Minimal tissue inputs produce a chromosome-scale genome assembly of the rusty patched bumble bee, an endangered North American pollinator

Running Title: Genome assembly of *Bombus affinis*

Jonathan Berenguer Uhuad Koch[1*], Sheina B. Sim[2], Brian Scheffler[3], Scott M. Geib[2], Tamara A. Smith[4]

[1]U.S. Department of Agriculture, Agricultural Research Service, Pollinating Insect Research Unit, 1410 N 800 E, Logan, UT 84341, USA

[2]U.S. Department of Agriculture, Agricultural Research Service, U.S. Pacific Basin Agricultural Research Center, Tropical Pest Genetics and Molecular Biology Research Unit, 64 Nowelo Street, Hilo, HI 96720, USA

[3]U.S. Department of Agriculture, Agricultural Research Service, Jamie Whitten Delta States Research Center, Genomics and Bioinformatics Research Unit, 141 Experiment Station Road, Stoneville, MS 38776, USA

[4]U.S. Fish and Wildlife Service, Minnesota/Wisconsin Ecological Services Field Office, 3815 American Boulevard East, Bloomington, MN 55425

* Correspondence: jonathan.koch@usda.gov

Abstract

The rusty patched bumble bee, *Bombus affinis*, is an important pollinator in North America and a federally listed endangered species. Due to habitat loss and large declines in population size, *B. affinis* is facing imminent extinction unless human intervention and recovery efforts are implemented. To better understand *B. affinis* biology and population genetic and genomic landscapes, we sequenced and assembled the *B. affinis* genome from a single male. Whole genome HiFi sequencing on PacBio coupled with HiC sequencing resulted in a complete and highly contiguous contig assembly that was scaffolded into a chromosomal context, resolving 18 chromosomes for this species. All material for both HiFi and HiC sequencing was derived from a single abdominal tissue segment from the one male. These assembly results, coupled with the minimal amount of tissue destructively sampled, demonstrates methods for generating contiguous and complete genomic resources for a rare and endangered species with limited material available and highlights the importance of sample preservation. Precise methods and applications of these methods are presented for potential applications in other species with similar limitations in specimen availability and curation considerations.

Key words: bees, pollinator, biodiversity, conservation, insect decline, genome assembly

Introduction

The rusty patched bumble bee, *Bombus affinis* (Cresson), was once an abundant and widespread pollinator in Canada and the U.S. (Figure 1) (Colla and Packer 2008; Cameron *et al.* 2011). Like its close relatives in the subgenus *Bombus* of North America, *B. affinis* underwent significant population decline and range collapse throughout its known historic distribution over the past several decades. Hypotheses proposed on the cause of their decline include the transmission of pathogens, specifically *Varimorpha bombi* (=*Nosema bombi*) (Tokarev *et al.*

2020), from managed to wild bumble bee populations (Cameron *et al.* 2011, 2016), habitat loss and degradation (Mola *et al.* 2021), small population biology, and climate change (Soroye *et al.* 2020). Compounding evidence for their decline at multiple spatial scales prompted the Committee on the Status of Endangered Wildlife in Canada (COSEWIC) and the United States (U.S.) Fish and Wildlife Service (USFWS) to list *B. affinis* as "Endangered" in 2010 and 2017, respectively (COSEWIC 2010; U.S. Fish and Wildlife Service 2017). This action enables independent *B. affinis* recovery plans to be developed in Canada and the U.S.

Historically, *B. affinis* has been recorded from coastal regions of Maine, U.S. to the western Dakotas, U.S. Its historic latitudinal distribution includes Georgia, U.S. in the south, and Ontario and Quebec, Canada in the north (Milliron 1971). Like most non-cleptoparasitic and non-tropical bumble bees, *B. affinis* forms annual colonies that begin with a gyne (i.e., primary reproductive female caste) emerging in the spring to establish a nest. Once the nest is established, the gyne (now queen) will produce workers (i.e., non-reproductive female caste) who will eventually take over a diversity of activities including foraging and nest defense. Colony sizes have reported to be large as 1341 bees and small as 251 bees (Boone *et al.* 2022). In the late spring, the queen will produce sexuals (males and gynes), who will go on mating flights. Mated gynes will then overwinter in a hibernaculum (i.e., overwinter refuge) until the following spring. Workers and males pass away at the conclusion of the colony cycle in the late summer/early autumn. Over the course of both the queen's and colony lifespan, a diversity of challenges may arise including limited or contaminated floral resources, unfavorable overwintering and foraging conditions, pathogens, parasites, and natural enemies. Despite these challenges, there is evidence to suggest that *B. affinis* can persist in human built environments, where human-modified

65 landscapes may account for between 75% - 100% of *B. affinis* nesting, overwintering, and
66 foraging habitat (Boone *et al.* 2022).

