Title

Genome divergence and reproductive incompatibility among populations of Ganaspis near brasiliensis.

Journal

G3: Genes, Genomes, Genetics, 14(7)

Authors

Hopper, Keith

Wang, Xingeng

Kenis, Marc

et al.

Publication Date

2024-07-08

DOI

10.1093/g3journal/jkae090

Peer reviewed

G3, 2024, 14(7), jkae090
https://doi.org/10.1093/g3journal/jkae090
Advance Access Publication Date: 8 May 2024
Investigation

Genome divergence and reproductive incompatibility among populations of *Ganaspis* near *brasiliensis*

Keith R. Hopper [1,*], Xingeng Wang,[1] Marc Kenis [2], M. Lukas Seehausen,[2] Paul K. Abram,[3] Kent M. Daane,[4] Matthew L. Buffington,[5] Kim A. Hoelmer,[1] Brewster F. Kingham [6], Olga Shevchenko,[6] Erin Bernberg[6]

[1]United States Department of Agriculture, Agricultural Research Service, 501 South Chapel Street, Newark, DE 19713, USA
[2]CABI, Rue des Grillons 1, CH-2800 Delémont, Switzerland
[3]Agriculture and Agri-Food Canada, Agassiz Research and Development Centre, 6947 Highway 7, PO Box 1000, Agassiz, BC V0 M 1A2, Canada
[4]Department of Environmental Science, Policy and Management, University of California, Berkeley, Berkeley, CA 94720-3114, USA
[5]United States Department of Agriculture, Agricultural Research Service, Washington, c/o Smithsonian Institution, National Museum of Natural History, 10th and Constitution NW, MRC-168, Washington, DC 20013-7012, USA
[6]DNA Sequencing & Genotyping Center, Delaware Biotechnology Institute, 590 Avenue 1743, Newark, DE 19713, USA

*Corresponding author: United States Department of Agriculture, Agricultural Research Service, 501 South Chapel Street, Newark, DE 19713, USA.
Email: Keith.Hopper@usda.gov

During the last decade, the spotted wing drosophila, *Drosophila suzukii*, has spread from eastern Asia to the Americas, Europe, and Africa. This fly attacks many species of cultivated and wild fruits with soft, thin skins, where its serrated ovipositor allows it to lay eggs in undamaged fruit. Parasitoids from the native range of *D. suzukii* may provide sustainable management of this polyphagous pest. Among these parasitoids, host-specificity testing has revealed a lineage of *Ganaspis* near *brasiliensis*, referred to in this paper as G1, that appears to be a cryptic species more host-specific to *D. suzukii* than other parasitoids. Differentiation among cryptic species is critical for introduction and subsequent evaluation of their impact on *D. suzukii*. Here, we present results on divergence in genomic sequences and architecture and reproductive isolation between lineages of *Ganaspis* near *brasiliensis* that appear to be cryptic species. We studied five populations, two from China, two from Japan, and one from Canada, identified as the G1 vs G3 lineages based on differences in cytochrome oxidase I sequences. We assembled and annotated the genomes of these populations and analyzed divergences in sequence and genome architecture between them. We also report results from crosses to test reproductive compatibility between the G3 lineage from China and the G1 lineage from Japan. The combined results on sequence divergence, differences in genome architectures, ortholog divergence, reproductive incompatibility, differences in host ranges and microhabitat preferences, and differences in morphology show that these lineages are different species. Thus, the decision to evaluate the lineages separately and only import and introduce the more host-specific lineage to North America and Europe was appropriate.

Keywords: genome assembly; genome annotation; reproductive incompatibility; biological control; parasitoid; Hymenoptera

Introduction

During the last decade, the spotted wing drosophila, *Drosophila suzukii* (Matsumura) (Diptera: Drosophilidae), has established widely in the Americas, Europe, and Africa (Asplen *et al.* 2015; Tait *et al.* 2021). *Drosophila suzukii* attacks many species of cultivated and wild fruits with soft, thin skins (Kirschbaum *et al.* 2020). Current management relies on frequent use of insecticides. Parasitoids from the native range of *D. suzukii* may provide sustainable management of this polyphagous pest without the use of insecticides (Lee *et al.* 2019; Wang *et al.* 2020). Recent exploration has revealed several species of larval parasitoids of *D. suzukii* in South Korea (Daane *et al.* 2016), China (Girod *et al.* 2018a; Giorgini *et al.* 2019), and Japan (Girod *et al.* 2018a). Among them, *Ganaspis brasiliensis* Ihering and *Ganaspis* cf. *brasiliensis* (Hymenoptera: Figitidae) were among the most common in intact fruits infested by *D. suzukii* and the closely related species, *Drosophila pulchrella* Tan, Hsu & Sheng and *Drosophila subpulchrella* Takamori. Host-specificity testing showed that a lineage of *Ganaspis* near *brasiliensis* was more specific to *D. suzukii* than other parasitoids (Wang *et al.* 2018; Girod *et al.* 2018b; Seehausen *et al.* 2020; Daane *et al.* 2021). This lineage appears to be a cryptic species similar morphologically to *G. brasiliensis* but differing in behavior and genetics. We consider species to be entities that are phylogenetically and genetically distinct, are reproductively isolated, and show important differences in biology, for example, differences in host specificity. A paper by one of us (MLB) in review at the Journal of Hymenoptera Research describes subtle morphological differences between these lineages and gives them species names. Here, we present results on divergence in genomic sequences and reproductive isolation between these cryptic species of *Ganaspis* near *brasiliensis*.

Ganaspis brasiliensis was first recorded in the Caribbean and Brazil and later redescribed as the new combination of *G. brasiliensis* based on morphology (Buffington and Forshage 2016). Prior to the redescription, *Ganaspis* species found parasitizing *D. suzukii* were referred to a "*D. suzukii*-specific-strain" of *Ganaspis xanthopoda* Ashmead in Japan (Mitsui *et al.* 2007; Kasuya *et al.* 2013). Nomano *et al.* (2017) grouped *Ganaspis* from different geographical regions

Received on 24 October 2023; accepted on 12 April 2024
Published by Oxford University Press on behalf of The Genetics Society of America 2024.

into five lineages (called G1–G5) based mainly on sub-sequences of the mitochondrial cytochrome oxidase I (COI) gene. The G2 lineage included individuals from the subtropical Japanese island of Iriomote-Jima reared from *Drosophila ficusphila* Kikkawa & Peng, the G4 lineage included individuals from Indonesia reared from *Drosophila eugracilis* Bock and Wheeler, and the G5 lineage included individuals from Japan and regions outside Asia from unknown hosts. Nomano *et al.* (2017) suggested that other specimens previously described as *G. xanthopoda* or *Ganaspis* sp. from Thailand and the Philippines (Schilthuizen *et al.* 1998), Hawaii and Uganda (Kacsoh and Schlenke 2012), Indonesia (Kimura and Suwito 2012; 2015), Benin and the Caribbean Sea, Brazil (Buffington and Forshage 2016), and Malaysia (Nomano *et al.* 2017) belong to the G5 lineage. Specimens collected from *D. suzukii* and two co-occurring hosts, *D. pulchrella* and *D. subpulchrella*, during recent exploration in Asia were assigned to the G1 and G3 lineages and possibly the G4 lineage (Giorgini *et al.* 2019; Seehausen *et al.* 2020).

The G1 and G3 lineages are endemic to East Asia and coexist on host plants infested by *D. suzukii* and two closely related fly species (Giorgini *et al.* 2019). For example, parasitoids collected in Yunnan, China, in 2016 consisted of 23% G3 and 77% G1 based on sequencing COI in a subsample of 30 individuals (Giorgini *et al.* 2019). Parasitoids collected in South Korea in 2017 were 25% G3 and 75% G1 based on sequencing COI of a subsample of 48 individuals (unpublished data). Laboratory tests of host specificity showed that G1 and G3 can attack *D. suzukii* and closely related species like *Drosophila melanogaster* and *Drosophila simulans* (Daane *et al.* 2021). However, G1 seems more specific to *D. suzukii* (Nomano *et al.* 2017; Girod *et al.* 2018b; Giorgini *et al.* 2019). G5 from Asia and other regions differs in host specificity from G1 (Nomano *et al.* 2017; Seehausen *et al.* 2020). In laboratory and field cage tests, G1 almost exclusively parasitized *Drosophila* larvae feeding on intact fruits, whereas G3 readily parasitized *D. melanogaster* and *D. simulans* in rotting fruits as well as *D. suzukii* (Girod *et al.* 2018b; Seehausen *et al.* 2020; Seehausen *et al.* 2022). G5 from Hawaii and Uganda had low success in development on *D. suzukii* in the laboratory (Kacsoh and Schlenke 2012). *Ganaspis brasiliensis* was reported in Mexico, but this population was unable to develop on *D. suzukii* (Sanchez-Gonzalez *et al.* 2020), and *G. brasiliensis* was collected from a single *D. suzukii* puparium in Argentina (Gallardo *et al.* 2022). Both the Mexican and Argentinian *G. brasiliensis* are likely in the G5 lineage (Sanchez-Gonzalez *et al.* 2020; Gallardo *et al.* 2022). Recent surveys in British Columbia (BC), Canada, and Washington State, USA, found *Ganaspis* near *brasiliensis*, and these populations were assigned to the G1 lineage based on COI sequences (Abram *et al.* 2020; Beers *et al.* 2022). The G1 lineage has been recently approved for field release against *D. suzukii* in France, Italy, and the USA.

