JHR 95: 31–44 (2023)
doi: 10.3897/jhr.95.97654
https://jhr.pensoft.net

RESEARCH ARTICLE

JOURNAL OF Hymenoptera RESEARCH
A peer-reviewed open-access journal
The International Society of Hymenopterists

The genome of the egg parasitoid *Trissolcus basalis* (Wollaston) (Hymenoptera, Scelionidae), a model organism and biocontrol agent of stink bugs

Zachary Lahey[1,2*], Huayan Chen[3,4*], Mark Dowton[5], Andrew D. Austin[6], Norman F. Johnson[1,3]

1 *Department of Evolution, Ecology, and Organismal Biology, Museum of Biological Diversity, The Ohio State University, Columbus, Ohio 43212, USA* **2** *United States Department of Agriculture, Agricultural Research Service, U.S. Vegetable Laboratory, Charleston, South Carolina 29414, USA* **3** *Department of Entomology, The Ohio State University, Columbus, Ohio 43212, USA* **4** *Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China* **5** *Centre for Medical and Molecular Bioscience, School of Biological Sciences, University of Wollongong, Wollongong, New South Wales 2522, Australia* **6** *Environment Institute, School of Biological Sciences, The University of Adelaide, Adelaide, South Australia 5005, Australia*

Corresponding authors: Zachary Lahey (zachary.lahey@usda.gov); Norman F. Johnson (johnson.2@osu.edu)

Academic editor: Elijah Talamas | Received 18 November 2022 | Accepted 12 January 2023 | Published 17 February 2023

https://zoobank.org/D4BCA9D4-A91B-4965-A7A0-8034808639C4

Citation: Lahey Z, Chen H, Dowton M, Austin AD, Johnson NF (2023) The genome of the egg parasitoid *Trissolcus basalis* (Wollaston) (Hymenoptera, Scelionidae), a model organism and biocontrol agent of stink bugs. Journal of Hymenoptera Research 95: 31–44. https://doi.org/10.3897/jhr.95.97654

Abstract

Trissolcus basalis (Wollaston) is a minute parasitic wasp that develops in the eggs of stink bugs. Over the past 30 years, *Tr. basalis* has become a model organism for studying host finding, patch defense behavior, and chemical ecology. As an entry point to better understand the molecular basis of these factors, in addition to filling a critical gap in the genomic resources available for parasitic Hymenoptera, we sequenced and assembled the genome of *Tr. basalis* using short (454, Illumina) and long read (Oxford Nanopore) sequencing technologies. The three sequencing methods produced 32 million reads (4.10 Gb; 27.9×), which were assembled into 7,586 scaffolds. The 147 Mb (N50: 42.8 kb) assembly contains complete sequences for 93.1% of the insect BUSCO dataset, and an extensive annotation protocol resulted in 14,158 protein-coding gene models, 12,197 (86%) of which had a blast hit in GenBank. Repetitive elements comprised 13.8% of the genome, and a phylogenomic analysis recovered *Tr. basalis* as sister to Chalcidoidea, a result in line with other studies. We identified 174 rapidly evolving gene families in *Tr. basalis*,

* These authors contributed equally to this work.

including olfactory receptors and pheromone/general odorant binding proteins. These genetic elements are an obligatory portion of the parasitoid-host relationship, and the draft genome of *Tr. basalis* has and will continue to be useful in elucidating these relationships at finer resolution.

Keywords
assembly, biological control, insect genomics, nanopore, Telenominae

Introduction

Trissolcus basalis (Wollaston) (Hymenoptera: Scelionidae) is a minute, solitary parasitoid of stink bug eggs (Hemiptera: Pentatomoidea), principally the cosmopolitan pest *Nezara viridula* (L.) (Pentatomidae). This parasitoid is found primarily in tropical and subtropical regions, where it has been used effectively in the biological control of its host (Davis 1964; Clarke 1990; Corrêa-Ferreira and Moscardi 1996). Given the economic importance of its host, considerable effort has been expended to elucidate how female *Tr. basalis* locate *N. viridula* eggs in the narrow window of time during which they are susceptible to attack (Bin et al. 1993; Colazza et al. 1999, 2004; Salerno et al. 2006; Laumann et al. 2009). Host location and acceptance by female wasps are known to be mediated by chemical cues, some of which have been isolated and identified (Mattiacci 1993; Colazza et al. 2004, 2007). This effort to sequence the genome of *Tr. basalis* was undertaken as a step in characterizing its repertoire of chemoreceptor proteins (Chen et al. 2021a) and to better understand the mechanisms of host finding in platygastroid wasps and their evolutionary consequences.

Methods

Whole-genome sequencing

454 Life sciences

Sequencing followed the protocol of Mao et al. (2012). Briefly, DNA was extracted from 25 adult male *Tr. basalis* from a colony maintained at the Università di Perugia (Perugia, Italy). Sequencing was conducted at the University of Pennsylvania Perelman School of Medicine on a Roche/454 GS FLX sequencer using Titanium chemistry, which generated 5,080,113 reads (1,535,920,544 bp).