67 *Bombus affinis* is estimated to have declined by 87% across their geographic distribution
68 (Cameron *et al.* 2011). From a collection of 16,788 bumble bees collected across 3 years at 382
69 field sites, *B. affinis* was only represented by 22 individuals, or 0.13% of the total collected
70 specimens, emphasizing their rarity across the contiguous U.S (Cameron *et al.* 2011; Koch *et al.*
71 2015). These results support the hypothesis that *B. affinis* will be facing imminent extinction
72 unless management and conservation actions are applied to support the species in diverse habitats
73 (Canada 2020; Smith *et al.* 2020; U.S. Fish and Wildlife Service 2021). A recent study of *B.*
74 *affinis* corbicular pollen loads preserved in natural history collections demonstrates no evidence
75 of significant change in the plant species they consumed over the last 100 years. The results
76 suggest that changes or declines in floral resources do not necessarily precede their decline
77 (Simanonok *et al.* 2021; Mola *et al.* 2021). Conversely, the fungal pathogen, *V. bombi*, has been
78 found to increase in detection frequency in the mid-1990s, corresponding with wild population
79 decline (Colla and Packer 2008) and the advent of infectious outbreaks in commercial bumble
80 bee rearing stocks (Cameron *et al.* 2016). However, there is no evidence of a novel *V. bombi*
81 genotype to be associated with the decline of *B. affinis* or other North American bumble bees.
82 Focused monitoring of extant *B. affinis* populations are currently underway across federal,
83 provincial, state, and city governments in Canada and the U.S (Otto *et al.* 2022).

84 To support the genetic assessment of *B. affinis* populations, we present the first high-
85 quality, chromosome-scale genome assembly of the species from a single wild caught male in
86 Minnesota, U.S. in 2020 and provide high quality gene annotations of this genome. The rationale
87 for this approach is to retain vouchered tissue of the specimen that will support downstream

taxonomic diagnoses using both molecular and morphological techniques. We implemented the best practices developed by the U.S. Department of Agriculture Ag100Pest (Childers *et al.* 2021) and the Beenome100 (https://www.beenome100.org/) initiatives to achieve assembly and high fidelity using Pacific Bioscience (PacBio) HiFi (circular consensus) reads.

Materials and Methods

Specimen sampling. *B. affinis* specimens used to develop the reference genome were collected by Dr. Elaine Evans of the University of Minnesota from an active colony in a residential area from Redwing, Goodhue County, Minnesota, U.S on 04 August 2020. The coordinates of the *B. affinis* colony are rounded to the 100th's place to maintain anonymity of the private homeowners (Latitude: 44.56, Longitude: -92.53, 236 m). Lethal collection of the specimen was conducted under the authority of the USFWS Endangered Species Permit sub-permit (16-07-3a). In total, six specimens were flash frozen in liquid nitrogen and maintained at -80 C until they were shipped to the U.S. Department of Agriculture - Agricultural Research Service (USDA-ARS) - Pollinating Insect Research Unit (PIRU) in Logan, UT.

DNA extraction, library preparation, and sequencing. Two *B. affinis* males were sent to the USDA-ARS Tropical Pest Genetics and Molecular Biology Research Unit in Hilo, HI to undergo DNA extraction and PacBio HiFi and HiC library preparation. Genomic DNA was extracted from a slice of abdominal tissue from a single *B. affinis* male (ToLID iyBomAffi1). The fresh or frozen tissue protocol of the Qiagen MagAttract HMW DNA Kit (Qiagen, Hilden Germany) was followed to obtain DNA that was sufficiently of high-molecular weight. Following isolation, genomic DNA was subjected to a 2.0x bead clean-up to improve sample purity and quantified using the dsDNA Broad Range (BR) Qubit assay (Thermo Fisher Scientific, Waltham, MA, USA) and the fluorometer of a DS-11 Spectrophotometer and Fluorometer

