melissococcus plutonius*), as well as fungal pathogens associated with stonebrood disease (*Aspergillus fumigatus* and *Aspergillus flavus*). In addition, we uncovered evidence of bacterial taxa that have been reported as opportunistic pathogens (*Serratia* sp. and *Hafnia* sp.), as well as the presence of Enterobacteriaceae, which if found at high levels can be signs of gut microbial

Table 1 Comparison of the *B. dahlbomii* genome (in bold text) to some other recently sequenced *Bombus* genomes (cited in references)

Species	Length (Mb)	No. of chromosomes	No. of scaffolds	Chromosome N50 (Mb)	BUSCO complete (hym_odb10)			BUSCO missing (hym_odb10) (%)	Gene total	Assembly sequencing tech	Citation
					Single (%)	Duplicated (%)	Total (%)				
Bombus dahlbomii	**265.33**	**18**	**20**	**15.8**	**97.4**	**0.3**	**97.7**	**1.8**	**15,767**	**PacBio; Illumina Hi-C**	**This study**
Bombus pyrosoma	254.8	18	4727	15.2	98.7	0.3	99.0	0.1	12,481	Illumina HiSeq	Sun et al. 2021
Bombus hortorum	296	18	43	17	97.2	0.3	97.5	2.0	No annotation	PacBio; Illumina Hi-C	Crowley et al. 2021
Bombus campestris	275.3	25	92	12.2	97.4	0.3	97.7	1.9	18,321	PacBio; Illumina Hi-C	Crowley et al. 2023a
Bombus pascuorum	307.5	17	81	17.6	98.7	0.4	99.1	1.8	18,442	PacBio; Illumina Hi-C	Crowley et al. 2023b
Bombus pratorum	285.1	18	50	16.5	97.4	0.3	97.7	1.9	19,424	PacBio; Illumina Hi-C	Crowley et al. 2023c
Bombus sylvestris	302.6	24	140	12.4	97.2	0.3	97.5	2.0	18,823	PacBio; Illumina Hi-C	Crowley et al. 2023d
Bombus terrestris	393	18	249	14.6	98.5	0.4	99.9	2.0	19,720	PacBio; Illumina Hi-C	Crowley et al. 2023e
Bombus hypnorum	297.3	12	53	24.3	97.4	0.2	97.6	1.9	19,634	PacBio; Illumina Hi-C	Crowley et al. 2023f
Bombus affinis	365.12	18	858	12.3	98.7	0.4	99.1	0.9	15,252	PacBio	Koch et AL. 2023
Bombus huntii	317.4	18	225	17.5	98.4	0.3	98.7	1.2	15,072	PacBio Sequel	Koch et al. 2024
Range	**254 to 393**	**12 to 25**	**20 to 4,727**	**12.2 to 24.3**	**97.2 to 98.7**	**0.2 to 0.4**	**97.5 to 99.1**	**0.1 to 2.0**	**12,481 to 29,720**	…	…

dysregulation (Hammer et al. 2021). We note that many of these microbes are commonly carried by and detected in healthy bees; thus, our data cannot be interpreted to indicate acute infection or disease in either of the two male *B. dahlbomii* we sequenced for this project.

Conclusion

This study presents state-of-the-art genomic resources for an endangered bumble bee species, *B. dahlbomii*. We capitalized on third-generation sequencing platforms, integrating long reads of genomic DNA using PacBio HiFi sequencing, with Hi-C chromosome conformation capture technology, bolstered by high-depth short-read RNA sequencing data. This has resulted in a chromosome-level genome assembly and highly complete gene annotation for this species. This valuable resource can be used to inform future evolutionary, genetic, and physiological studies on this species and its taxonomic relatives. Given that there was no prior genomic information available for this species, this represents a major leap forward in terms of resource development to inform the basic biology, population status, and conservation of this endangered South American species. We encourage the use of this resource to assess genetic diversity (modern and historical) and population structure, in order to inform future population management and possible reintroduction efforts.

Materials and Methods

Sample Collections and Sequencing

For genome sequencing, a single male (haploid) was collected at Puerto Blest, Parque Nacional Nahuel Huapi, in Río Negro Province, Argentina, on 2021 April 13 (supplementary fig. S1b, Supplementary Material online). The male was placed into 96% ethanol and exported for DNA isolation and sequencing at Cantata Bio, LLC, in Scotts Valley, California (formerly Dovetail Genomics). DNA was extracted from whole tissue using a Qiagen DNA Extraction Kit (manufacturer's protocol). A sequencing library was prepared using a PacBio Sequencing Library Prep Kit (SMRTbell Express Template Prep Kit 2.0, manufacturer's protocol). The library was further processed using a PacBio Sequel II Binding Kit 2.0 and sequenced on the PacBio Sequel II platform in 8M SMRT cells.