Illumina

To correct homopolymer errors in the 454 reads, an Illumina sequencing library was prepared from five female *Tr. basalis* in the same culture. The DNA extract was prepared for Illumina sequencing using a Nextera DNA Sample Preparation Kit (Epicen-

tre Biotechnologies, Madison, Wisconsin, USA). Sequencing was conducted on an Illumina Genome Analyzer IIx (Illumina, San Diego, California, USA) at the Nucleic Acid Shared Resource (College of Medicine, The Ohio State University, Columbus, Ohio, USA). In total, 29,780,645 51-bp reads (1,518,812,895 bp) were generated.

Oxford nanopore

High molecular weight DNA was extracted from approximately 100 unsexed *Tr. basalis* using a Gentra Puregene Tissue Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol. DNA quality was estimated using an Agilent Bioanalyzer. The DNA library was prepared using a Ligation Sequencing Kit 1D. Sequencing was performed on a R9.5 flow cell using an Oxford Nanopore MinION (Oxford Nanopore, Oxford, United Kingdom). The 48-hour MinION sequencing run generated 341,751 reads (1,047,061,835 bp). All steps, excluding DNA extraction, were conducted at The Molecular and Cellular Imaging Center (MCIC; The Ohio State University, Wooster, Ohio, USA).

Processing of sequencing reads

Pyrosequencing reads are particularly susceptible to the accumulation of homopolymer errors (Huse et al. 2007). These were corrected using HECTOR (version 1.0.0; Wirawan et al. 2014). Adapter sequences were removed from nanopore reads with Porechop (version 0.2.3; https://github.com/rrwick/Porechop) and reads with internal adapter sequences were split into two reads.

Genome assembly

The *Tr. basalis* genome was assembled following a hybrid approach that utilized short (454, Illumina) and long read (Oxford Nanopore) sequencing technologies. 454 and nanopore reads were assembled with SPAdes (version 3.11.1; Bankevich et al. 2012), with the initial assembly (assembled in 2010 by NFJ) treated as 'trusted contigs' and the '-careful' flag turned on to minimize misassemblies. The assembly was polished with single-end 51 bp Illumina reads for 4 iterations using Pilon (version 1.22; Walker et al. 2014). The polished assembly was then scaffolded with RNA-seq reads from pooled tissues of male and female *Tr. basalis* using rascaf (version 20161129; Song et al. 2016; Chen et al. 2021a) to produce the final assembly.

Assembly statistics and quality

Genome statistics were calculated with QUAST (version 4.5; Gurevich et al. 2013). Genome assembly completeness was assessed with BUSCO (version 4.0.6; Simão et al. 2015) using the Metazoa, Arthropoda, Insecta, Endopterygota, and Hymenoptera datasets.

Genome annotation

The *Tr. basalis* genome was annotated following the protocol of Daren Card (Department of Organismic & Evolutionary Biology, Harvard University), with modifications (https://gist.github.com/zjlahey/3c400c3039eef674e335d3d850ad595f).

Repetitive elements

Repetitive elements were identified and annotated with RepeatModeler (version open-2.0.1; Flynn et al. 2020) and RepeatMasker (version 4.1.0; Smit et al. 2014). First, a custom repeat library was generated for *Tr. basalis* using RepeatModeler. This repeat library was then combined with a curated arthropod repeat library from RepBase (Bao et al. 2015), which was used to mask complex repetitive elements in the *Tr. basalis* genome using RepeatMasker.

Protein-coding genes

Protein-coding genes were annotated in an iterative fashion with MAKER (version 3.01.03; Campbell et al. 2014). MAKER utilizes external evidence in the form of protein and transcript sequences from other organisms to train *ab-initio* gene prediction software to annotate genes within a genome. In the first iteration, external evidence was supplied to MAKER as (1) TransDecoder-derived coding sequences (CDS) from each *Tr. basalis* transcriptome assembly; (2) CDS from *Telenomus remus* Nixon (Huayan Chen, unpublished data); (3) TransDecoder-derived protein sequences from each *Tr. basalis* transcriptome assembly; (4) all 170 arthropod proteomes in OrthoDBv10.1 (http://www.orthodb.org/); and (5) the UniProtKB/Swiss-Prot protein database (Bateman et al. 2020). Subsequent rounds utilized SNAP (version 2006-07-28; Korf 2004) and Augustus (version 3.3.3; Stanke and Waack 2003) to improve the gene models from the first iteration and identify new genes in the assembly. MAKER was run for three iterations, until the number of gene models and average length of each gene declined. Conserved domains within proteins of the final gene set were identified using InterProScan (version 5.46-81.0; Blum et al. 2020), and conserved functions were determined by performing a BLASTp (version 2.6.0) of the gene set against all metazoan proteins in the Swiss-Prot Uniprot database (Bateman et al. 2020). Finally, COGNATE (version 1.0; Wilbrandt et al. 2017) was employed to generate summary statistics of the annotated protein set (no. of exons/introns, avg. gene length, etc.).

Non-coding RNAs

We followed the protocol on Rfam (https://docs.rfam.org/en/latest/genome-annotation.html) to identify and annotate non-coding RNAs with Infernal (version 1.1.3; Nawrocki and Eddy 2013) and Rfam (version 13.0; Kalvari et al. 2018). Nuclear transfer RNAs were annotated with tRNAscan-SE (version 2.0.6; Lowe and Eddy 1997).