(DeNovix Inc, Wilmington, DE, USA). Purity was determined using the UV-Vis spectrometer feature of the DS-11 which reports OD 260/230 and 260/280 ratios. Following the first bead clean-up, the high-molecular weight DNA sample was sheared to a mean size of 20 kb with the Diagenode Megaruptor 2 (Denville, NJ, USA) and subsequent size distribution was assessed with the High Sensitivity (HS) Large fragment kit run on the Fragment Analyzer (Agilent Technologies, Santa Clara, California, USA). A PacBio SMRTBell library was prepared using the sheared DNA using the SMRTBell Express Template Prep Kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA). The prepared library was bound and sequenced at the USDA-ARS Genomics and Bioinformatics Research Unit in Stoneville, MS on a Pacific Biosciences 8M SMRT Cell on a Sequel IIe system (Pacific Biosciences, Menlo Park, CA, USA) beginning with a 2-hour pre-extension followed by a 30-hour movie collection time. After sequencing, circular consensus sequences from the PacBio Sequel IIe subreads were obtained using the SMRTLink v8.0 software.

Concurrent to the PacBio HiFi library prep and sequencing, a HiC library was prepared from a slice of abdominal tissue from the same male *B. affinis* (ToLID iyBomAffi1) as was used for the PacBio WGS HiFi library. The proximity-ligated sequencing library was prepared using the Arima HiC kit (Arima Genomics, San Diego, California, USA) from crosslinked tissue prepared following the Arima HiC low input protocol. Following proximity ligation, DNA was sheared using a Bioruptor Pico (Diagenode, Dennville, NJ, USA) and DNA fragments in the range of 200-600 bp were selected as the input for Illumina library prep using the Swift Accel NGS 2S Plus kit (Integrated DNA Technologies, Coralville, IA, USA). Illumina 2x150 bp sequencing was performed on a NovaSeq 6000 at the Hudson Alpha Genome Sequencing Center (Huntsville, AL, USA), and adapter trimming after sequence collection was performed using

BaseSpace software (Illumina, San Diego, CA, USA). All remaining tissue samples and DNA extractions are housed at the *B. affinis* genetic resources repository at the USA-ARS PIRU.

Genome assembly. Prior to genome assembly, HiFi reads containing artifact adapter sequences were removed from the HiFi read pool using the program HiFiAdapterFilt v2.0 (Sim *et al.* 2022). This filtered read set was assembled into a contig assembly using HiFiASM v0.16.1-r375 (Cheng *et al.* 2021) using the default parameters. The output of HiFiASM was an assembly in .gfa format which was converted to a .fasta format using any2fasta (Seeman 2018 https://github.com/tseemann/any2fasta). The primary contig assembly was scaffolded following the Arima Genomics mapping pipeline (Ghurye et al. 2019, https://github.com/ArimaGenomics/mapping_pipeline) and YaHS scaffolding software (Zhou *et al.* 2022) (https://github.com/c-zhou/yahs). The Arima Genomics mapping pipeline uses BWA mem (Li 2013) to align the paired Illumina R1 and R2 reads separately to the reference contig assembly and applies the filtering script `filter_five_end.pl` to only retain reads that are mapped in the 5' orientation. Following filtering, the independently mapped R1 and R2 reads were paired using the script `two_read_bam_combiner.pl` which resulted in a sorted and quality filtered paired-end file in .bam format. The `MarkDuplicates` function of Picard Tools (Picard Tools, 2019, https://broadinstitute.github.io/picard/) was used to remove PCR duplicate artifacts from the mapped and paired .bam which along with the reference contig assembly served as the input files for the YaHS scaffolding software. The YaHS software was implemented using the `no contig error correcting` option and YaHS outputs were converted using the `juicer_pre` function of YaHS to Juicebox (Durand *et al.* 2016) compatible files for manual curation visually within Juicebox. Following manual curation, edits were applied to the scaffold assembly using

`juicebox_assembly_converter.py` from the Phase Genomics suite of juicebox_scripts (https://github.com/phasegenomics/juicebox_scripts).