The species status of these lineages of *Ganaspis* near *brasiliensis* has not yet been resolved. The G1 and G3 lineages differ in acid-soluble insect protein spectra (Reeve and Seehausen 2019), and Seehausen *et al.* (2020) found that Japanese G1 and G3 were reproductively incompatible and never observed copulation between females of one lineage with males of the other, suggesting that they are cryptic species. Here, we report genome assemblies, annotations, and sequence divergence of material from five populations of *Ganaspis* near *brasiliensis*. These include a Chinese population identified as G1 and another identified as G3, a Japanese population identified as G1 and another identified as G3, and a Canadian population identified as G1 (Supplementary Table 1). We also report results from crosses to test reproductive compatibility between the G3 population from China and the G1 population from Japan. Our results support the status of the G1 and G3 lineages as different species.

Materials and methods

Insect sources

A colony of *D. suzukii* was started with field collections of infested cherries during 2010 in Davis, CA, USA, which was used for rearing parasitoid colonies. We studied material from colonies of *Ganaspis* near *brasiliensis* established from field collections of five populations. Collection and voucher details for the parasitoid material are given in Supplementary Table 1.

A colony of *Ganaspis* near *brasiliensis* from Yunnan was started from field collections in Kunming, Yunnan, China, in 2016. Wild berries of *Rubus foliosus* Weihe, *Rubus niveus* Thunberg, *Fragaria moupinensis* Cardot (Rosaceae), and *Sambucus adnata* Wallich (Adoxaceae) were collected in the suburbs of Kunming. The berries were often infested by *D. suzukii* and the closely related *D. pulchrella*. About 600 adult parasitoids emerged from imported puparia at the quarantine facility of the University of California, Berkeley (UCB). These specimens were assigned to two lineages, G1 and G3, based on COI sequences. A colony of the G3 lineage from Yunnan was started at USDA-ARS Beneficial Insects Introduction Research Unit (BIIRU), Newark, DE, USA, from about 100 females and 50 males received from the UCB in 2018. In this paper, we will designate this material G3-Yunnan.

A colony of parasitoids from Tokyo [referred to as *Ganaspis* cf. *brasiliensis* in Girod *et al.* (2018a) and Seehausen *et al.* (2020)] was started from collections in 2016 from *D. suzukii* on wild cherry *Prunus serrulata* in Naganuma Park, Hachioji, Tokyo. This population was assigned to the G1 lineage based on its COI sequence (Nomano *et al.* 2017; Seehausen *et al.* 2020). The colony is maintained in the quarantine laboratory at CABI in Delémont, Switzerland (Girod *et al.* 2018a). An Italian colony of the parasitoids from Tokyo was started at Fondazione Edmund Mach, San Michele all'Adige, Italy, in 2020 from 150 wasps from the CABI colony, and the BIIRU colony from this source was established from about 500 wasps from Italy in 2021. In this paper, we will designate this material G1-Tokyo.

A colony of parasitoids was established from material collected in 2017 in Xining Park, Kunming, Yunnan, China, parasitizing *D. suzukii* and *D. pulchrella* on *Prunus* sp. and identified as a G1 lineage (Girod *et al.* 2018a; Xining strain in Seehausen *et al.* 2020). In this paper, we will designate this material G1-Yunnan.

A colony of parasitoids was established from material collected in 2017 in Hasuike—Shiga Kogen, Nagano Prefecture, Japan, parasitizing *D. suzukii* and *D. subpulchrella* on *Vaccinium* sp. and identified as a G3 lineage (Girod *et al.* 2018a; Hasuike strain in Seehausen *et al.* 2020). In this paper, we will designate this material G3-Nagano.

A colony of parasitoids was established from material collected in 2020 in BC, Canada (Abram *et al.* 2020, 2022; Beers *et al.* 2022). In this paper, we will designate this material G1-BC. Some of the material from BC identified as G3 did not provide sufficient data for genome assembly and annotation, so we do not include the results from this material here.

Insect rearing

Colonies of *D. suzukii* and parasitoids were maintained with the methods described by Rossi-Stacconi *et al.* (2022). Briefly, *D. suzukii* was maintained on an artificial diet in 250-ml flasks. The parasitoid populations were maintained on blueberries infested by *D. suzukii*. Fruits were exposed to *D. suzukii* for 1–2 days for

oviposition in screen cages (30 × 30 × 30 cm). The parasitoids were reared in clear plastic containers (9 × 12 × 8 cm) by exposing 5–10 female wasps to 10–20 infested blueberries for 4–5 days, with droplets of honey streaked on the container's screen as a food source. Following exposure, infested fruits were removed from cages and kept in new plastic containers with filter paper at the bottom to absorb leaking fruit juice. Newly emerged wasps were collected in plastic vials (95 × 25 mm) and provided honey. Reared and parasitoid crosses done in plant growth chambers (23 ± 1°C, 14-hour light:10-hour dark, 40–60% relative humidity) at the containment facility at USDA-ARS, Newark, DE, USA.

Genomic DNA libraries and sequencing

DNA was extracted, and libraries were prepared and sequenced at the DNA Sequencing and Genotyping Center, Delaware Biotechnology Institute, University of Delaware, Newark, DE, USA. Briefly, genomic DNA was extracted from an individual male from each population by cryo-pulverizing it in a 1.5-ml tube. After pulverizing, DNA was separated from other material with HMW MagAttract kits (Qiagen, Germantown, MD, USA). Lysis was done at 56°C for 30 minutes, but the shaking steps were replaced by 15 minutes of rotation to reduce DNA shearing. Extracted DNA was quantified with a Qubit fluorometer (Invitrogen, Waltham, MA, USA), and DNA length distributions were determined on a Femto Pulse electrophoresis system (Agilent Technologies, Santa Clare, CA, USA). PacBio HiFi libraries were prepared from the extracted DNA with low-input and ultra-low-input kits (Pacific Biosciences, CA, USA). The resultant libraries were sequenced on a PacBio Sequel IIe.

Genome assemblies

For all programs used in this paper, citations, URL locations, versions, parameter settings, and template scripts for running the programs are given in Supplementary Document 1. We made assemblies of the DNA from G1-Tokyo and G3-Yunnan males with seven assemblers: CSA_assemble (Kuhl *et al.* 2020) (pipeline that runs wtdbg2, Ruan and Li 2020), Flye (Kolmogorov *et al.* 2019), Hifiasm (Cheng *et al.* 2021), IPA (https://github.com/PacificBiosciences/pbbioconda/wiki/Improved-Phased-Assembler), Mecat2 (Xiao *et al.* 2017), Miniasm (Li 2016), and Raven (Vaser and Šikić 2020) with raw and error-corrected reads, where appropriate. Some assemblies were run through the HERA pipeline for gap closing (Du and Liang 2019). Metrics of these assemblies are given in Supplementary Table 2. Three assemblers (Mecat2, Miniasm, and Raven) gave assembly sizes smaller than expected based on the flow cytometry estimate of 971 Mb for a closely related species, *G. xanthopoda* (Gokhman *et al.* 2011), so we did not consider these assemblers further. Among the assemblies made with remaining four assemblers (CSA-assemble, Flye, Hifiasm, and IPA), Hifiasm gave an assembly of G1-Tokyo with N_{50} 3–26-fold higher and L_{50} 3–18-fold lower than the other assemblies and an assembly of G3-Yunnan with N_{50} 0.9–7-fold higher and L_{50} 0.9–4-fold lower than the other assemblies (Supplementary Table 2). Therefore, we report here assemblies of all the genomes made with Hifiasm. We compared assemblies using contiguity statistics from QUAST (Mikheenko *et al.* 2018), in particular N_{50}, which is the length of shortest contig at which over half the genome is covered, and L_{50}, which is the number of contigs covering 50% of the genome. We also analyzed the assemblies for the number of complete genes in the Benchmark Universal Single-Copy Orthologs (BUSCO) set of 5,991 genes from Hymenoptera (Hymenoptera_odb10) (Manni *et al.* 2021).

Bacterial contamination

To find bacterial contamination in the *Ganaspis* assemblies, we ran blastn with the assembly contigs against the NCBI nt database (Altschul *et al.* 1997; NCBI Resource Coordinators 2016). We restricted our searches to Bacteria and Insecta because initial searches with all of nt showed that most contigs with hits to Bacteria also had hits to Eukaryota, which can be explained by many DNA database accessions identified as Eukaryota being contaminated with sequences from Bacteria (Merchant *et al.* 2014; Lu and Salzberg 2018). We determined the taxa of Bacteria that were among the contaminant sequences and removed the contigs with bacterial sequences from the assemblies for downstream analyses.

Mitochondrial genomes

To find the mitochondrial genomes in these assemblies, we used blastn to search for matches in the assemblies with sequences in the published mitochondrial genomes of *Leptopilina boulardi* (Oliveira *et al.* 2016) and *Leptopilina syphax* (Zhang *et al.* 2021). This approach revealed a single contig in each *Ganaspis* assembly, which we submitted to Mitos2 (Donath *et al.* 2019) to identify mitochondrial features and determine their lengths and order. We compared feature orders with those reported for other figitid species (Shu *et al.* 2022).

To confirm lineage assignments of the males used for assemblies, we captured the COI sequences from the mitochondrial genomes and compared them with one another and with the sub-sequences of COI from *G. brasiliensis* in the NCBI nr nucleotide database (ncbi.nlm.nih.gov; queried November 2022). To make these comparisons, we did multiple sequence alignment, calculated distances between the aligned sequences, and made a neighboring-joining tree with these distances using functions in the following R packages msa (Bodenhofer *et al.* 2015), seqinr (Charif and Lobry 2007), and ape (Paradis 2012).