Gene family analysis

Taxon sampling and protein datasets

To estimate gene gains, losses, and rapidly evolving gene families within *Tr. basalis*, we conducted a gene family analysis using the *Tr. basalis* proteome and the protein sequences of six additional hymenopterans. Taxa were chosen based on the availability of hymenopteran proteomes and included three members of Proctotrupomorpha [*Belonocnema kinseyi* Weld (Cynipidae), *Nasonia vitripennis* (Walker) (Pteromalidae), and *Trichogramma pretiosum* Riley (Trichogrammatidae)]; one member of Ichneumonoidea [*Microplitis demolitor* Wilkinson (Braconidae)]; one member of Orussoidea [*Orussus abietinus* (Scopoli) (Orussidae)]; and the turnip sawfly, *Athalia rosae* (L.) (Tenthredinidae). Protein sequences of *A. rosae*, *O. abietinus*, *M. demolitor*, *N. vitripennis*, and *T. pretiosum* were downloaded from OrthoDB v10 (Kriventseva et al. 2019). The *B. kinseyi* proteome (then under the name *B. treatae* (Mayr) (Zhang et al. 2021)) was downloaded from NCBI. Redundant isoforms of multicopy genes in the *B. kinseyi* proteome were removed prior to analysis. Proteomes downloaded from OrthoDB did not require this step.

Gene family identification and clustering

Orthogroup inference was conducted with OrthoFinder (version 2.5.2; Emms and Kelly 2019) at default parameters (DIAMOND, MAFFT, FastTree). Due to computational limitations associated with using IQ-TREE at the tree inference step of OrthoFinder, we performed a separate phylogenetic analysis on the same 4,510 orthologues (SpeciesTreeAlignment.fa) identified during the initial run using IQ-TREE (version 2.1.2; Minh et al. 2020). The final step of OrthoFinder was then rerun with the species tree produced by IQ-TREE as input (orthofinder.py -ft RESULTS_DIR -s IQ-TREE_SPECIES_TREE). We then converted the species tree to a time-calibrated ultrametric tree using the OrthoFinder accessory script make_ultrametric.py, with the root node calibrated at 265 mya based on the estimated divergence time between *Athalia* Leach and *Orussus* Latreille in the TimeTree database (Kumar et al. 2017).

Gene family evolution

Rates of gene gain and loss (λ) were estimated with CAFE (version 4.2.1; Han et al. 2013) using the orthogroup count data and ultrametric time-tree produced by OrthoFinder as input. Prior to running CAFE, we modified the orthogroup count data file by removing gene family clusters where only a single species was present (Prost et al. 2019). This step reduced the number of gene families from 11,205 to 10,190. Finally, we accounted for possible deviation in the number of observed vs true gene family counts by estimating an error model (ε) to optimize the value of λ.

Data availability

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JAAMPD000000000. Raw DNA sequencing reads (454, Illumina, Nanopore) are available at the Sequence Read Archive by searching for BioProject Accession PRJNA49235.

Results and discussion

Genome sequencing and assembly statistics

We assembled the genome of *Tr. basalis de novo* using sequence data from second- and third-generation sequencing technologies. The combined read output from all sequencing platforms totaled 4.10 Gb (27.9× coverage). These reads were assembled into 7,568 scaffolds, totaling 147 Mb in length (34.7% GC content). The scaffold N50 was 42.8 kb, and the longest scaffold measured 349,262 kb. Given low read coverage, we were unable to estimate genome size *in silico*. However, the *Tr. basalis* draft assembly size falls within the range of genome size estimates of other platygastroids, which are typically between 200 and 400 Mb (data not shown), in addition to the average genome size range of other hymenopterans. We assessed genome assembly completeness with BUSCO, using the Insecta odbv10 database (N = 1,367) in genome mode and with the 'long' flag enabled to perform a more thorough search. We recovered 93.1% complete, 1.8% duplicated, 1.4% fragmented, and 3.7% missing Insecta BUSCOs in the *Tr. basalis* genome. These values compare favorably with other parasitoid Hymenoptera with more contiguous genome assemblies (Fig. 1).

Genome annotation

Repetitive elements

RepeatMasker annotated 13.8% of the *Tr. basalis* genome as composed of repeats, approximately half of which were unclassified repeats (7.1%). The most abundant classified repetitive elements were various LINE and LTR retroelements (3.0%); DNA transposons (1.3%); and simple repeats (1.5%). The repeat landscape of *Tr. basalis* shows a relatively uniform distribution of repeat classes, with a gradual decline in the proportion of LTR retroelements and an increase in the proportion of DNA transposons (Fig. 1). Subtle increases in the proportion of LINEs and rolling circle transposons are evident between Kimura distances 0.04 and 0.12. SINEs contribute little to the overall repeat content in *Tr. basalis*, and other Hymenoptera, in general (Petersen et al. 2019). A complete list of the repeats found in the *Tr. basalis* genome is in the Suppl. material 1.