The chromosome scale assembly was assessed for completeness in terms of gene content using a Benchmark of Universal Single Copy Orthologs (BUSCOs) using all relevant taxonomic databases for the genome (Eukaryota, Metazoa, Arthropoda, Insecta, and Endopterygota) and only the most derived database, Endopterygota for the protein set. *Ab initio* annotations on the scaffold assembly were performed using Metaeuk v.4.a0f584d (Levy Karin *et al.* 2020) for the Eukaryota, Arthropoda, Insecta, and Endopterygota odb10 databases and Augustus v3.4.0 (Stanke *et al.* 2008) were used to detect the Metazoa odb10 orthologs. Designation of genes as complete single copy, duplicated, fragmented, or missing were determined using BUSCO v5.2.2 (Manni *et al.* 2021) in `genome` mode for the genome assembly and `protein` for the annotated protein set. Identification for off-target (non-*B. affinis*) contigs in the assembly was performed by aligning all contigs to the NCBI nucleotide (NT) database (accessed 2022-02-14) using the `blastn` function of BLAST+ v2.5.9+ (Camacho *et al.* 2009). The contigs were secondarily aligned to the UniProt protein database (accessed 2020-03) using Diamond (Buchfink *et al.* 2021). Local alignments to the nucleotide and protein databases were then used to assign the *B. affinis* contigs to a taxon using the rule `bestsumorder` of blobtoolkit v.2.6.1 (Challis *et al.* 2020) which assigns contigs to a taxon first based on alignments to the nucleotide database and then followed by alignments to the protein database if there were no hits to the nucleotide database. Taxonomic assignment of assembled scaffolds was tertiarily conducted using NCBI Foreign Contamination Screen (FCS, https://github.com/ncbi/fcs/wiki) tool suite using the fcs-gx function which uses the genome cross-species aligner (GX) to identify contaminants of which there were none in the final assembly. Coverage per scaffold and contig record was calculated using minimap2 v2.2-r1101

(Li 2021). Coverage, taxonomic assignment, and BUSCO results were aggregated using blobtoolkit and subsequently summarized using blobblurb v2.0.1(Sim 2022). Expected genome size was estimated using GenomeScope v2.0 (Ranallo-Benavidez *et al.* 2020) which uses k-mer frequency analysis of k-mer counts performed by KMC v3.2.1 (Kokot *et al.* 2017). Level of duplicate artifacts in the assembly was assessed using BUSCO results for both the genome and the protein set and using k-mer abundance in the raw HiFi reads relative to their representation in the final assembly as determined by K-mer Analysis Toolkit v2.4.2 (KAT) (Mapleson *et al.* 2017).

Genome annotation and synteny. The *B. affinis* genome was submitted to the National Center for Biotechnology Information (NCBI) RefSeq (Rajput *et al.* 2019) using the NCBI Eukaryotic Genomic Annotation Pipeline v10.0. This annotation method has been used in other bumble bee species and insects associated with the AgPest100 and Beenome100 initiatives. Adhering to the NCBI pipelines provides standardization in the annotation methodology. Annotation of the *B. affinis* was supported by n = 184 transcript resources (RNA-Seq) for *B. terrestris*, a closely related bumble bee species in the subg. *Bombus* endemic to Europe, available on GenBank. We were unable to acquire high quality *B. affinis* specimens for subsequent RNA sequencing to support annotation. Finally, synteny of the *B. affinis* genome with a chromosome resolution assembly of *B. terrestris* was investigated by developing idiograms with RIdeogram (Hao *et al.* 2020). Idiograms provide the opportunity to visualize chromosomal distribution patterns and identify potential regions of the genome that are associated with genomic gaps, inversions, and repeats.

Data availability. The *B. affinis* genome assembly is associated with NCBI Bioproject #PRJNA827177 (ToLID: iyBomAffi1) in the Beenome100 umbrella Bioproject #PRJNA923301.

RefSeq raw sequence reads of the *B. affinis* genome is associated with NCBI Bioproject #PRJNA880876.

Results

Genome assembly. Extracted DNA from a single tissue sample of the *B. affinis* abdomen was successfully used to achieve a highly contiguous contig assembly with chromosomal resolution. After preliminary assembly, we determined that 92.5% of the contigs had BLAST hits to Arthropoda (Kingdom: Animalia) (Figure 2A). The remaining 7.5% of the contigs were categorized as follows: 4.5% as Microsporidia (Kingdom: Fungi), 0.25% matched either Chordata (Kingdom: Animalia), Streptophyta (Kingdom: Plantae), or Ciliophora (Kingdom: Chromista) (combined = 0.25%); and 2.75% did not match any taxonomic group (i.e., "No-Hit"). It is of special note that the 4.5% of the contigs matched with the Microsporidia. Microsporidia, namely *V. bombi* is a significant fungal parasite associated with bumble bee decline. In total 109 sequences of Microsporidia accounted for a total of 18 Mb in the metagenome (Figure 2A). Further filtering with Blobtools and alignment to HiC sequences resulted in the removal of non-Arthropod contaminants in the remaining scaffolds (Figure 2B).