Repeats and transposable elements

We analyzed the reads used to make these assemblies for repetitive DNA with Jellyfish (Marçais and Kingsford 2011) and GenomeScope (Ranallo-Benavidez *et al.* 2020), which uses input from Jellyfish. We also searched the assemblies for transposons with the Extensive de novo TE Annotator (EDTA) pipeline (Ou *et al.* 2019). The EDTA pipeline identifies repeats de novo based on their sequences and the attributes of known transposable element (TE) families. This pipeline runs LTR_FINDER (Xu and Wang 2007), LTRharvest (Ellinghaus *et al.* 2008), and LTR_retriever (Ou and Jiang 2018) to find all of the long-terminal repeats (LTRs) in the genome. It also runs Generic Repeat Finder (Shi and Liang 2019), HelitronScanner (Xiong *et al.* 2014), and TIR-Learner (Su *et al.* 2019) to find non-LTR transposons. It then runs a series of filters and RepeatModeler (Flynn *et al.* 2020) to remove duplicates, identify full-length TEs, and create a final TE library.

Sequence divergence between genome assemblies

To find sequence differences among our assemblies of *Ganaspis* near *brasiliensis*, we mapped reads from each sample to assemblies of the other samples using minimap2 with the hifi option to set parameters for PacBio HiFi data (Li 2018). We analyzed the minimap2 output with BEDtools (genomecov) (Quinlan and Hall 2010) to find the proportion of the target assemblies covered by reads. We processed the minimap2 output through a pipeline

including routines in SAMtools (sort, index, flagstat) (Li *et al.* 2009), BCFtools (mpileup, call) (Li 2011; Danecek *et al.* 2021), and VCFtools (Danecek *et al.* 2011) to find the numbers and densities of single-nucleotide polymorphic (SNP) loci that differed between assemblies (see Supplementary Document 1 for links and commands). Although indels (insertion–deletions) can be very important in evolution, determining indel homology is problematic, so we did not analyze the differences of indels among assemblies.

Gene discovery and annotation

We used Augustus with the *Nasonia* gene model (Stanke *et al.* 2008) to identify protein-coding genes in these assemblies. To discover the function of the proteins expressed by these genes, we used blastp (blosum62 scoring matrix, *E*-value = 0.001, default values for other parameters) (Altschul *et al.* 1997; NCBI Resource Coordinators 2016) to search for sequence homology between our predicted amino acid sequences and those in the NCBI nr protein database (version 5, downloaded April 2020; ncbi.nlm.nih.gov) in the hope of finding ones with known function. We used blast2GO (Götz *et al.* 2008) to categorize the top hits and assign gene ontology (GO) annotations to our query sequences. Because sequence-based homology can occur between genes with different functions and miss similarities between genes with similar functions, we also used with InterProScan (Jones *et al.* 2014) to assess protein domain-based homology. Domain annotation reflects the known domain architectures associated with protein functions and can provide a more reliable way to categorize the functions of unannotated sequences.

Orthologs and ortholog divergence

We determined orthology among proteins across assemblies of these populations of *Ganaspis* near *brasiliensis* using Orthofinder (Emms and Kelly 2019). This program provided the numbers of orthogroups that overlapped between populations, as well as the number of genes in orthogroups, the number of orthogroups in each assembly, the number of assembly-specific orthogroups, and the number of genes in assembly-specific orthogroups. Orthofinder also generated a molecular phylogeny based on the genes in universal orthogroups, i.e. those that occur in all populations, using Species Tree inference from All Genes (; Emms and Kelly 2018) and Species Tree Root Inference from gene Duplication Events (; Emms and Kelly 2017).

Population crosses

We made crosses to test reproductive compatibility between G3-Yunnan and G1-Tokyo. For these crosses, parasitized *D. suzukii* puparia from the parasitoid colonies were isolated in plastic vials (95 × 25 mm). A piece of moisturized tissue paper was placed in each vial to provide humidity. When individuals emerged, they were supplied with a streak of honey on the bottom of the vial plug and paired within 48–72 h with an individual of either the same or different population with the same emergence date. We made four crosses, two within populations and two between populations: G1♀ × G1♂, G3♀ × G3♂, G1♀ × G3♂, and G3♀ × G1♂. To control for thelytoky, that is, females developing from unfertilized eggs (e.g. from *Wolbachia* infection), virgin females were also tested for each parasitoid population. For all crosses and controls, each female was provided with two infested blueberries containing approximately 10 first- and second-instar *D. suzukii* larvae, based on counts of initial host eggs laid in berries. After 3 days, females were removed and placed in 95% ethanol. Exposed host larvae were kept for 6 weeks during which adult flies should emerge in about 2 weeks or parasitoids should emerge in about 30 days. All emerged insects were counted and sexed. The remaining (assumed dead) host puparia were reconstituted in water for 1 day and then dissected under a microscope to determine the presence or absence of parasitoids. The numbers of parasitoid offspring produced and the percent parasitism were estimated from numbers of flies and wasps that emerged as adults as well as the dissected hosts with vs without parasitoids. Progeny sex ratios of parasitoids were estimated from the genders of adult wasps that emerged. Each cross was replicated 20 times, except for G3 ♀ × G3 ♂ which was replicated 10 times and unmated female controls which were replicated 5 times.

We used analyses of deviance to test the effects of cross type, including unmated females, on sex ratio, proportion parasitism, number of offspring per parasitoid female, and total emergences (flies plus parasitoids), using generalized linear models with the glm function in the STATS R package (R_Core_Team 2020) or the glm.nb function in the MASS R package (version 7.3-48; Venables and Ripley 2002). For these analyses, we chose the error distributions that gave the highest model probability calculated from the residual deviance divided by residual degrees of freedom (dfs) compared to a chi-square distribution (Littell *et al.* 1996). The binomial distribution gave the best fit for sex ratio, the normal distribution for proportion parasitism, and the negative binomial distribution for number of offspring per parasitoid female and total emergences. We calculated means and 95% asymptotic confidence intervals of the means with the emmeans R package (Lenth 2019).

Results

DNA sequencing

Sequencing of PacBio HiFi libraries provided 9–15 Gb of data in 1–2 million reads with mean lengths from 8 to 10 kb (Table 1). For two samples, G1-Tokyo and G3-Yunnan, this included data from ultra-low-input libraries, which included a PCR step during library preparation, as well as from low-input libraries, which did not include a PCR step. For the other samples, data were from low-input libraries alone.

Genome assemblies

Assembly of the genomes of these *Ganaspis* samples using the HiFiASM assembler gave sizes that varied from 1,015 to 1,067 Mb (Table 2), which are 4–9% larger than the flow cytometry estimate of 971 Mb for a closely related species, *G. xanthopoda* (Gokhman *et al.* 2011). The assemblies of G3-Nagano and G3-Yunnan were 4–5% larger than those of G1-BC, G1-Tokyo, and G1-Yunnan. N_{50} ranged from 295 to 2,223 kb and L_{50} from 132 to 882 contigs. N_{50} values were 1.1–8-fold higher and L_{50} values were 1.1–7-fold lower for the G1 assemblies than for the G3 assemblies, indicating greater contiguity for the G1 assemblies. Numbers of complete genes among the 5,991 genes in the BUSCO set for Hymenoptera varied from 4,798 to 5,364 among assemblies, which represents 80–90%

Table 1. Amount of sequence data in PacBio HiFi libraries made with single males from five populations of Ganaspis near brasiliensis.

	Gigabases				
Population	**Low**	**Ultra-low**	**Total**	***n* reads**	**Read length**
G1-BC	11	-	11	1,433,541	8,017
G1-Tokyo	6	9	15	1,987,486	7,567
G1-Yunnan	9	-	9	923,584	9,744
G3-Nagano	10	-	10	1,178,236	8,344
G3-Yunnan	5	7	12	1,544,834	7,873

Table 2. Metrics for assemblies of five genomes of *Ganaspis* near *brasiliensis* made with the HiFiASM assembler.

	Assembly				
Metric	**G1-BC**	**G1-Tokyo**	**G1-Yunnan**	**G3-Nagano**	**G3-Yunnan**
Total length (bp)	1,025,447,244	1,026,409,626	1,015,402,716	1,066,929,324	1,062,643,061
N_{50} (bp)	2,223,115	2,064,752	861,079	808,385	295,240
N_{90} (bp)	574,919	520,348	246,207	199,011	76,016
Largest contig (bp)	9,576,695	12,048,020	4,031,133	4,427,312	2,711,619
n contigs	1,037	1,059	2,135	2,653	5,635
L_{50}	132	140	354	396	882
L_{90}	479	523	1,175	1,391	3,918
Percent BUSCO	90	90	89	89	80

BUSCO percent is for complete genes in the Hymenoptera set of 5,999 genes

of the BUSCO set (Table 2). However, the percent complete, single-copy BUSCO genes varied from 68 to 89 (Supplementary Table 3). The lowest values of were for the G3-Yunnan assembly, which had higher numbers of duplicated and missing BUSCO genes than the other assemblies. The G3-Yunnan assembly also had the lowest N_{50} and highest L_{50}, which may have affected the percent complete, single-copy BUSCO genes.

Bacterial contamination

Using blastn with these assemblies and the NCBI nr database and then filtering hits for eukaryote vs bacterial sequences, we revealed that the assemblies had 3–13 contigs comprising 0.2–0.4% of the assembly lengths that harbored sequences that matched those from Bacteria (Supplementary Table 4). It is worth noting that, even though we restricted our searches to Bacteria and Insecta, 1–7 contigs had sequences that matched sequences identified as both Bacteria and Eukaryota, which can be explained by some NCBI accessions identified as Insecta being contaminated with sequences from Bacteria.

In the G1 assemblies, the bacterial sequences matched *Wolbachia* only. In the G3 assemblies, 61–85% of the bacterial sequences matched those from *Wolbachia*, and the other matches were with *Rickettsia* for the G3-Nagano assembly and *Rickettsia* plus six other bacterial genera for the G3-Yunnan assembly. The maximum contig sizes (0.9–2.3 Mb) for those that mapped to *Wolbachia* were close to the size range of the entire *Wolbachia* genome (0.75–1.7 Mb; Scholz *et al.* 2020). We removed the 3–13 contigs with bacterial sequences from the assemblies for all other analyses.