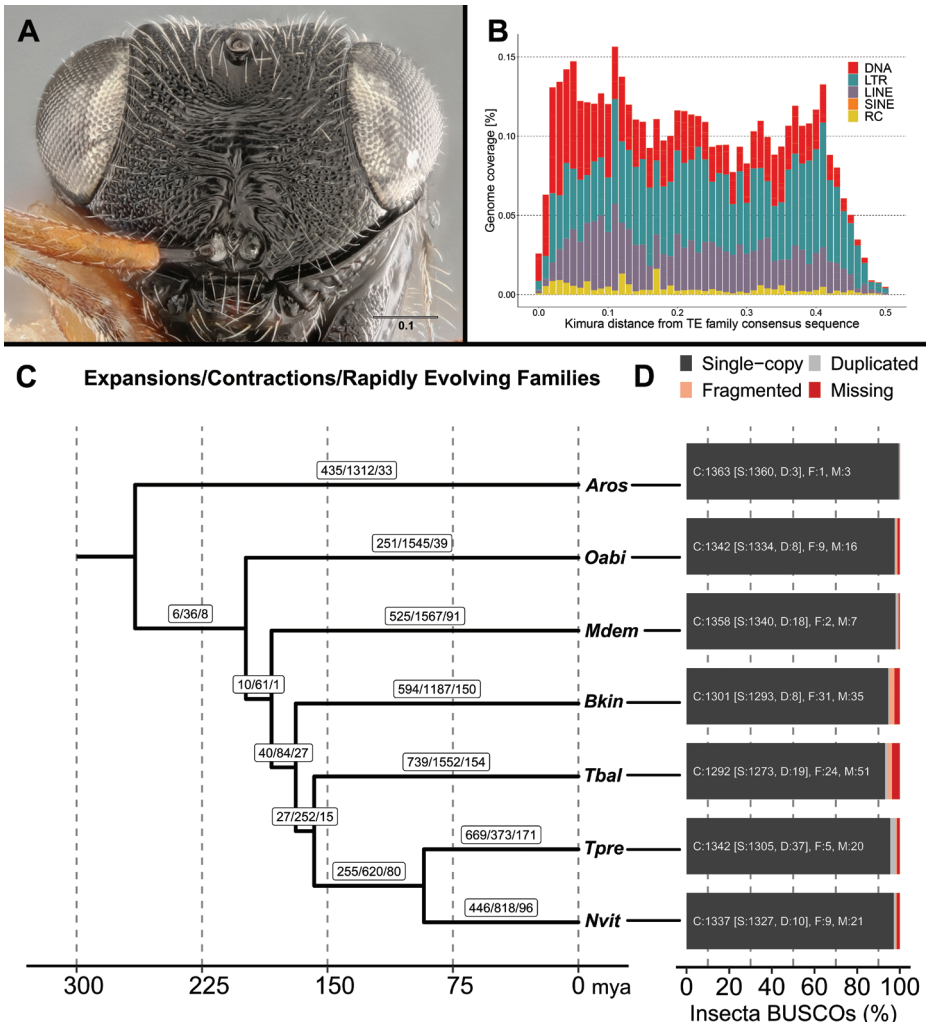

Figure 1. Morphological and genomic traits of *Trissolcus basalis* **A** head of female *Tr. basalis* reared from BMSB eggs in Tuscaloosa, Alabama, USA (FSCA 00090269 **B** repeat landscape plot of different TE classes within the *Tr. basalis* genome. Nucleotide sequence divergence in each TE copy was calculated as the Kimura distance between the annotated TE copies in the genome and the consensus sequence of each TE family **C** ultrametric timetree depicting the position of *Tr. basalis* relative to six other hymenopterans inferred from a phylogenetic analysis of 4,510 single-copy protein-coding genes identified by OrthoFinder. Numbers above branches (left to right, separated by forward slashes) indicate gene family expansions, gene family contractions, and the number of rapidly evolving gene families in each lineage. Each branch received 100% SH-aLRT and UFBoot2 support values **D** genome assembly completeness comparison based on the proportion of BUSCOs recovered in each genome using the Insecta odbv10 dataset (N = 1367). Abbreviations: BMSB, brown marmorated stink bug; C, complete; D, duplicated; DNA, DNA transposon; F, fragmented; LINE, long interspersed nuclear element; LTR, long terminal repeat; M, missing; mya, million years ago; RC, rolling circle transposon; S, single-copy; SINE, short interspersed nuclear element; TE, transposable element; Aros, *Athalia rosae*; Bkin, *Belonocnema kinseyi*; Mdem, *Microplitis demolitor*; Nvit, *Nasonia vitripennis*; Oabi, *Orussus abietinus*; Tbal, *Trissolcus basalis*; Tpre, *Trichogramma pretiosum*.