The assembled *B. affinis* genome resulted in 858 contigs with a contig N50 of 12.33 Mb (Table 1) (Figure 3A). Our assembly thus exceeds the minimum reference standard of the 6.7.Q40 quality category (>1.0 Mb contig and 10.0 Mb scaffold N50) suggested by the Earth BioGenome Project (Lewin *et al.* 2018; Lawniczak *et al.* 2022). The highly contiguous *B. affinis* genome assembly is comparable to the *B. ignitus* genome assembly (No. contigs = 250) but was 243.2% more fractionated than the *B. terrestris* assembly. The *B. affinis* genome assembly was less fractionated than the *B. impatiens* and *B. terricola* genomes as represented by the number of contigs (Table 1). The size of the scaffolded *B. affinis* assembly is 365.1 Mb, which is smaller

than the most recent *B. terrestris* genome assembly (v1.2), but larger than other published genomes of species in the subg. *Bombus* [*B. ignitus* (v1) = 242.6 Mb and *B. terricola* (Kent et al. (2018)) = 239.9] (Figure 3B). The count of the smallest number of contigs whose length makes up half the genome size of the current *B. affinis* assembly is L50 = 11 and is comparable to other high-quality genomes of the subg. *Bombus* currently published in NCBI (Table 1). GC content of the *B. affinis* assembly is comparable to the species assemblies included in our study at ~37.4% (average = 37.8% +/- 0.24%) (Figure 3B) (Table 1).

Total haploid assembly size is estimated to be 365.50 Mb based on k-mer analysis with GenomeScope v2 (k-mer = 21, k-cov = 18) (Figure 3C). The assembly is hypothesized to have minimal error (=0.121%), likely due to low sequencing error and low repetitive content of sequences (=0.126%). HiC contact mapping was able to construct 18 chromosome-length scaffolds (Figure 3D). Continued demonstration of the highly complete genome assembly is exemplified by BUSCO scores (Figure 3B) (Table 1). The *B. affinis* assembly performed exceptionally better than all other published assemblies in the subg. *Bombus*, with 98.9% of the 5,991 benchmarking universal single-copy orthologs represented in OrthoDB v1.10 Hymenoptera lineage dataset (hymenoptera_odb10). Most of the genes were single copy (98.1%) with 0.4% duplicated or 1.1% missing (Figure 3B) (Table 1).

Genome annotation and synteny. In total, the NCBI Eukaryotic Genomic Annotation Pipeline predicted 15,252 genes and pseudogenes, of which 14,971 genes gave rise to 35,888 transcripts. The number of predicted genes in *B. affinis* exceeds the number of genes predicted in *B. terrestris* and *B. impatiens* by 13.84% to 15.89% (Table 2). This increase in the number of predicted genes is mostly associated with non-coding genes in the *B. affinis* genome. Specifically, *B. affinis* is estimated to have 44.68% and 85.43% more non-coding genes than *B. terrestris* and

B. impatiens, respectively (Table 2). Non-coding genes do not result in proteins but have been demonstrated to be important in serving roles in the regulation of gene expression, whereas other studies continue to demonstrate no known function understood by science. The statistics of the genome annotation in *B. affinis* is largely consistent with the available bumble bee genomes we compare in our study that are available on RefSeq (Table 2).

We analyzed the synteny of the *B. affinis* genome with *B. terrestris*. The new *B. affinis* assembly is highly collinear with the 18 *B. terrestris* linkage groups (Figure 4). However, we were able to visualize several rearrangements in the genomic architecture of *B. affinis* relative to *B. terrestris*. For example, our synteny analysis revealed evidence for multiple inversions (*e.g., B. affinis* CM044943.1 vs. *B. terrestris* OU342922.1 = 1 inversion) (Figures 4A, 4B), gaps (e.g., *B. affinis* CMO44953.1 vs. *B. terrestris* OU342937.1) (Figure 4A), and repeats (e.g., *B. affinis* CM044946.1 vs. *B. terrestris* OU342932.1) (Figure 4B). The structural rearrangements may represent a small number of assembly artifacts, or may represent rearrangements that are hypothesized to have occurred over the last 5 millions years since the last common ancestor at the Pliocene and Miocene boundary (Hines 2008). Overall, our synteny analysis continues to demonstrate that bumble bee genomes are highly conserved, even across deep timescales, particularly within subgenera (Heraghty *et al.* 2020).