Mitochondrial genomes

Using blastn with our assemblies against the published mitochondrial genomes of *L. boulardi* (Oliveira *et al.* 2016) and *L. syphax* (Zhang *et al.* 2021), we found single contigs in each *Ganaspis* assembly with sequences that mapped to the mitochondrial genomes of the *Leptopilina* species with mapped lengths that matched the reported lengths for mitochondrial genomes of chalcidoids (Shu *et al.* 2022). We submitted these putative mitochondrial genomes to Mitos2 (Donath *et al.* 2019) and found complete mitochondrial genomes on single contigs from each assembly. The contigs in two assemblies were too large because of duplications and had to be trimmed based on the results from Mitos2. These mitochondrial genomes had sequences coding for 13–14 mitochondrial genes, 4–5 ribosomal RNAs, and 20–22 transfer RNAs (Supplementary Table 5), and the orders and lengths of these features matched well those in other figitids (Supplementary Fig. 1; Shu *et al.* 2022).

We made a neighbor-joining tree with distances between the COI sequences captured from our mitochondrial assemblies and sequences/sub-sequences of the COI gene for *G. brasiliensis* from the NCBI nr nucleotide database (ncbi.nlm.nih.gov; queried November 2022). This analysis placed the COI sequences from the mitochondrial genomes next to those from the expected lineages (Supplementary Fig. 2). The reads and assembly of G1-BC were closest to those from G1-Tokyo. However, another sample from BC that did not provide sufficient quality sequence for assembly had a COI sequence closest to those from G3-Nagano. Thus, the hypothesis, based on previous COI sequencing, that *Ganaspis* near *brasiliensis* populations in BC are composed only of the G1 lineage should be revisited.

Repeats and TEs

Analyses of repetitive DNA in our assemblies of *Ganaspis* near *brasiliensis* with Jellyfish (Marçais and Kingsford 2011) and GenomeScope (Ranallo-Benavidez *et al.* 2020) identified 218–433 Mb (21–42% of the assembly sequences) as repetitive (Supplementary Table 6), much of which is likely to be relic transposons.

Analysis of TEs in these assemblies with the EDTA pipeline (Ou *et al.* 2019) identified 484–598 Mb of sequence as TEs, which comprised 48–50% of the assembly sequences (Supplementary Table 7). These TEs included 217–318 Mb of LTR retrotransposons, 107–190 Mb of terminal inverted repeat (TIR) DNA TEs, 39–55 Mb of non-TIR DNA TEs, and 14–22 Mb of miniature inverted repeat transposable element DNA. The two most common transposon families were LTR *Gypsy*, comprising 10–11% of the assembly sequences, and TIR *Mutator*, comprising 5–7% of the assembly sequences. However, unknown LTR TEs comprised 13–16% on the assembly sequences. Among the 1.5–1.8 million TEs in each assembly, only 18–21 thousand were intact (containing the sequences needed for transposition), i.e. 1% of all TEs, and these comprised 2–3% of the assembly lengths and thus about one-tenth of the total in TE sequences. Intact TE sequences were found on 37–80% of assembly contigs. Most of the TE sequences appear to be inactive and may eventually be removed from the genomes by natural selection. However, all the TE sequences may have roles beyond transposition that can affect evolution (Bourque *et al.* 2018; Gilbert *et al.* 2021). The G3 assemblies harbored 5–17% more TEs and 5–19% greater total TE sequence, which explains in part the differences in assembly sizes between the lineages.

Sequence divergence between genome assemblies

When we mapped reads from each sample to assemblies of the other samples using minimap2 with the hifi option for PacBio HiFi data (Li 2018) and analyzed the mapping results with

genomecov in BEDtools (Quinlan and Hall 2010), we found that the proportion of the assembly with mapped reads was higher within each lineage than between lineages (Fig. 1a). Within the G1 lineage, 96–97% of each assembly had mapped reads, and within the G3 lineage, 91–93% of each assembly had mapped reads (Supplementary Table 8a). However, only 70–77% of each assembly had mapped reads between lineages. We found more SNP loci and higher density of SNP loci between lineages than within lineages (Fig. 1, b and c; Supplementary Table 8, b and c). The assemblies of the G1 lineages differed by 5 SNP loci per kilobase, and the assemblies of the G3 lineages differed by 13 SNP loci per kilobase. However, the assemblies of G1 vs G3 lineages differed by 18–24 SNP loci per kilobase, which is two to four times greater than the within-lineage differences.

Gene detection and annotation

Using Augustus with the *Nasonia* gene model (Stanke *et al.* 2008), we identified 61–69 thousand genes in the assemblies of *Ganaspis* near *brasiliensis* (Table 3). There were 10–12% more genes in the assemblies of the G3 lineage than those of the G1 lineage, which is not surprising given that the assemblies of the G3 lineage were 4–5% longer. The genes comprised 184–199 Mb which was 18–19% of each of the assemblies. The mean gene lengths were 2,807–3,038 bp with 2.4–2.7 exons, mean coding sequence lengths 1,214–1,231 bp, and mean intronic lengths 3,598–3,820 bp.

Using blastp with these genes and the NCBI nr database, we found 81–89% had homologs (Table 3). Using blast2GO (Götz *et al.* 2008) to categorize the top hits and assign GO annotations, we found 54–66% with GO mappings and 21–36% with GO annotations of molecular function, biological process, and/or cellular location. We also used InterProScan (Jones *et al.* 2014) to assess protein domain homology and found 76–99% of the genes had protein domain homologies. The assemblies of the G3 lineage had 5–23% more genes with protein domain homologies than in the assemblies of the G1 lineage.

Orthologs and ortholog divergence

Using Orthofinder (Emms and Kelly 2019), we identified 33–41k orthogroups per assembly in these *Ganaspis* populations with (Supplementary Fig. 3 and Table 9). These orthogroups contained 50–56k genes per assembly, which is 73–91% of the genes in each assembly. There were more orthogroups, and more genes were assigned to them in the G1 lineage than in the G3 lineages. There were 111–480 assembly-specific orthogroups that harbored 591–1,375 genes. There were 5,664–5,713 unassigned genes (not in orthogroups) in the G1 assemblies, which is 9% of all genes in these assemblies, but there were 16,049–18,606 unassigned genes in the G3 assemblies, which is 24–27% of all genes in these assemblies. Within lineages, 88–93% of orthogroups were shared, but between lineages, only 51–63% of orthogroups were shared (Fig. 2; Supplementary Table 10).

Orthofinder generated a molecular phylogeny based on genes in 17,300 universal orthogroups, i.e. those present in all assemblies (Fig. 3). This phylogeny supports the separation of the G1 and G3 genome assemblies into different clades.

Population crosses

The sex ratio of offspring differed among crosses within vs between the G1-Tokyo and G3-Yunnan populations (model deviance = 130.6, df = 5; residual deviance = 32.8, df = 68; P [χ^2] < 0.0001). Within-population crosses produced 50:50 sex ratios, but between-population crosses produced no females, as did unmated females (Fig. 4a). The ratio of adult parasitoids to adult flies differed between crosses (model deviance = 69.5, df = 5; residual deviance = 230.5, df = 71; P [χ^2] < 0.0001), as did number of parasitoid offspring per female (model deviance = 18.5, df = 5; residual deviance = 96.7, df = 74; P [χ^2] = 0.002). The proportion parasitism was lower for between-population crosses than for the respective within-population crosses and was like that produced by unmated females (Fig. 4b), and the number of offspring was lower for between-population crosses than for the respective within-population crosses (Fig. 4c). However, total emergences of flies plus parasitoids did not differ among crosses (model deviance = 0.2, df = 5; residual deviance = 90.2, df = 68; P [χ^2] = 0.99), which shows no overall difference in total survival among crosses (Fig. 4d).

Discussion

The amount of sequence divergence between the assemblies of the G1 and G3 lineages reported here suggests they are different species. The percent of assemblies with mapped reads was 90–97 within lineages but only 70–77 between lineages, showing much greater divergence between lineages than within lineages. Furthermore, the assemblies differed by 5–13 SNP loci per kilobase within lineages but by 18–24 SNP loci per kilobase between lineages. The assemblies of G1 and G3 lineages from China were made with insects originally collected 12 km apart and thus sympatric. The assemblies of the G1 and G3 lineages from Japan were made with insects originally collected 143 km apart, which, although further apart than the collection sites of G1 and G3 in China, appear sufficiently close that gene flow likely. Indeed, the distances between the samples of the G1 lineage were >3,000 km and between the samples of the G3 lineage were also 3,000 km, including hundreds of kilometers of sea, yet the assemblies were much more similar within lineages than between lineages, suggesting gene flow across much longer distances than tens to hundreds of kilometers.