Protein-coding genes

The MAKER genome annotation pipeline resulted in 14,158 protein-coding gene models. Approximately 95% (13,507) of the 14,158 gene models have an annotation edit distance (AED) score of less than 0.5, and 70% (9,915) contain at least one recognizable InterPro domain. AED is a quality control metric that explains how well the gene annotations produced by MAKER match external evidence (i.e., proteomes from other species). The AED values and proportion of gene annotations with a recognizable InterPro domain for the *Tr. basalis* genome are indicative of a well-annotated assembly (Holt & Yandell, 2011). In addition, nearly half of the protein set (6,929 or 48.9%) was assigned at least one gene ontology (GO) term. To determine how well our annotated protein set compares with external protein databases, we queried our protein annotations against those of the metazoan portion of the Swiss-Prot/UniProt database and all Hymenoptera protein sequences deposited in GenBank (last accessed March 17, 2021). A total of 9,303 (65%) and 12,197 (86%) of the annotated proteins in *Tr. basalis* were supported by a best BLASTp hit in the Swiss-Prot/UniProt database and GenBank, respectively. A table of the most frequently recovered InterPro domains, GO terms, and Pfam entries associated with the *Tr. basalis* protein set is available in the Suppl. material 1.

Non-coding RNAs

Seventy-five different RNA families were annotated in the *Tr. basalis* genome. The top 5 most common families belong to the tRNA (RF00005), Histone3 (RF00032), 5S_rRNA (RF00001), SSU_rRNA_eukarya (RF01960), and LSU_rRNA_eukarya (RF02543) RNA sequence families. We also identified both conserved regions of the *Sphinx* long non-coding RNA gene, which plays a role in the regulation of male mating behavior in the fruit fly *Drosophila melanogaster* Meigen (Wang et al. 2002; Dai et al. 2008). Within Hymenoptera, *Sphinx* has been reported from 15 taxa including three species of parasitoid in the pteromalid genus *Nasonia* Ashmead (Werren et al. 2010). Its role in regulating mating behavior in Hymenoptera is not known. Additional statistics of the ribosomal DNA within the *Tr. basalis* genome are in the Suppl. material 1.

Gene family analyses

We compared the annotated proteome of *Tr. basalis* with those of six other hymenopterans with well-annotated genomes. Orthogroup clustering performed with OrthoFinder assigned 81,474 (93.4%) of the 87,222 protein sequences into 11,205 orthogroups. The number of orthogroups with all species present was 6,295 and 4,510 of these were identified as single-copy orthologues. Regarding *Tr. basalis*, 81.8% (11,582) of its genes were assigned to an orthogroup, and 76.7% (8,599) of orthogroups contained *Tr. basalis*. The number of orthogroups specific to *Tr. basalis* was 173, and the number of genes within these 173 species-specific orthogroups was 1,026 (7.2% of the 11,582

genes assigned to an orthogroup). The number of unassigned genes in *Tr. basalis* was much higher than the taxa with which it was compared. Potential explanations for this discrepancy are (1) the fragmentary nature of the *Tr. basalis* draft assembly leading to truncated protein models and (2) inaccurate gene annotations. Increasing genome contiguity using additional long-read sequencing technologies and chromosome confirmation capture would decrease the incidence of truncated protein models, and manual curation of the gene models would aid in the identification of false positives.

The orthogroup count data and ultrametric timetree produced by OrthoFinder were used to estimate the rate of gene family evolution with CAFE. We estimated the rate of gene family evolution (gains and losses) in this group of Hymenoptera at 0.0008, after accounting for possible genome assembly/annotation error. This result is in line with a recent multi-order gene family analysis that reported the rate of gene family gain and loss in 24 hymenopteran taxa at 0.0009 (Thomas et al. 2020), a gene turnover rate slower than Coleoptera (0.001), Diptera (0.001), and Lepidoptera (0.0014). In total, 638 gene families were identified as rapidly evolving among the 7 hymenopterans included in this study.

We identified 174 (99 expansions and 55 contractions) rapidly evolving gene families in *Tr. basalis*, with most (91) rapidly expanding families containing at least one member with an InterPro, PANTHER, or Pfam annotation (Suppl. material 1), and slightly fewer than half with at least one corresponding GO term (48). Notable examples of gene families undergoing rapid evolution in *Tr. basalis* are three groups of olfactory receptors (contracting in OG0000089; expanding in OG0000163 and OG0000567), one group of 7-transmembrane chemoreceptors (expanding, OG0000365), and one group of pheromone/general odorant binding proteins (expanding, OG0009810). The chemoreceptor repertoire of *Tr. basalis* was recently treated by Chen et al. (2021a) who employed sex- and tissue-specific transcriptome assemblies, in addition to the *Tr. basalis* genome, to annotate its gustatory, olfactory, and ionotropic receptor genes. One family of proteins not treated by Chen et al. (2021a), yet integral in the recognition and delivery of odorant molecules to their respective odorant receptors, are the odorant binding proteins (OBPs) (Pelosi and Maida 1995). OBPs are small, water-miscible polypeptides that solubilize and deliver volatile, hydrophobic compounds to the membrane of chemosensory receptor neurons for further processing (Pelosi et al. 2018). Therefore, OBPs are the first component in a multistep process that begins with semiochemical binding and culminates in a behavioral response. We are only beginning to investigate the OBP repertoire in *Tr. basalis*; however, given the quality of the *Tr. basalis* draft genome, we have identified, annotated, and characterized 18 putative OBPs, and determined those that exhibit antennal-biased expression patterns (King et al. 2021).