Discussion

Highly contiguous contig assemblies of invertebrates are becoming increasingly available, especially considering recent initiatives by the Ag100Pest and Beenome100 initiatives. In our study, we demonstrated the promise of using limited tissue samples from a single insect specimen to produce a high-quality genome assembly that achieves chromosomal-level resolution. This achievement was possible in large part to the capacity for PacBio HiFi sequencing technology to

capture long and accurate read lengths during the sequencing process. Using a fraction of the tissue available in constructing a high-quality assembly enables scientists to retain a physical voucher of the tissue for additional research and confirmation of species identity.

In addition to producing a high-quality genome assembly, our sequencing approach also recovered evidence for a significant bumble bee pathogen – Microsporidia. From the preliminary assembly of the *B. affinis* genome, we determined that 4.5% (18 Mb) of the contigs matched sequences associated with Microsporidia. Microsporida such as *V. bombi* are an obligate, intracellular gut pathogen that can produce systemic infection in many invertebrates, including the bumble bees (Macfarlane et al. 1995). In our tissue sampling approach, we specifically targeted a segment of the abdomen to reduce unnecessary specimen destruction. Organs within the abdomen that can be readily infected by *V. bombi* include Malpighian tubules, midgut, and fat body (Macfarlane *et al.* 1995). Thus, the capture of Microsporidia genetic information in our minimally destructive tissue sampling of the insect host highlights the value of our molecular, sequencing, and bioinformatic framework in detecting parasites of significant interest.

The mechanisms associated with *B. affinis* decline, and potential solutions for recovery is an active area of research. Recovery of the species would benefit from an assessment of the underlying genetic diversity, including the development of a reference genome (Smith *et al.* 2020). The high-quality reference genome of *B. affinis* we present in our study will enable scientists and stakeholders to compare DNA sequenced from tissue samples associated with hundreds of specimens using high-throughput short-read sequencing. Sequence variation can then be compared to the high resolution reference genome to make inferences about diverse population-level phenomena (Webster *et al.* 2022). Specific research objectives that can be achieved with a reference genome include reconstruction of population size fluctuations (*Ne*),

estimations of deleterious genetic variation, identification of the genetic mechanisms associated with environmental stressors, including pesticides (Kent *et al.* 2018), and predicting evolution in response to the stressors that drive insect decline (Grozinger and Zayed 2020; Webster *et al.* 2022). Ultimately, the results of this research agenda can support evidence-based decisions on *B. affinis* management across diverse habitat types and *ex situ* scenarios.

Acknowledgments

This work was supported by the U.S. Fish and Wildlife Service (USFWS) and the U.S. Department of Agriculture, Agricultural Research Service (USDA-ARS). The genome assembly was generated as part of the USDA-ARS Beenome100 Initiative (https://www.beenome100.org/). Specimens were collected by Dr. Elaine Evans with permission from USFWS (Sub-Permit #: USFWS 16-07-03a). This research used resources provided by the SCINet project of the USDA-ARS project number 0500-00093-001-00-D. The authors thank the members of the USDA-ARS Beenome100 and Ag100Pest Team for sequencing and analysis support. All opinions expressed in this paper are the authors' and do not necessarily reflect the policies and views of USDA and USFWS. Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Government. USDA and USFWS are an equal opportunity provider and employer.

Author Contributions

JBK conceived and provided supervision of the research. JBK and SBS designed and performed research, analyzed data, and wrote the initial paper. SG and BS supported research with new analytical tools, reagents, and resources. TS supported funding and specimen acquisition for the research. All authors contributed to review and editing the final paper.

References

Boone, M. L., E. Evans, A. Wolf, H. Minser, J. Watson *et al.*, 2022 Notes from rusty patched bumble bee (*Bombus affinis* Cresson) nest observations. Insect Conserv. Divers. 15: 380–384.