It would be useful to compare the genomic-level sequence divergences reported here with differences between genomes of other pairs or groups of closely related parasitoids. Unfortunately, out of 25 recent papers we found on the genetic divergence between closely related parasitoid species, few compared divergence at the genome level or across many loci with most involving comparisons among sequences of one or a few genes or of length polymorphisms, not sequence differences. However, research on closely related species in the *varipes* complex of the genus *Aphelinus* (Hymenoptera: Aphelinidae) does provide an apt comparison. Analysis of sequence divergences among 17 Asian populations of *Aphelinus certus* Yasnosh using restriction site-associated DNA sequencing) found 892 SNP loci distributed across the *A. certus* genome that gave a SNP density of 12 loci per kilobase (Hopper *et al.* 2019). The amount of divergence among *A. certus* populations was the same as that between populations of the G3 lineage of *Ganaspis* near *brasiliensis*, greater than the divergence between populations of the G1 lineage, and much less than the divergence between the G1 and G3 lineages. Mapping reads used to an assembly the 334 Mb genome of *A. certus* (Wittmeyer *et al.* 2021) to the genomes of four other species in the *varipes* complex showed divergences ranging from 12 to 19 SNP loci per kilobase (unpublished data). *Aphelinus certus* differed from two species, *Aphelinus atriplicis* Kurdjumov and *Aphelinus varipes* Förster, at 12 SNP loci per kilobase and was partially reproductively compatible with these species in the laboratory (Heraty *et al.* 2007). However, they are phylogenetically distinct and allopatric from *A. certus*, being separated from it by more than 5,000 km of deserts and mountains (Heraty *et al.* 2007), and have different host ranges than *A. certus* (Hopper *et al.* 2017). Two other *Aphelinus*

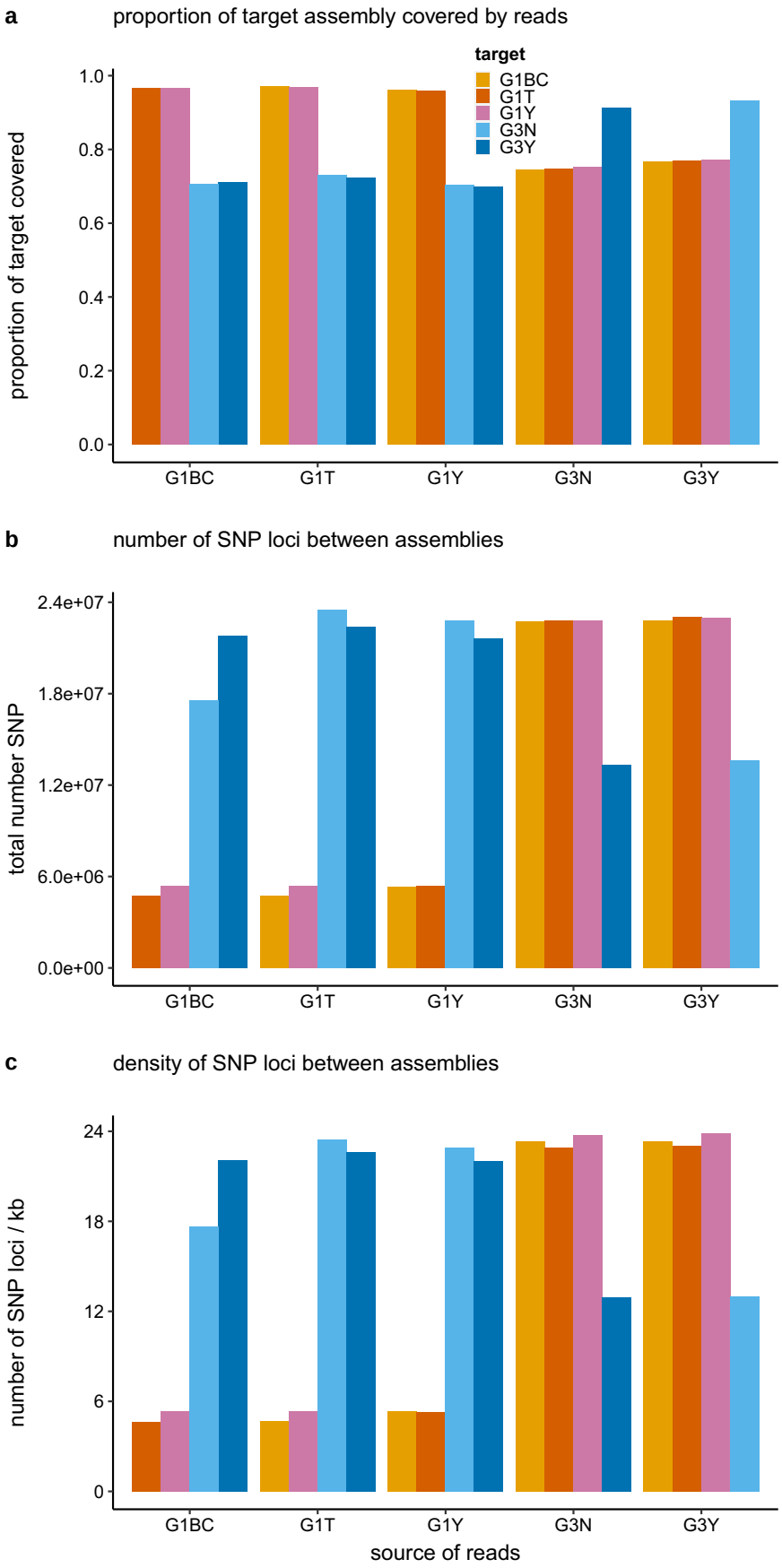

Fig. 1. Sequence divergence between genomes of *Ganaspis* near *brasiliensis*. a) percentage of each assembly with mapped reads from other populations; b) number of SNP loci for reads that mapped to each assembly; c) density of SNP loci per kilobase of sequence for reads that mapped to each assembly.

Table 3. Metrics of gene number, structure, and annotation for assemblies of the genomes of *Ganaspis* near *brasiliensis*.

	Assembly				
Metric	**G1-BC**	**G1-Tokyo**	**G1-Yunnan**	**G3-Nagano**	**G3-Yunnan**
Number of genes	61,261	61,706	60,746	69,033	68,309
Total length in genes (bp)	186,140,805	185,555,010	183,989,980	198,931,898	191,765,255
Percent of assembly	18%	18%	18%	19%	18%
Mean:					
Gene length (bp)	3,038	3,007	3,029	2,882	2,807
Exons per gene	2.7	2.7	2.7	2.4	2.4
Exon length (bp)	455	457	453	502	510
Total exon length (bp)	1,224	1,215	1,214	1,226	1,231
Introns per gene	3.4	3.4	3.4	3.3	3.2
Intron length (bp)	1,073	1,080	1,078	1,145	1,107
Total intron length (bp)	3,692	3,673	3,676	3,820	3,589
n proteins with:					
GenBank homologs	49,613	53,412	52,412	57,570	60,640
GO mappings	40,425	33,325	32,652	42,402	40,801
GO annotations	21,938	14,007	13,886	19,846	14,257
InterProScan annotations	46,728	46,722	46,216	55,726	67,624
Percent proteins with:					
GenBank homologs	81%	87%	86%	83%	89%
GO mappings	66%	54%	54%	61%	60%
GO annotations	36%	23%	23%	29%	21%
InterProScan annotations	76%	76%	76%	81%	99%

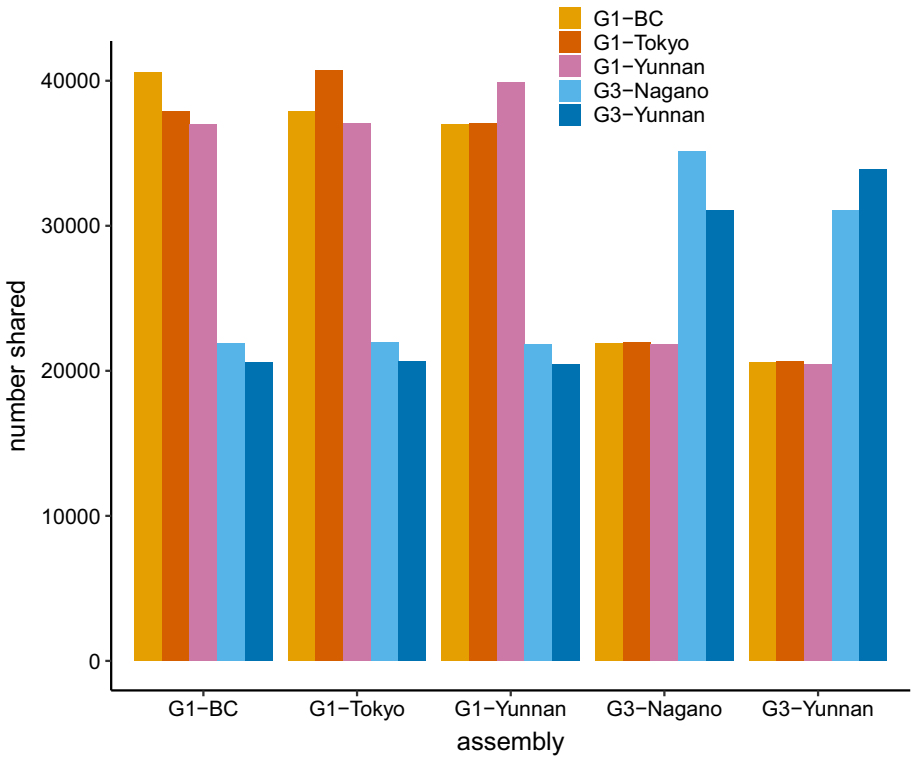

Fig. 2. Number of orthogroups shared among assemblies of genomes of *Ganaspis* near *brasiliensis*.

species, *Aphelinus hordei* (Kurdjumov) and *Aphelinus kurdjumovi* Mercet, that differ from *A. certus* at 17 and 19 SNP loci per kilobase, respectively (comparable to the divergence between the G1 and G3 lineages of *Ganaspis* near *brasiliensis*), are reproductively incompatible with *A. certus* (Heraty *et al.* 2007) and have different host ranges than *A. certus* (Hopper *et al.* 2017). It is worth noting that several species in the *varipes* complex were synonymized based on morphology (Ferrière 1965; Graham 1976) but later found to be reproductively incompatible in the laboratory and phylogenetically distinct, based on sequences of six genes (Heraty *et al.* 2007). Two sibling species of *Oobius* (Hymenoptera: Encyrtidae) provide another example of sequence divergence between closely related parasitoids (Yao *et al.* 2016). Mapping reads from *Oobius agrili* Zhang & Huang from China to coding sequences from *Oobius primorskyensis* Yao & Duan from Russia revealed 38 SNP loci per kilobase. These species also differed in diapause patterns and showed subtle differences in morphology, which together with sequences differences led them to being described as different species.