Author's note

While this manuscript was in preparation, Xu et al. (2021) published a highly contiguous, chromosome-scale genome assembly of *Te. remus* (Platygastroidea: Scelionidae), a telenomine egg parasitoid of the fall armyworm *Spodoptera frugiperda* (J. E. Smith)

(Lepidoptera: Noctuidae). *Telenomus* Haliday occupies an important phylogenetic position within the family Scelionidae as the sister taxon to *Trissolcus* Ashmead (Taekul et al. 2014; Chen et al. 2021b), and thus serves as an ideal candidate with which to compare the *Tr. basalis* genome assembly reported here. A preliminary investigation into some commonly reported genome metrics corroborates several features that may be characteristic of telenomine genomes: (1) small genome size (< 150 Mb); (2) low repetitive element content; (3) approximately 15,000 protein-coding genes; and (4) similar rates of gene family evolution (Xu et al. 2021). These differences were discernible between the two genomes despite the stark contrast in assembly contiguity (i.e., 11.9 Mb scaffold N50 for *Te. remus*; 42.8 kb scaffold N50 for *Tr. basalis*). This suggests that even highly fragmented genome assemblies can be of sufficient quality to infer genome-scale parameters accurately. We anticipate comparative genomic analyses between *Te. remus* and *Tr. basalis* will result in major discoveries related to genome evolution within Hymenoptera and the genomic factors implicated in host location, host acceptance, and the biological control potential of both species.

Acknowledgements

We thank Dr. Malte Petersen (Max Planck Institute of Immunobiology and Epigenetics) for sharing the R script used to generate the repeat landscape plot. Thanks to Dr. Jason Mottern (USDA-APHIS) and an anonymous reviewer for their careful and thoughtful review of the manuscript. This material is based upon work supported in part by the National Science Foundation under grant No. DEB-0614764 to N.F. Johnson and A.D. Austin and by funding from The Ohio State University, and the National Natural Science Foundation of China (31900346) to Huayan Chen.

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19: 455–477. https://doi.org/10.1089/cmb.2012.0021

Bao W, Kojima KK, Kohany O (2015) Repbase Update, a database of repetitive elements in eukaryotic genomes. Mobile DNA-UK 6: 1–11. https://doi.org/10.1186/s13100-015-0041-9

Bateman A, Martin MJ, Orchard S, Magrane M, Agivetova R, Ahmad S, Alpi E, Bowler-Barnett EH, Britto R, Bursteinas B, Bye-A-Jee H (2020) UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Research 49: D480–D489. https://doi.org/10.1093/nar/gkaa1100

Bin F, Vinson SB, Strand MR, Colazza S, Jones Jr WA (1993) Source of an egg kairomone for *Trissolcus basalis*, a parasitoid of *Nezara viridula*. Physiological Entomology 18: 7–15. https://doi.org/10.1111/j.1365-3032.1993.tb00443.x

Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L (2020) The InterPro protein families and

domains database: 20 years on. Nucleic Acids Research 49: D344–D354. https://doi.org/10.1093/nar/gkaa977

Campbell MS, Holt C, Moore B, Yandell M (2014) Genome annotation and curation using MAKER and MAKER-P. Current Protocols in Bioinformatics 48: 4–11. https://doi.org/10.1002/0471250953.bi0411s48

Chen H, Lahey Z, Talamas EJ, Johnson NF (2021a) Identification and expression of chemosensory receptor genes in the egg parasitoid *Trissolcus basalis*. Comparative Biochemistry and Physiology Part D 37: e100758. https://doi.org/10.1016/j.cbd.2020.100758

Chen H, Lahey Z, Talamas EJ, Valerio AA, Popovici OA, Musetti L, Klompen H, Polaszek A, Masner L, Austin AD, Johnson NF (2021b) An integrated phylogenetic reassessment of the parasitoid superfamily Platygastroidea (Hymenoptera: Proctotrupomorpha) results in a revised familial classification. Systematic Entomology 46: 1088–1113. https://doi.org/10.1111/syen.12511

Clarke AR (1990) The control of *Nezara viridula* L. with introduced egg parasitoids in Australia. A review of a 'landmark' example of classical biological control. Australian Journal of Agricultural Research 41: 1127–1146. https://doi.org/10.1071/AR9901127

Colazza S, Aquila G, De Pasquale C, Peri E, Millar JG (2007) The egg parasitoid *Trissolcus basalis* uses n-nonadecane, a cuticular hydrocarbon from its stink bug host *Nezara viridula*, to discriminate between female and male hosts. Journal of Chemical Ecology 33: 1405–1420. https://doi.org/10.1007/s10886-007-9300-7

Colazza S, McElfresh JS, Millar JG (2004) Identification of volatile synomones, induced by *Nezara viridula* feeding and oviposition on bean spp., that attract the egg parasitoid *Trissolcus basalis*. Journal of Chemical Ecology 305: 945–964. https://doi.org/10.1023/B:JOEC.0000028460.70584.d1

Colazza S, Salerno G, Wajnberg E (1999) Volatile and contact chemicals released by *Nezara viridula* (Heteroptera: Pentatomidae) have a kairomonal effect on the egg parasitoid *Trissolcus basalis* (Hymenoptera: Scelionidae). Biological Control 16: 310–317. https://doi.org/10.1006/bcon.1999.0763