Buchfink, B., K. Reuter, and H.-G. Drost, 2021 Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat. Methods 18: 366–368.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos *et al.*, 2009 BLAST+: architecture and applications. BMC Bioinformatics 10: 421.

Cameron, S. A., H. C. Lim, J. D. Lozier, M. A. Duennes, and R. Thorp, 2016 Test of the invasive pathogen hypothesis of bumble bee decline in North America. Proc. Natl. Acad. Sci. U.S.A. 113: 4386–4391.

Cameron, S. A., J. D. Lozier, J. P. Strange, J. B. Koch, N. Cordes *et al.*, 2011 Patterns of widespread decline in North American bumble bees. Proc. Natl. Acad. Sci. U.S.A. 108: 662–667.

Canada, E. A. C., 2020 Recovery Strategy for the Rusty-patched Bumble Bee (*Bombus affinis*) in Canada: Environment and Climate Change Canada, Ottawa.

Challis, R., E. Richards, J. Rajan, G. Cochrane, and M. Blaxter, 2020 BlobToolKit - interactive quality assessment of genome assemblies. G3 10: 1361–1374.

Cheng, H., G. T. Concepcion, X. Feng, H. Zhang, and H. Li, 2021 Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods 18: 170–175.

Childers, A. K., S. M. Geib, S. B. Sim, M. F. Poelchau, B. S. Coates *et al.*, 2021 The USDA-ARS Ag100Pest Initiative: High-Quality Genome Assemblies for Agricultural Pest Arthropod Research. Insects 12: 626.

Colla, S. R., and L. Packer, 2008 Evidence for decline in eastern North American bumblebees (Hymenoptera: Apidae), with special focus on *Bombus affinis* Cresson. Biodivers. Conserv. 17: 1379–1391.

COSEWIC. 2010. COSEWIC assessment and status report on the Rusty–patched Bumble Bee *Bombus affinis* in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. vi + 34 pp.

Durand, N. C., J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov *et al.*, 2016 Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst 3: 99–101.

Grozinger, C. M., and A. Zayed, 2020 Improving bee health through genomics. Nat. Rev. Genet. 21: 277–291.

Hao, Z., D. Lv, Y. Ge, J. Shi, D. Weijers *et al.*, 2020 RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. PeerJ Comput Sci 6: e251.

Heraghty, S. D., J. M. Sutton, M. L. Pimsler, J. L. Fierst, J. P. Strange *et al.*, 2020 De Novo Genome Assemblies for Three North American Bumble Bee Species: *Bombus bifarius*, *Bombus vancouverensis*, and *Bombus vosnesenskii*. G3 10: 2585–2592.

Hines, H. M., 2008 Historical biogeography, divergence times, and diversification patterns of bumble bees (Hymenoptera: Apidae: *Bombus*). Syst. Biol. 57: 58–75.

Kent, C. F., A. Dey, H. Patel, N. Tsvetkov, T. Tiwari *et al.*, 2018 Conservation Genomics of the Declining North American Bumblebee *Bombus terricola* Reveals Inbreeding and Selection on Immune Genes. Front. Genet. 9: 316.

Koch, J. B., J. Lozier, J. P. Strange, H. Ikerd, T. Griswold *et al.*, 2015 USBombus, a database of contemporary survey data for North American Bumble Bees (Hymenoptera, Apidae, *Bombus*) distributed in the United States. Biodivers Data J e6833.

Kokot, M., M. Dlugosz, and S. Deorowicz, 2017 KMC 3: counting and manipulating k-mer statistics. Bioinformatics 33: 2759–2761.

Lawniczak, M. K. N., R. Durbin, P. Flicek, K. Lindblad-Toh, X. Wei *et al.*, 2022 Standards recommendations for the Earth BioGenome Project. Proc. Natl. Acad. Sci. U. S. A. 119.:

Levy Karin, E., M. Mirdita, and J. Söding, 2020 MetaEuk—sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. Microbiome 8: 48.

Lewin, H. A., G. E. Robinson, W. J. Kress, W. J. Baker, J. Coddington *et al.*, 2018 Earth BioGenome Project: Sequencing life for the future of life. Proc. Natl. Acad. Sci. U.S.A. 115: 4325–4333.

Li, H., 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv [q-bio.GN].

Li, H., 2021 New strategies to improve minimap2 alignment accuracy. Bioinformatics 37: 4572–4574.