Although entities with highly divergent sequences in the above studies were valid species, so were species that showed the same level of divergence as among populations within species. Thus, it appears that sequence divergence alone is not adequate for determining species status. However, a combination of sequence divergence, reproductive incompatibility, and differences in behavior and ecology can distinguish species hard to tell apart with morphology.

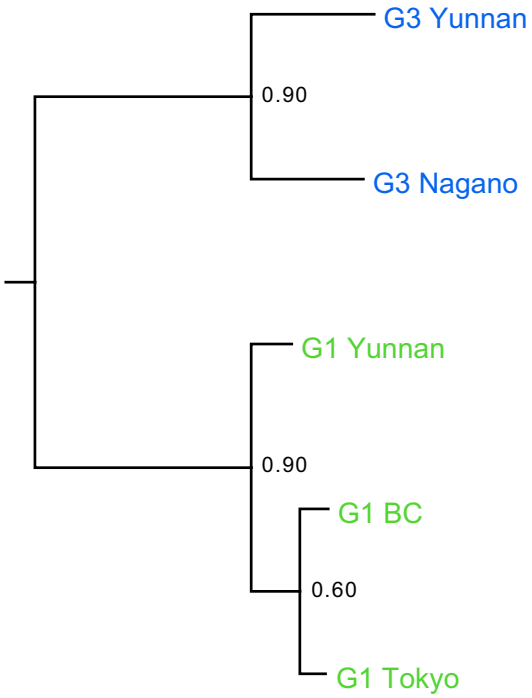

Fig. 3. Molecular phylogeny of five populations of *Ganaspis* near *brasiliensis* based on genes in 17,300 universal orthogroups. Numbers on the nodes are the percentages of gene trees that supported the node.

Besides sequence divergences between the assemblies of these lineages of *Ganaspis* near *brasiliensis*, several other comparisons of their genomes show differences between the lineages. The assemblies of the G3 lineage were longer than for the G1 lineage, which can be explained in part by greater content of TEs. The values for N_{50} were higher and those for L_{50} were lower for the G1 assemblies than for the G3 assemblies, indicating greater contiguity for the G1 assemblies. The identities of bacterial contaminants differed between the G1 and G3 assemblies. More genes were identified in the G3 assemblies than in the G1 assemblies, and a higher percentage of the genes in the G3 assemblies had InterProScan annotations of protein domains. A larger percentage of orthogroups were shared within lineages than between lineages. There were more orthogroups and more genes assigned to them in the G1 assemblies than in the G3 assemblies.

Two aspects of these assemblies that did not differ between lineages were the lengths and organization of the mitochondrial genomes, but these are conserved not only among figitid species but among cynipoid species (Shu *et al.* 2022). However, COI sequences did differ between lineages and were consistent with previous results.

Perhaps, the strongest evidence that the G1 and G3 lineages are indeed different species are the results from laboratory crosses. The crosses reported here showed reproductive incompatibility between the G3-Yunnan and G1-Tokyo populations of *Ganaspis* near *brasiliensis*. Seehausen *et al.* (2020) crossed the G1-Tokyo population with merged G3-4 populations from Nagano and found no female progeny. Our results combined with theirs suggest that

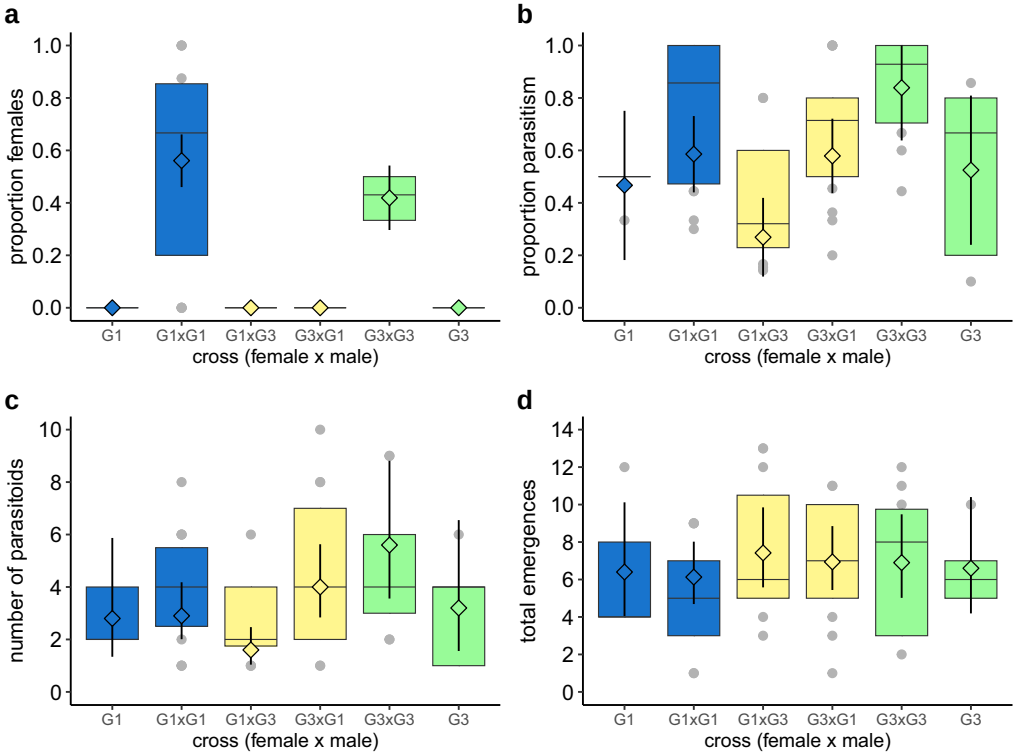

Fig. 4. Results of crosses between G1-Tokyo and G3-Yunnan of *Ganaspis* near *brasiliensis*: a) proportion female offspring, b) proportion of flies parasitized, c) offspring per female, d) total number of flies and parasitoids; G1 = G1-Tokyo; G3 = G3-Yunnan. Boxes are quantiles of values, horizontal lines are medians, gray dots are outliers, diamonds are means, and vertical bars are 95% confidence intervals of means.

both sympatric and allopatric G1 and G3 populations are reproductively incompatible. Given that the assemblies of both lineages harbor *Wolbachia*, the original isolation may have involved *Wolbachia* infections because *Wolbachia* can cause reproductive isolation between insects in general and parasitoids in particular (for review, see Shropshire *et al.* 2020). However, Turelli *et al.* (2022) argued, based on the recency of many *Wolbachia* infections compared to the age of the species infected, it is unlikely that *Wolbachia* infections were involved in speciation, although such infections may reinforce isolation subsequent to speciation. Males and females from different lineages did not mate in the crosses reported here or those reported by Seehausen *et al.* (2020), showing that reproductive incompatibility is behavioral, with females rejecting courtship by heterospecific males.

Conclusions

One of us (MLB) is describing subtle morphological differences between the G1 and G3 lineages of *Ganaspis* near *brasiliensis* in order to assign species names. Such morphological differences alone cannot determine species status, but our combined data on sequence divergence, differences in genome size, ortholog divergence, reproductive incompatibility, and differences in host ranges and microhabitat preferences show that these lineages are different species.

For biological control introductions, the most important differences between these lineages are the differences in host and microhabitat specificities. The G1 lineage of *Ganaspis* near *brasiliensis* is more specific to *D. suzukii* than other lineages and is specialized on *Drosophila* larvae feeding in intact fruits. Because *D. suzukii* in Europe and North America is the only drosophilid to attack intact fruit, choice of the G1 lineage for biological control introductions, which was the case in the USA and Europe, was wise. More broadly, our results provide support for the growing recognition that cryptic species that differ in important biological traits must be considered in biological control introductions.

Data availability

DNA sequence data and assemblies are archived at NCBI (www.ncbi.nlm.nih.gov) under BioProject PRJNA1035088 with individual Biosample and Assembly accessions as follows: G1-BC = SAMN38082132, GCA_037103525.1; G1-Tokyo = SAMN38082384, GCA_037103515.1; G1-Yunnan = SAMN38082405; GCA_037103535.1; G3-Nagano = SAMN38082482, GCA_037103505.1; and G3-Yunnan = SAMN38082483, GCA_037103545.1. Gene annotation (gff) files and variant calls (vcf) are archived at figshare (https://doi.org/10.25387/g3.24585591). The data from the crosses are archived in a spreadsheet (xlsx) at Ag Data Commons DOI (https://doi.org/10.15482/USDA.ADC/25208948.v1).

Supplemental material is available at G3 online.

Acknowledgments

The authors thank Karol Miaskiewicz at University of Delaware Bioinformatics Core Facility for help with using the BIOMIX computational cluster and Amanda Stout at BIIRU for assistance with rearing and crosses. 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 USDA. The USDA is an equal opportunity provider and employer. The CABI as an international intergovernmental not-for-profit organization gratefully acknowledges the generous support received from our many donors, sponsors, and partners. In particular, the authors thank their Member Countries for their vital financial and strategic contributions.

Funding

This research was supported by the United States Department of Agriculture, Agricultural Research Service, Project Numbers 8010-22000-032-00D and 8010-22000-033-00D, and the USDA Specialty Crop Research Initiative Award No. 2020-51181-32140. Support from the University of Delaware Bioinformatics Core Facility, including use of the BIOMIX computational cluster, was made possible by the Delaware IDeA Network of Biomedical Research Excellence (National Institutes of Health grant P20 GM103446).