Corrêa-Ferreira BS, Moscardi F (1996) Biological control of soybean stink bugs by inoculative releases of *Trissolcus basalis*. Entomologia Experimentalis et Applicata 79: 1–7. https://doi.org/10.1111/j.1570-7458.1996.tb00802.x

Dai H, Chen Y, Chen S, Mao Q, Kennedy D, Landback P, Eyre-Walker A, Du W, Long M (2008) The evolution of courtship behaviors through the origination of a new gene in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America 105: 7478–7483. https://doi.org/10.1073/pnas.0800693105

Davis CJ (1964) The introduction, propagation, liberation, and establishment of parasites to control *Nezara viridula* variety *smaragdula* (Fabricius) in Hawaii (Heteroptera: Pentatomidae). Proceedings of the Hawaiian Entomological Society 18: 369–375.

Emms DM, Kelly S (2019) OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology 20: e238. https://doi.org/10.1186/s13059-019-1832-y

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF (2020) RepeatModeler2 for automated genomic discovery of transposable element families. Proceedings of the National Academy of Sciences of the United States of America 117: 9451–9457. https://doi.org/10.1073/pnas.1921046117

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: quality assessment tool for genome assemblies. Bioinformatics 29: 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Han MV, Thomas GWC, Lugo-Martinez J, Hahn MW (2013) Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Molecular Biology and Evolution 30: 1987–1997. https://doi.org/10.1093/molbev/mst100

Holt C, Yandell M (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. BMC Bioinformatics 12: 1–4. https://doi.org/10.1186/1471-2105-12-491

Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biology 8: R143. https://doi.org/10.1186/gb-2007-8-7-r143

Kalvari I, Argasinska J, Quinones-Olvera N, Nawrocki EP, Rivas E, Eddy SR, Bateman A, Finn RD, Petrov AI (2018) Rfam 13.0: shifting to a genome-centric resource fornon-coding RNA families. Nucleic Acids Research 46: D335–D342. https://doi.org/10.1093/nar/gkx1038

King K, Meuti ME, Johnson NF (2021) Identification and expression of odorant binding proteins in the egg-parasitoid *Trissolcus basalis* (Wollaston) (Hymenoptera, Scelionidae, Telenominae). Journal of Hymenoptera Research 87: 251–266. https://doi.org/10.3897/jhr.87.68954

Korf I (2004) Gene finding in novel genomes. BMC Bioinformatics 5: 1–59. https://doi.org/10.1186/1471-2105-5-59

Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM (2019) OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic Acids Research 47: D807–D811. https://doi.org/10.1093/nar/gky1053

Kumar S, Stecher G, Suleski M, Hedges SB (2017) TimeTree: a resource for timelines, timetrees, and divergence times. Molecular Biology and Evolution 34: 1812–1819. https://doi.org/10.1093/molbev/msx116

Laumann RA, Aquino MF, Moraes MC, Pareja M, Borges M (2009) Response of the egg parasitoids *Trissolcus basalis* and *Telenomus podisi* to compounds from defensive secretions of stink bugs. Journal of Chemical Ecology 35: 8–19. https://doi.org/10.1007/s10886-008-9578-0

Lowe TM, Eddy SR (1997) tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research 25: 955–964. https://doi.org/10.1093/nar/25.5.955

Mao M, Valerio A, Austin AD, Dowton M, Johnson NF (2012) The first mitochondrial genome for the wasp superfamily Platygastroidea: the egg parasitoid *Trissolcus basalis*. Genome 55: 194–204. https://doi.org/10.1139/g2012-005

Mattiacci L, Vinson SB, Williams HJ, Aldrich JR, Bin F (1993) A long-range attractant kairomone for egg parasitoid *Trissolcus basalis*, isolated from defensive secretion of its host, *Nezara viridula*. Journal of Chemical Ecology 19: 1167–1181. https://doi.org/10.1007/BF00987378

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R (2020) IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37: 1530–1534. https://doi.org/10.1093/molbev/msaa015

Nawrocki EP, Eddy SR (2013) Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29: 2933–2935. https://doi.org/10.1093/bioinformatics/btt509

Pelosi P, Maida R (1995) Odorant-binding proteins in insects. Comparative Biochemistry and Physiology Part B 111: 503–514. https://doi.org/10.1016/0305-0491(95)00019-5

Pelosi P, Iovinella I, Zhu J, Wang G, Dani FR (2018) Beyond chemoreception: diverse tasks of soluble olfactory proteins in insects. Biological Reviews 93: 184–200. https://doi.org/10.1111/brv.12339

Peters RS, Krogmann L, Mayer C, Donath A, Gunkel S, Meusemann K, Kozlov A, Podsiadlowski L, Petersen M, Lanfear R, Diez PA, Heraty J, Kjer KM, Klopfstein S, Meier R, Polidori C, Schmitt T, Liu S, Zhou X, Wappler T, Rust J, Misof B, Niehuis O (2017) Evolutionary history of the Hymenoptera. Current Biology 27: 1013–1018. https://doi.org/10.1016/j.cub.2017.01.027