At Cantata Bio, to generate each Dovetail Omni-C library, chromatin was fixed in place with formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DNAse I, chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter-containing ends, crosslinks were reversed, and purified DNA was treated to remove biotin not internal to ligated fragments. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The library was sequenced on an Illumina HiSeqX platform to an approximate depth of 30× coverage.

For RNA sequencing, a single male was collected at Puerto Blest, Parque Nacional Nahuel Huapi on 2022 March 12, cut into sections, stored in RNA-later, and placed in a −20 °C freezer. The sample was exported to Iowa State University (ISU) at room temperature, and RNA was extracted from thoracic tissue using a Qiagen RNeasy Kit with DNAse digestion. RNA was checked for quality and quantity using a Bioanalyzer electrophoretic system (Agilent Technologies), and libraries were prepared at the ISU DNA Facility using a standard Illumina TruSeq mRNA library prep kit. The library was sequenced on one flow cell using an Illumina NovaSeq 6000 at the ISU DNA Facility as paired-end 150 bp reads, generating >500 M reads.

This work complies with the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization to the Convention on Biological Diversity. All *B. dahlbomii* samples were collected and exported with permission from the Argentine government, under permit "Autorización de investigación DRPN 1615, IF-2019-42089751-APN-DRPN#APNAC, República Argentina, Poder Ejecutivo Nacional, Parques Nacionales" issued to A.L.T., C.L.M., M.A.A., E.E.Z., and M.P.A. Samples were legally imported via international shipment to the United States.

Bioinformatic Analysis

Code used for all bioinformatic analyses is provided at https://github.com/ISUgenomics/B_dahlbomii_Genome_And_Annotation. An initial de novo assembly was performed by Cantata Bio (formerly Dovetail LLC) based on 198.6 Gb of PacBio CLR reads, which were used as input for WTDBG2 v2.5 (Ruan and Li 2020). Parameters for WTDBG2 were set as genome size 0.2 g, minimum read length 20,000, and minimum alignment length 8,192. In addition, realignment was enabled with the -R option and read type was set with the option -x sq. This assembly was then used as input for the HiRise pipeline together with Omni-C library reads with MQ > 50 (Putnam et al. 2016). Omni-C library sequences were aligned to the draft input assembly using BWA (Li 2013). The separations of Omni-C read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a threshold.

Additional scaffolding was conducted after receiving the *B. dahlbomii* initial assembly from Cantata Bio. Hi-C reads were extracted from bam files using Picard 2.17.0 (Picard Toolkit 2018), which were aligned to scaffolds with BWA 0.7.17 (Li 2013) and processed with Juicer 1.5.7, 3D-DNA

180114, and Juicebox 1.11 (Durand et al. 2016). Assembly quality was assessed using Busco 5.12 (Simão et al. 2015) with eukaryota_odb10 and hymenoptera_odb10 data sets. Due to the lower-than-expected BUSCO scores for such a contiguous assembly (hymenoptera_odb10 C: 91.9% [S: 91.6%, D: 0.3%], F: 0.2%, M: 7.9%, n: 5991 and eukaryota_odb10 C: 89.7% [S: 65.7%, D: 24.0%], F: 1.4%, M: 8.9%, n: 954), we used Minimap2 (2.1) (Li 2018) to identify PacBio HiFi reads that did not map to this genome assembly. While 97.8% of these reads mapped to the genome, the 2.2% that did not were reassembled with Flye 2.9-b1768. This resulted in an assembly of 197 scaffolds at a length of ~9.7 Mb and an N50 of 68.8 kb. We identified that 185 of these scaffolds matched known *Bombus* sequences using BLASTn 2.13.0 (Altschul et al. 1990) searches against the NCBI nr/nt database (downloaded 2023 January 1). All reassembled contigs were scaffolded into the assembly using the aforementioned BWA, Juicer, 3D-DNA, and Juicebox pipeline, resulting in an assembly of 32 scaffolds at 275.14 Mb with an N50 of 17.3 Mb. We inspected these scaffolds with Blobtools2, using Minimap2 (2.1) mapping of PacBio reads and BLASTn to NCBI nucleotide database. We removed 12 scaffolds of bacterial origin, resulting in an assembly of 20 scaffolds at 274.05 Mb with an N50 of 17.3 Mb.