Conflicts of interest

The author(s) declare no conflicts of interest.

Literature cited

Abram PK, Franklin MT, Hueppelsheuser T, Carrillo J, Grove E, Eraso P, Acheampong S, Keery L, Girod P, Tsuruda M, *et al.* 2022. Adventive larval parasitoids reconstruct their close association with spotted-wing *Drosophila* in the invaded North American range. Environ Entomol. 51(4):670–678. doi:10.1093/ee/nvac019.

Abram PK, McPherson AE, Kula R, Hueppelsheuser T, Thiessen J, Perlman SJ, Curtis CI, Fraser JL, Tam J, Carrillo J, *et al.* 2020. New records of *Leptopilina*, *Ganaspis*, and *Asobara* species associated with *Drosophila suzukii* in North America, including detections of L. japonica and G. brasiliensis. J Hymenopt Res. 78:1–17. doi:10.3897/jhr.78.55026.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17):3389–3402. doi:10.1093/nar/25.17.3389.

Asplen MK, Anfora G, Biondi A, Choi DS, Chu D, Daane KM, Gibert P, Gutierrez AP, Hoelmer KA, Hutchison WD, *et al.* 2015. Invasion biology of spotted wing Drosophila (*Drosophila suzukii*): a global perspective and future priorities. J Pest Sci (2004). 88(3):469–494. doi:10.1007/s10340-015-0681-z.

Beers EH, Beal D, Smytheman P, Abram PK, Schmidt-Jeffris R, Moretti E, Daane KM., Looney C, Lue C-H, Buffington M, *et al.* 2022. First records of adventive populations of the parasitoids *Ganaspis brasiliensis* and *Leptopilina japonica* in the United States. J Hymenopt Res. 91:11–25. doi:10.3897/jhr.91.82812.

Bodenhofer U, Bonatesta E, Horejš-Kainrath C, Hochreiter S. 2015. Msa: an R package for multiple sequence alignment. Bioinformatics. 31(24):3997–3999. doi:10.1093/bioinformatics/btv494.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, *et al.* 2018. Ten things you should know about transposable elements. Genome Biol. 19(1):199. doi:10.1186/s13059-018-1577-z.

Buffington ML, Forshage M. 2016. Redescription of *Ganaspis brasiliensis* (ihering, 1905), new combination, (Hymenoptera: Figitidae) a natural enemy of the invasive *Drosophila suzukii* (Matsumura, 1931) (Diptera: Drosophilidae). Proc Entomol Soc Wash. 118 (1):1–13. doi:10.4289/0013-8797.118.1.1.

Charif D, Lobry JR. 2007. Seqinr 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Roman HE,

Vendruscolo M, editors. Structural Approaches to Sequence Evolution: Molecules, Networks, Populations. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 207–232.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 18(2):170–175. doi:10.1038/s41592-020-01056-5.

Daane KM, Wang XG, Biondi A, Miller B, Miller JC, Riedl H, Shearer PW, Guerrieri E, Giorgini M, Buffington M, *et al.* 2016. First exploration of parasitoids of Drosophila suzukii in South Korea as potential classical biological agents. J Pest Sci (2004). 89(3):823–835. doi:10.1007/s10340-016-0740-0.

Daane KM, Wang XG, Hogg BN, Biondi A. 2021. Potential host ranges of three Asian larval parasitoids of *Drosophila suzukii*. J Pest Sci (2004). 94(4):1171–1182. doi:10.1007/s10340-021-01368-1.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* 2011. The variant call format and VCFtools. Bioinformatics. 27(15):2156–2158. doi:10.1093/bioinformatics/btr330.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al.* 2021. Twelve years of SAMtools and BCFtools. GigaScience. 10(2):giab008. doi:10.1093/gigascience/giab008.

Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, Middendorf M, Bernt M. 2019. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res. 47(20):10543–10552. doi:10.1093/nar/gkz833.

Du H, Liang C. 2019. Assembly of chromosome-scale contigs by efficiently resolving repetitive sequences with long reads. Nat Commun. 10(1):1–10. doi:10.1038/s41467-019-13355-3.

Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics. 9(1):18. doi:10.1186/1471-2105-9-18.

Emms DM, Kelly S. 2017. STRIDE: species tree root inference from gene duplication events. Mol Biol Evol. 34(12):3267–3278. doi:10.1093/molbev/msx259.

Emms DM, Kelly S. 2018. STAG: species tree inference from all genes. bioRxiv 267914. https://doi.org/10.1101/267914, preprint: not peer reviewed.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20(1):238. doi:10.1186/s13059-019-1832-y.

Ferrière C. 1965. Hymenoptera Aphelinidae d'Europe et du Bassin Mediterraneen. Paris: Masson and Associates.

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

Gallardo FE, Funes CF, Reche V, Kirschbaum DS, Ovruski SM, Buffington ML. 2022. First record and distribution of *Ganaspis brasiliensis* (Hymenoptera: Figitidae: Eucoilinae), a parasitoid of *Drosophila suzukii* (Diptera: Drosophilidae) in Argentina. Neotrop Entomol. 51(1):164–169. doi:10.1007/s13744-021-00912-z.

Gilbert C, Peccoud J, Cordaux R. 2021. Transposable elements and the evolution of insects. Annu Rev Entomol. 66(1):355–372. doi:10.1146/annurev-ento-070720-074650.

Giorgini M, Wang XG, Wang Y, Chen FS, Hougardy E, Zhang H-M, Chen Z-Q, Chen H-Y, Liu C-X, Cascone P, *et al.* 2019. Exploration for native parasitoids of *Drosophila suzukii* in China reveals a diversity of parasitoid species and narrow host range of the dominant parasitoid. J Pest Sci (2004). 92(2):509–522. doi:10.1007/s10340-018-01068-3.

Girod P, Borowiec N, Buffington M, Chen GH, Fang Y, Kimura MT, Peris-Felipo FJ, Ris N, Wu H, Xiao C, *et al.* 2018a. The parasitoid complex of *D. suzukii* and other fruit feeding *Drosophila* species in Asia. Sci Rep. 8(1):11839. doi:10.1038/s41598-018-29555-8.

Girod P, Lierhmann O, Urvois T, Turlings TCJ, Kenis M, Haye T. 2018b. Host specificity of Asian parasitoids for potential classical biological control of *Drosophila suzukii*. J Pest Sci (2004). 91(4):1241–1250. doi:10.1007/s10340-018-1003-z.

Gokhman VE, Johnston JS, Small C, Rajwani R, Hanrahan SJ, Govind S. 2011. Genomic and karyotypic variation in *Drosophila* parasitoids (Hymenoptera, Cynipoidea, Figitidae). Comp Cytogenet. 5(3):211–221. doi:10.3897/compcytogen.v5i3.1435.

Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talon M, Dopazo J, Conesa A. 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36(10):3420–3435. doi:10.1093/nar/gkn176.

Graham MWRDV. 1976. British species of *Aphelinus* with notes and descriptions of other European Aphelinidae (Hymenoptera). Syst Entomol. 1(2):123–146. doi:10.1111/j.1365-3113.1976.tb00345.x.

Heraty JH, Woolley JB, Hopper KR, Hawks DL, Kim J-W, Buffington M. 2007. Molecular phylogenetics and reproductive incompatibility in a complex of cryptic species of aphid parasitoids. Mol Phylogenet Evol. 45(2):480–493. doi:10.1016/j.ympev.2007.06.021.

Hopper KR, Lanier K, Rhoades JH, Coutinot D, Mercadier G, Ramualde N, Roche M, Woolley JB, Heraty JM. 2017. Host specificity of *Aphelinus* species considered for introduction to control *Diuraphis noxia*. Biol Control. 107:21–32. doi:10.1016/j.biocontrol.2017.01.006.

Hopper KR, Oppenheim SJ, Kuhn KL, Lanier K, Hoelmer KA, Heimpel GE, Meikle WG, O'Neil RJ, Voegtlin DG, Wu K, *et al.* 2019. Counties not countries: variation in host specificity among populations of an aphid parasitoid. Evol Appl. 12(4):815–829. doi:10.1111/eva.12759.

Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, *et al.* 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. 30(9):1236–1240. doi:10.1093/bioinformatics/btu031.

Kacsoh BZ, Schlenke TA. 2012. High hemocyte load is associated with increased resistance against parasitoids in *Drosophila suzukii*, a relative of *D. melanogaster*. PLoS One. 7(4):e34721. doi:10.1371/journal.pone.0034721.

Kasuya N, Mitsui H, Ideo S, Watada M, Kimura MT. 2013. Ecological, morphological and molecular studies on Ganaspis individuals (Hymenoptera: Figitidae) attacking Drosophila suzukii (Diptera: Drosophilidae). Appl Entomol Zool (Jpn). 48(1):87–92. doi:10.1007/s13355-012-0156-0.

Kimura MT, Suwito A. 2012. Diversity and abundance of frugivorous drosophilids and their parasitoids in Bogor, Indonesia. J Nat Hist. 46(31–32):1947–1957. doi:10.1080/00222933.2012.707239.

Kimura MT, Suwito A. 2015. Altitudinal patterns of abundances and parasitism in frugivorous drosophilids in west Java, Indonesia. J Nat Hist. 49(27–28):1627–1639. doi:10.1080/00222933.2015.1005709.

Kirschbaum DS, Funes CF, Buonocore-Biancheri MJ, Suárez L, Ovruski SM. 2020. The biology and ecology of Drosophila suzukii (Diptera: Drosophilidae). In: Garcia FRM, editors. Drosophila suzukii Management. Cham: Springer International Publishing. p. 41–91.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 37(5):540–546. doi:10.1038/s41587-019-0072-8.