Macfarlane, R. P., J. J. Lipa, H. J. Liu, 1995 Bumble bee pathogens and internal enemies. Bee World, 76, 130–148.

Manni, M., M. R. Berkeley, M. Seppey, F. A. Simão, and E. M. Zdobnov, 2021 BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Mol. Biol. Evol. 38: 4647–4654.

Mapleson, D., G. Garcia Accinelli, G. Kettleborough, J. Wright, and B. J. Clavijo, 2017 KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics 33: 574–576.

Milliron, H. E., 1971 A monograph of the western hemisphere bumblebees (Hymenoptera: Apidae; Bombinae). I. Mem. Entomol. Soc. Can. 103: 1–80.

Mola, J. M., L. L. Richardson, G. Spyreas, D. N. Zaya, and I. S. Pearse, 2021 Long-term surveys support declines in early season forest plants used by bumblebees. J. Appl. Ecol. 58: 1431–1441.

Otto, C. R. V., A. C. Schrage, L. L. Bailey, J. M. Mola, T. A. Smith *et al.*, 2022 Addressing Detection Uncertainty in *Bombus affinis* (Hymenoptera: Apidae) Surveys Can Improve Inferences Made From Monitoring. Environ. Entomol. https://doi.org/10.1093/ee/nvac090.

Rajput, B., K. D. Pruitt, and T. D. Murphy, 2019 RefSeq curation and annotation of stop codon recoding in vertebrates. Nucleic Acids Res. 47: 594–606.

Ranallo-Benavidez, T. R., K. S. Jaron, and M. C. Schatz, 2020 GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat. Commun. 11: 1432.

Sim, S. B., 2022 *blobblurb*. https://github.com/sheinasim/blobblurb.

Simanonok, M. P., C. R. V. Otto, R. S. Cornman, D. D. Iwanowicz, J. P. Strange *et al.*, 2021 A century of pollen foraging by the endangered rusty patched bumble bee (*Bombus affinis*): inferences from molecular sequencing of museum specimens. Biodivers. Conserv. 30: 123–137.

Sim, S. B., R. L. Corpuz, T. J. Simmonds, and S. M. Geib, 2022 HiFiAdapterFilt, a memory efficient read processing pipeline, prevents occurrence of adapter sequence in PacBio HiFi reads and their negative impacts on genome assembly. BMC Genomics 23: 157.

Smith, T.A., J.P. Strange, E.C. Evans, B.M. Sadd, J.C. Steiner, J.M. Mola and K. Traylor-Holzer

406 (Ed.), 2020 Rusty Patched Bumble Bee, *Bombus affinis*, Ex Situ Assessment and Planning
407 Workshop: Final Report: IUCN SSC Conservation Planning Specialist Group, Apple Valley,
408 MN, USA.
409 Soroye, P., T. Newbold, and J. Kerr, 2020 Climate change contributes to widespread declines
410 among bumble bees across continents. Science 367: 685–688.
411 Stanke, M., M. Diekhans, R. Baertsch, and D. Haussler, 2008 Using native and syntenically
412 mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24: 637–644.
413 Tokarev, Y. S., W.-F. Huang, L. F. Solter, J. M. Malysh, J. J. Becnel *et al.*, 2020 A formal
414 redefinition of the genera *Nosema* and *Varimorpha* (Microsporidia: Nosematidae) and
415 reassignment of species based on molecular phylogenetics. J. Invertebr. Pathol. 169: 107279.
416 U.S. Fish and Wildlife Service. 2021. Recovery Plan for the Rusty Patched Bumble Bee (*Bombus*
417 *affinis*). Midwest Regional Office, Bloomington, MN.
418 U.S. Fish and Wildlife Service. 2016. Endangered and Threatened Wildlife and Plants;
419 Endangered Species Status for Rusty Patched Bumble Bee – Final Rule – FWS-R3-ES-2015-
420 0112; 4500030113. Federal Register 82(7): 3186–3208.
421 Webster, M. T., A. Beaurepaire, P. Neumann, and E. Stolle, 2023 Population Genomics for Insect
422 Conservation. Annu Rev Anim Biosci. https://doi.org/10.1146/annurev-animal-122221-
423 075025.
424 Zhou, C., S. A. McCarthy, and R. Durbin, 2022 YaHS: yet another Hi-C scaffolding tool.
425 bioRxiv 2022.06.09.495093.