To annotate *B. dahlbomii* genes, repeats were identified in the genome using RepeatModeler 2.0.2 (Flynn et al. 2020) and softmasked with RepeatMasker 4.1.2-p1 (Chen 2004). RNA-seq reads were aligned to the genome using STAR 2.7.6 (Dobin et al. 2013), with bam files subsequently merged, sorted, and indexed using Samtools 1.14 (Li et al. 2009). Using an Augustus 3.4 (Hoff and Stanke 2019) model trained with BUSCO 5.12 on hymenoptera_odb10, we annotated genes in the genome with Braker 3.0.2 (Hoff et al. 2016) and GeneMark 4.69 (Lomsadze et al. 2014). Genes were functionally annotated by querying predicted proteins with Diamond 2.0.25 (Buchfink et al. 2021) to the NCBI NR database (downloaded 2021 May 3) and Swiss-Prot database (downloaded 2023 January 23).

We investigated the genomic reads (PacBio HiFi and Hi-C) and transcriptomic reads (Illumina RNA-seq) for microbial content, including potential pathogens and parasites. To do this, we queried the taxonomy of each read using a readily available Kraken2 (Wood et al. 2019) database created (2021 August 4) from standard libraries of viral, bacterial, fungal, protozoan, and nematodal sequences (the database also included the *Heterodera glycines* genome GCA_004148225.2, not relevant to this project; thus, any *H. glycines* hits were discounted).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Funding

This work was supported by a Fulbright Science and Technology Award, Argentina Commission (to A.L.T.), Iowa State University Faculty Professional Development Leave (to A.L.T.), and the National Geographic Society NGS-57001R-19 and NGS-97579C-22 (to M.P.A.).

Data Availability

The final genome assembly has been deposited at DDBJ/ENA/GenBank under the accession JAZEWW000000000. The version described in this paper is version JAZEWW010000000. HiFi PacBio, Hi-C Illumina, and RNA-seq reads have been deposited in SRA under accessions SRR28013339, SRR27026714, and SRR28005379, respectively. The predicted gene sequences have been deposited at NCBI under accession GKSC00000000. All scripts used to assemble and evaluate the genomes can be found at https://github.com/ISUgenomics/B_dahlbomii_Genome_And_Annotation.

Literature Cited

Aizen MA, Morales CL, Morales JM. Invasive mutualists erode native pollination webs. PLoS Biol. 2008:6(2):e31. https://doi.org/10.1371/journal.pbio.0060031.

Aizen MA, Smith-Ramírez C, Morales CL, Vieli L, Sáez A, Barahona-Segovia RM, Arbetman MP, Montalva J, Garibaldi LA, Inouye DW, et al. Coordinated species importation policies are needed to reduce serious invasions globally: the case of alien bumblebees in South America. J Appl Ecol. 2019:56(1):100–106. https://doi.org/10.1111/1365-2664.13121.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990:215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Arbetman MP, Gleiser G, Morales CL, Williams P, Aizen MA. Global decline of bumblebees is phylogenetically structured and inversely related to species range size and pathogen incidence. Proc R Soc B Biol Sci. 2017:284(1859):20170204. https://doi.org/10.1098/rspb.2017.0204.

Arbetman MP, Meeus I, Morales CL, Aizen MA, Smagghe G. Alien parasite hitchhikes to Patagonia on invasive bumblebee. Biol Invasions. 2013:15:489–494.

Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021:18(4):4. https://doi.org/10.1038/s41592-021-01101-x.

Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinform. 2004:5(1):4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s05.

Crowley LM, University of Oxford; Wytham Woods Genome Acquisition Lab; Wellcome Sanger Institute Tree of Life; and Darwin Tree of Life Consortium. The genome sequence of the garden bumblebee, *Bombus hortorum* (Linnaeus, 1761). Wellcome Open Res. 2021:6:270. https://doi.org/10.12688/wellcomeopenres.17187.1.

Crowley LM, University of Oxford; Wytham Woods Genome Acquisition Lab; Wellcome Sanger Institute Tree of Life; and Darwin Tree of Life Consortium. The genome sequence of the Field Cuckoo-bee, *Bombus campestris* (Panzer, 1801). Wellcome Open Res. 2023a:8:77. https://doi.org/10.12688/wellcomeopenres.18985.1.