Kuhl H, Li L, Wuertz S, Stöck M, Liang XF, Klopp C. 2020. CSA: a high-throughput chromosome-scale assembly pipeline for vertebrate genomes. GigaScience. 9(5):1–14. doi:10.1093/gigascience/giaa034.

Lee JC, Wang XG, Daane KM, Hoelmer KA, Isaacs R, Sial AA, Walton VM. 2019. Biological control of spotted-wing *Drosophila* (Diptera: Drosophilidae)-current and pending tactics. J Integr Pest Manag. 10(1):1–9. doi:10.1093/jipm/pmz012.

Lenth RV. 2019. emmeans: Estimated marginal means, aka least-squares means. R package version 1.3.4. [accessed 31 Jan 2024]. https://CRAN.R-project.org/package=emmeans

Li H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 27(21): 2987–2993. doi:10.1093/bioinformatics/btr509.

Li H. 2016. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics. 32(14):2103–2110. doi:10.1093/bioinformatics/btw152.

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

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

Littell RC, Milliken GA, Stroup WW, Wolfinger RD. 1996. Generalized linear mixed models, chapter 10. In: SAS System for Mixed Models. Cary, NC: SAS Institute Inc. p. 423–460.

Lu J, Salzberg SL. 2018. Removing contaminants from databases of draft genomes. PLoS Comput Biol. 14(6):e1006277. doi:10.1371/journal.pcbi.1006277.

Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 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(10): 4647–4654. doi:10.1093/molbev/msab199.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 27(6): 764–770. doi:10.1093/bioinformatics/btr011.

Merchant S, Wood DE, Salzberg SL. 2014. Unexpected cross-species contamination in genome sequencing projects. PeerJ. 2:e675. doi:10.7717/peerj.675.

Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics. 34(13):i142–i150. doi:10.1093/bioinformatics/bty266.

Mitsui H, Van Achterberg K, Nordlander G, Kimura MT. 2007. Geographical distributions and host associations of larval parasitoids of frugivorous Drosophilidae in Japan. J Nat Hist. 41(25–28): 1731–1738. doi:10.1080/00222930701504797.

NCBI Resource Coordinators. 2016. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 44(D1):D7–D19. doi:10.1093/nar/gkv1290.

Nomano FY, Kasuya N, Matsuura A, Suwito A, Mitsui H, Buffington ML, Kimura MT. 2017. Genetic differentiation of *Ganaspis brasiliensis* (Hymenoptera: Figitidae) from east and Southeast Asia. Appl Entomol Zool (Jpn). 52(3):429–437. doi:10.1007/s13355-017-0493-0.

Oliveira DS, Gomes T, Loreto ELS. 2016. The rearranged mitochondrial genome of *Leptopilina boulardi* (Hymenoptera: Figitidae), a parasitoid wasp of *Drosophila*. Genet Mol Biol. 39(4):611–615. doi:10.1590/1678-4685-gmb-2016-0062.

Ou S, Jiang N. 2018. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. Plant Physiol. 176(2):1410–1422. doi:10.1104/pp.17.01310.

Ou S, Su W, Liao Y, Chougule K, Agda JRA, Hellinga AJ, Lugo CSB, Elliott TA, Ware D, Peterson T, *et al.* 2019. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. Genome Biol. 20(1):275. doi:10.1186/s13059-019-1905-y.

Paradis E. 2012. Analysis of Phylogenetics and Evolution with R. New York (NY): Springer.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26(6):841–842. doi:10.1093/bioinformatics/btq033.

R_Core_Team. 2020. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 11(1):1432. doi:10.1038/s41467-020-14998-3.

Reeve MA, Seehausen ML. 2019. Discrimination between Asian populations of the parasitoid wasp Ganaspis cf. brasiliensis using a simple MALDI-TOF MS-based method for use with insects. Biol Methods Protoc. 4(1):bpz002. doi:10.1093/biomethods/bpz002.

Rossi-Stacconi MVR, Wang XG, Tonina L, Daane KM, Biondi A, Stahl J, Buffington ML, Anfora G, Hoelmer KA. 2022. Methods for rearing the parasitoid *Ganaspis brasiliensis*, a promising biological control agent for the invasive *Drosophila suzukii*. J Vis Exp. 184: e63898.

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

Sanchez-Gonzalez JA, Lomeli-Flores JR, Rodriguez-Leyva E, Arredondo-Bernal HC, Gonzalez-Cabrera J. 2020. *Drosophila suzukii* response to *Leptopilina boulardi* and *Ganaspis brasiliensis* parasitism. B Insectol. 73(2):209–215.

Schilthuizen M, Nordlander G, Stouthamer R, Van Alphen JJM. 1998. Morphological and molecular phylogenetics in the genus Leptopilina (Hymenoptera: Cynipoidea: Eucoilidae). Syst Entomol. 23(3):253–264. doi:10.1046/j.1365-3113.1998.00049.x.

Scholz M, Albanese D, Tuohy K, Donati C, Segata N, Rota-Stabelli O. 2020. Large scale genome reconstructions illuminate Wolbachia evolution. Nat Commun. 11(1):5235. doi:10.1038/s41467-020-19016-0.

Seehausen ML, Ris N, Driss L, Racca A, Girod P, Warot S, Borowiec N, Toševski I, Kenis M 2020. Evidence for a cryptic parasitoid species reveals its suitability as a biological control agent. Sci Rep. 10(1): 19096. doi:10.1038/s41598-020-76180-5.

Seehausen ML, Valenti R, Fontes J, Meier M, Marazzi C, Mazzi D, Kenis M. 2022. Large-arena field cage releases of a candidate classical biological control agent for spotted wing drosophila suggest low risk to non-target species. J Pest Sci (2004). 95(3):1057–1065. doi:10.1007/s10340-022-01487-3.

Shi JM, Liang C. 2019. Generic repeat finder: a high-sensitivity tool for genome-wide de novo repeat detection. Plant Physiol. 180(4): 1803–1815. doi:10.1104/pp.19.00386.

Shropshire JD, Leigh B, Bordenstein SR. 2020. Symbiont-mediated cytoplasmic incompatibility: what have we learned in 50 years? Elife 9:e61989. doi:10.7554/eLife.61989.

Shu XH, Li ZK, Yuan RZ, Tang P, Chen XX. 2022. Novel gene rearrangements in the mitochondrial genomes of cynipoid wasps (Hymenoptera: Cynipoidea). Genes (Basel). 13(5):914. doi:10.3390/genes13050914.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 24(5):637–644. doi:10.1093/bioinformatics/btn013.

Su WJ, Gu X, Peterson T. 2019. TIR-Learner, a new ensemble method for TIR transposable element annotation, provides evidence for

abundant new transposable elements in the maize genome. Mol Plant. 12(3):447–460. doi:10.1016/j.molp.2019.02.008.

Tait G, Mermer S, Stockton D, Lee J, Avosani S, Abrieux A, Anfora G, Beers E, Biondi A, Burrack H, *et al.* 2021. *Drosophila suzukii* (Diptera: Drosophilidae): a decade of research towards a sustainable integrated pest management program. J Econ Entomol. 114(5): 1950–1974. doi:10.1093/jee/toab158.

Turelli M, Katznelson A, Ginsberg PS. 2022. Why *Wolbachia*-induced cytoplasmic incompatibility is so common. Proc Natl Acad Sci USA. 119(47):e2211637119. doi:10.1073/pnas.2211637119.

Vaser R, Šikić M. 2020. Raven: a de novo genome assembler for long reads. bioRxiv 2020.08.07.242461. https://doi.org/10.1101/2020.08.07.242461, preprint: not peer reviewed.

Venables WN, Ripley BD. 2002. Modern Applied Statistics with S. 4th ed. New York: Springer.

Wang XG, Lee JC, Daane KM, Buffington ML, Hoelmer KA. 2020. Biological control of *Drosophila suzukii*. CAB Rev. 15(54):1–19. doi: 10.1079/PAVSNNR202015054.

Wang XG, Nance AH, Jones JML, Hoelmer KA, Daane KM. 2018. Aspects of the biology and reproductive strategy of two Asian larval parasitoids evaluated for classical biological control of Drosophila suzukii. Biol Control. 121:58–65. doi:10.1016/j.biocontrol.2018.02.010.

Wittmeyer KT, Oppenheim SJ, Hopper KR. 2021. Assemblies of the genomes of parasitic wasps using meta-assembly and scaffolding with genetic linkage. G3 (Bethesda). 12(1):jkab386. doi:10.1093/g3journal/jkab386.

Xiao CL, Chen Y, Xie SQ, Chen KN, Wang Y, Han Y, Luo F, Xie Z. 2017. MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. Nat Methods. 14(11): 1072–1074. doi:10.1038/nmeth.4432.

Xiong WW, He LM, Lai JS, Dooner HK, Du CG. 2014. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. Proc Natl Acad Sci USA. 111(28):10263–10268. doi:10.1073/pnas.1410068111.

Xu Z, Wang H. 2007. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 35: W265–W268. doi:10.1093/nar/gkm286.

Yao YX, Duan JJ, Hopper KR, Mottern JL, Gates MW. 2016. A new species of *Oobius* Trjapitzin (Hymenoptera: Encyrtidae) from the Russian far east that parasitizes eggs of emerald ash borer (Coleoptera: Buprestidae). Ann Entomol Soc Am. 109(4):629–638. doi:10.1093/aesa/saw022.

Zhang QC, Zhang X, Wang Y, Chen JN, Zhou SC, Pang L, Wang Z, Wang Y, Chen X, Huang J. 2021. The complete mitochondrial genome of *Leptopilina syphax* (Hymenoptera: Figitidae). Mitochondrial DNA B Resour. 6(1):17–18. doi:10.1080/23802359.2020.1845106

Editor: R. Anholt