Prost S, Armstrong EE, Nylander J, Thomas GW, Suh A, Petersen B, Dalen L, Benz BW, Blom MP, Palkopoulou E, Ericson PG (2019) Comparative analyses identify genomic features potentially involved in the evolution of birds-of-paradise. GigaScience 8: giz003. https://doi.org/10.1093/gigascience/giz003

Salerno G, Conti E, Peri E, Colazza S, Bin F (2006) Kairomone involvement in the host specificity of the egg parasitoid *Trissolcus basalis* (Hymenoptera: Scelionidae). European Journal of Entomology 103: 311–318. https://doi.org/10.14411/eje.2006.040

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31: 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Smit AFA, Hubley R, Green P (2014) RepeatMasker. http://www.repeatmasker.org

Song L, Shankar DS, Florea L (2016) rascaf: Improving genome assembly with RNA sequencing data. Plant Genome-US 9: 1–12. https://doi.org/10.3835/plantgenome2016.03.0027

Stanke M, Waack S (2003) Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19: 215–225. https://doi.org/10.1093/bioinformatics/btg1080

Taekul C, Valerio AA, Austin AD, Klompen H, Johnson NF (2014) Molecular phylogeny of telenomine egg parasitoids (Hymenoptera: Platygastridae sl.: Telenominae): evolution of host shifts and implications for classification. Systematic Entomology 39: 24–35. https://doi.org/10.1111/syen.12032

Thomas GWC, Dohmen E, Hughes DST, Murali SC, Poelchau M, Glastad K, Anstead CA, Ayoub NA, Batterham P, Bellair M, Binford GJ, Chao H, Chen YH, Childers C, Dinh H, Doddapaneni HV, Duan JJ, Dugan S, Esposito LA, Friedrich M, Garb J, Gasser RB, Goodisman MAD, Gundersen-Rindal DE, Han Y, Handler AM, Hatakeyama M, Hering L, Hunter WB, Ioannidis P, Jayaseelan JC, Kalra D, Khila A, Korhonen PK, Lee CE, Lee SL, Li Y, Lindsey ARI, Mayer G, McGregor AP, McKenna DD, Misof B, Munidasa M, Munoz-Torres M, Muzny DM, Niehuis O, Osuji-Lacy N, Palli SR, Panfilio KA, Pechmann M, Perry T, Peters RS, Poynton HC, Prpic N-M, Qu J, Rotenberg D, Schal C, Schoville SD, Scully ED, Skinner E, Sloan DB, Stouthamer R, Strand MR, Szucsich NU, Wijeratne A, Young ND, Zattara EE, Benoit JB, Zdobnov EM, Pfrender ME, Hackett KJ, Werren JH, Worley KC, Gibbs RA, Chipman AD, Waterhouse RM, Bornberg-Bauer E,

Hahn MW, Richards S (2020) Gene content evolution in the arthropods. Genome Biology 21: 1–15. https://doi.org/10.1186/s13059-019-1925-7

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9: e112963. https://doi.org/10.1371/journal.pone.0112963

Wang W, Brunet FG, Nevo E, Long M (2002) Origin of *sphinx*, a young chimeric RNA gene in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America 99: 4448–4453. https://doi.org/10.1073/pnas.072066399

Werren JH, Richards S, Desjardins CA, Niehuis O, Gadau J, Colbourne JK, Nasonia Genome Working Group, Beukeboom LW, Desplan C, Elsik CG, Grimmelikhuijzen CJ (2010) Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. Science 327: 343–348. https://doi.org/10.1126/science.1178028

Wilbrandt J, Misof B, Niehuis O (2017) COGNATE: comparative gene annotation characterizer. BMC Genomics 18: e535. https://doi.org/10.1186/s12864-017-3870-8

Wirawan A, Harris RS, Liu Y, Schmidt B, Schröder J (2014) HECTOR: a parallel multistage homopolymer spectrum based error corrector for 454 sequencing data. BMC Bioinformatics 15: e131. https://doi.org/10.1186/1471-2105-15-131

Xu H, Ye X, Yang Y, Yang Y, Sun YH, Mei Y, Xiong S, He K, Xu L, Fang Q, Li F (2021) Comparative genomics sheds light on the convergent evolution of miniaturized wasps. Molecular Biology and Evolution 38: 5539–5554. https://doi.org/10.1093/molbev/msab273

Zhang YM, Egan SP, Driscoe AL, Ott JR (2021) One hundred and sixty years of taxonomic confusion resolved: *Belonocnema* (Hymenoptera: Cynipidae: Cynipini) gall wasps associated with live oaks in the USA. Zoological Journal of the Linnean Society 193: 1234–1255. https://doi.org/10.1093/zoolinnean/zlab001

Supplementary material 1

Genome of the egg parasitoid *Trissolcus basalis* (Wollaston) (Hymenoptera, Scelionidae), a model organism and biocontrol agent of stink bugs

Author: Zachary Lahey

Data type: genomic (excel document)

Explanation note: Bioinformatic data associated with the annotated *Trissolcus basalis* genome assembly.

Link: https://doi.org/10.3897/jhr.95.97654.suppl11