Crowley LM, University of Oxford; Wellcome Sanger Institute Tree of Life; and Darwin Tree of Life Consortium. The genome sequence of the Forest Cuckoo Bee, *Bombus sylvestris* (Lepeletier, 1832). Wellcome Open Res. 2023d;8:78. https://doi.org/10.12688/wellcomeopenres.18986.1.

Crowley LM, Sivell O, Darwin Tree of Life Consortium. The genome sequence of the tree bumblebee, *Bombus hypnorum* (Linnaeus, 1758). Wellcome Open Res. 2023f;8(21):21. https://doi.org/10.12688/wellcomeopenres.18778.1.

Crowley LM, Sivell O, Sivell D, University of Oxford; Wytham Woods Genome Acquisition Lab; Natural History Museum Genome Acquisition Lab; Wellcome Sanger Institute Tree of Life; and Darwin Tree of Life Consortium. The genome sequence of the Common Carder Bee, *Bombus pascuorum* (Scopoli, 1763). Wellcome Open Res. 2023b:8:142. https://doi.org/10.12688/wellcomeopenres.19251.1.

Crowley LM, Sivell O, Sivell D, University of Oxford; Wytham Woods Genome Acquisition Lab; Natural History Museum Genome Acquisition Lab; Wellcome Sanger Institute Tree of Life; and Darwin Tree of Life Consortium. The genome sequence of the Early Bumblebee, *Bombus pratorum* (Linnaeus, 1761). Wellcome Open Res. 2023c;8:143. https://doi.org/10.12688/wellcomeopenres.19250.1.

Crowley LM, Sivell O, Sivell D, University of Oxford; Wytham Woods Genome Acquisition Lab; Natural History Museum Genome Acquisition Lab; Wellcome Sanger Institute Tree of Life; and Darwin Tree of Life Consortium. The genome sequence of the Buff-tailed Bumblebee, *Bombus terrestris* (Linnaeus, 1758). Wellcome Open Res. 2023e;8:161. https://doi.org/10.12688/wellcomeopenres.19248.1.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013:29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. Cell Syst. 2016:3(1):99–101. https://doi.org/10.1016/j.cels.2015.07.012.

Estay PP, Wagner VA, Escaff GM. Evaluación de *Bombus dahlbomii* (Guér.) como agente polinizador de flores de tomate (*Lycopersicon esculentum* (Mill)), bajo condiciones de invernadero. Agricultura Técnica. 2001:61(2):113–119. https://doi.org/10.4067/S0365-28072001000200001.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020:117(17):9451–9457. https://doi.org/10.1073/pnas.1921046117.

Geslin B, Morales C. New records reveal rapid geographic expansion of *Bombus terrestris* Linnaeus, 1758 (Hymenoptera: Apidae), an invasive species in Argentina. Check List. 2015:11(3):Article 3. https://doi.org/10.15560/11.3.1620.

Goulson D. Bumblebees: behaviour, ecology, and conservation. 2nd ed. Oxford, UK: Oxford University Press; 2010.

Hammer TJ, Le E, Martin AN, Moran NA. The gut microbiota of bumblebees. Insectes Soc. 2021:68(4):287–301. https://doi.org/10.1007/s00040-021-00837-1.

Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. BRAKER1: unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. Bioinformatics. 2016:32(5):767–769. https://doi.org/10.1093/bioinformatics/btv661.

Hoff KJ, Stanke M. Predicting genes in single genomes with AUGUSTUS. Curr Protoc Bioinform. 2019:65(1):e57. https://doi.org/10.1002/cpbi.57.

Koch JB, Sim SB, Scheffler B, Geib SM, Smith TA. Chromosome-scale genome assembly of the rusty patched bumble bee, *Bombus affinis* (Cresson) (Hymenoptera: Apidae), an endangered North American pollinator. G3 (Bethesda). 2023:13(8):jkad119. https://doi.org/10.1093/g3journal/jkad119.

Koch JB, Sim SB, Scheffler B, Lozier JD, Geib SM. Chromosome-scale genome assembly of the Hunt bumble bee, *Bombus huntii* Greene, 1860, a species of agricultural interest. bioRxiv 591905. https://doi.org/10.1101/2024.04.30.591905, 2 May 2024, preprint: not peer reviewed.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997, preprint: not peer reviewed.

Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:34(18):3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. Bioinformatics (Oxford England). 2009:25(16):2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Lomsadze A, Burns PD, Borodovsky M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucl Acids Res. 2014:42(15):e119. https://doi.org/10.1093/nar/gku557.

Macfarlane RP, Lipa JJ, Liu HJ. Bumble bee pathogens and internal enemies. Bee World. 1995:76(3):130–148. https://doi.org/10.1080/0005772X.1995.11099259.

Montalva J, Dudley LS, Sepúlveda JE, Smith C. The giant bumble bee (*Bombus dahlbomii*) in Mapuche cosmovision. Ethnoentomology. 2020:4(1):1.

Morales C, Montalva J, Arbetman M, Aizen MA, Smith-Ramírez C, Vieli L, Hatfield R. Bombus dahlbomii. The IUCN red list of threatened species 2016: e.T21215142A100240441. https://dx.doi.org/10.2305/IUCN.UK.2016-3.RLTS.T21215142A100240441.en.

Morales CL. Introducción de abejorros (*Bombus*) no nativos: causas, consecuencias ecológicas y perspectivas. Ecología Austral. 2007:17(1):51–65.

Morales CL, Arbetman MP, Cameron SA, Aizen MA. Rapid ecological replacement of a native bumble bee by invasive species. Front Ecol Environ. 2013:11(10):529–534. https://doi.org/10.1890/120321.

Owen RE, Richards KW, Wilkes A. Chromosome numbers and karyotypic variation in bumble bees (Hymenoptera: Apidae; Bombini). J Kansas Entomol Soc. 1995:68(3):290–302.

Peat J, Darvill B, Ellis J, Goulson D. Effects of climate on intra- and interspecific size variation in bumble-bees. Funct Ecol. 2005:19(1):145–151. https://doi.org/10.1111/j.0269-8463.2005.00946.x.

Picard Toolkit. [Java]. Cambridge (MA): Broad Institute; 2018. https://broadinstitute.github.io/picard/.

Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, et al. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 2016:26(3):342–350. https://doi.org/10.1101/gr.193474.115.

Rendoll-Carcamo JA, Contador TA, Saavedra L, Montalva J. First record of the invasive bumblebee *Bombus terrestris* (Hymenoptera: Apidae) on Navarino Island, southern Chile (55 S). J Melittol. 2017:71(7 July):1–5. https://doi.org/10.17161/jom.v0i71.6520.

Roig-Alsina A, Aizen MA. *Bombus ruderatus* Fabricius, un nuevo *Bombus* para la Argentina (Hymenoptera: Apidae). Physis. 1996:5:49–50.

Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods 2020;17(2):155–158. https://doi.org/10.1038/s41592-019-0669-3.

Sang H, Li Y, Tan S, Gao P, Wang B, Guo S, Luo S, Sun C. Conservation genomics analysis reveals recent population decline and possible

causes in bumblebee *Bombus opulentus*. Insect Sci. 2024:0:1–15. https://doi.org/10.1111/1744-7917.13324.
Schmid-Hempel R, Eckhardt M, Goulson D, Heinzmann D, Lange C, Plischuk S, Escudero LR, Salathé R, Scriven JJ, Schmid-Hempel P. The invasion of southern South America by imported bumblebees and associated parasites. J Anim Ecol. 2014:83(4):823–837. https://doi.org/10.1111/1365-2656.12185.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:31(19):3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
Sun C, Huang J, Wang Y, Zhao X, Su L, Thomas GWC, Zhao M, Zhang X, Jungreis I, Kellis M, et al. Genus-wide characterization of bumblebee genomes provides insights into their evolution and variation in ecological and behavioral traits. Mol Biol Evol. 2021:38(2):486–501. https://doi.org/10.1093/molbev/msaa240.
Torretta JP, Medan D, Abrahamovich AH. First record of the invasive bumblebee *Bombus terrestris* (L.) (Hymenoptera, Apidae) in Argentina. Trans Am Entomol Soc. 2006:132(3):285–289. https://doi.org/10.3157/0002-8320(2006)132[285:FROTIB]2.0.CO; 2.
Williams PH, Cameron SA, Hines HM, Cederberg B, Rasmont P. A simplified subgeneric classification of the bumblebees (genus *Bombus*). Apidologie (Celle). 2008:39(1):46–74. https://doi.org/10.1051/apido:2007052.
Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019:20(1):257. https://doi.org/10.1186/s13059-019-1891-0.

Associate editor: John Wang