GBE

The Genome Sequence of the Hemlock Woolly Adelgid, *Adelges tsugae* Annand 1924

Bryan M.T. Brunet [1,*], Dustin T. Dial [2], Gaelen R. Burke [2], Carol D. von Dohlen [3], Julia Douglas Freitas [1], Haley Sanderson [1], Afiya R. Chida [4], Samantha J. Jones [4], Fergal J. Martin [5], Leanne Haggerty [5], Stephen W. Scherer [6,7], Ioannis Ragoussis [8], Steven J.M. Jones [4,9], Robert G. Foottit [1], Nathan P. Havill [10]

[1]Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON, Canada K1A 0C6
[2]Department of Entomology, University of Georgia, Athens, GA 30602, USA
[3]Department of Biology, Utah State University, Logan, UT 84322-0300, USA
[4]Canada's Michael Smith Genome Sciences Centre, Vancouver, BC, Canada V5Z 4S6
[5]European Molecular Biology Laboratory, European Bioinformatics Institute, Cambridge CB10 1SD, UK
[6]The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada M5G 0A4
[7]Department of Molecular Genetics, McLaughlin Centre, University of Toronto, Toronto, ON, Canada M5S 1A8
[8]McGill Genome Centre, Victor Phillip Dahdaleh Institute of Genomic Medicine, McGill University, Montreal, QC, Canada H3A 0G1
[9]Department of Medical Genetics, University of British Columbia, Vancouver, BC, Canada V6T 1Z4
[10]USDA Forest Service, Northern Research Station, Hamden, CT 06514, USA

*Corresponding author: Tel: 1-613-759-6861, E-mail: bryan.brunet@agr.gc.ca.

Accepted: August 19, 2025

Abstract

The genome of the hemlock woolly adelgid (*Adelges tsugae*) is presented in 10 chromosomal pseudomolecules along with 132 unplaced scaffolds for a combined length of 216.56 Mb. The genome is highly contiguous with an N50 of 20.54 Mb and its longest scaffold 37.41 Mb in length. The assembly recovered 97.7% of benchmarked universal single-copy Hemiptera orthologs (BUSCO) with an overall completion score of 98.6%. On the basis of chromosomal synteny with other aphid genomes, the hemlock woolly adelgid's genome is composed of eight autosomes and two sex (X) chromosomes. Annotation of the assembly identified 11,800 coding genes, 1,930 noncoding genes, and 20,403 mRNA transcripts. The mitogenome is also presented in a single, annotated, circular contig 25,980 bases in length.

Key words: *Adelges tsugae*, Adelgidae, genome assembly, chromosome-scale, mitogenome.

Significance

A chromosome-scale genome assembly is currently lacking for the hemlock woolly adelgid, *Adelges tsugae*, despite the fact that it is one of eastern North America's most destructive and invasive forest pests. Here, we use a combination of sequencing technologies to assemble the first chromosome-length genome for this species and the Adelgidae. We also identify its sex determination system through comparisons of chromosomal synteny with other aphids and assemble its mitogenome, the largest among adelgids sequenced thus far. These important genomic resources will help to facilitate investigation of the genetic basis of life history traits that allows this species to exploit new areas.

Introduction

Few adventive aphids (Hemiptera: Sternorrhyncha: Aphidomorpha) are as economically important to North American forests as the hemlock woolly adelgid (HWA), *Adelges tsugae* Annand, 1924. This species has devastating and disruptive effects on native eastern hemlock species and their biotic and abiotic associates, including the alteration of forest vegetation composition, declines in associated fauna, and ecosystem-level changes (Ellison et al. 2018). HWA was introduced to North America near Richmond, Virginia, before 1951 from its native range in southern Japan (Havill et al. 2016) and has since spread at a rate of 2.5 to 37 km per year to cover most of the native hemlock forest in the eastern United States (McClure 2001; Evans and Gregoire 2007; Morin et al. 2009; McAvoy et al. 2025). It is now also threatening hemlocks in eastern Canada, with recent detections in southern Ontario (Fidgen et al. 2021; Canadian Food Inspection Agency 2024) and an established population in Nova Scotia (Emilson et al. 2018; Limbu et al. 2018). Moreover, species distribution models forecast that HWA will infest all but a narrow band of hemlock forest north of the 45th parallel by 2050 (Fitzpatrick et al. 2012; Ellison et al. 2018).

Unlike other significant insect forest pests in North America, robust genomic resources remain to be developed for HWA. In fact, among the Aphidomorpha, only the Aphididae and Phylloxeridae are represented among species with available chromosome-length genome assemblies, and at present the only adelgid with an assembly, the Cooley spruce gall adelgid, *Adelges cooleyi* (Gillette, 1907) (Dial et al. 2023), is in 840 scaffolds (270.2 Mb, N50 = 14.87 Mb). High-quality genome assemblies provide a foundation for understanding the genetic architecture of complex traits and are an important resource in the molecular arsenal used in the development of management and biocontrol strategies (Li et al. 2019; Leung et al. 2020). Here, we present a chromosome-scale genome assembly for HWA, the first of this magnitude for the Adelgidae, to bolster resources available for future work on the comparative genomics of the Aphidomorpha and the genetic basis of life history traits that allow HWA to exploit new areas in its invasive range.

Results and Discussion

The genome of the HWA was assembled from multiple samples collected in Hamden Connecticut, USA (Table S1, Supplementary Material online), together representing the single multilocus lineage that predominates the northeast United States and Canada (Havill et al. 2016). Using a combination of newly generated PacBio continuous long-reads (101× coverage), 10× Chromium linked-reads (249×), and Hi-C data (24×), as well as existing Illumina short-reads (86×) from two previously published endosymbiont genomes (Weglarz et al. 2018), a preliminary assembly consisting of 145 scaffolds was generated. This was manually reordered to accommodate two sex chromosomes and merge X1 with two smaller unplaced scaffolds through translocation corrections (Fig. 1a). A 5 Kb gap associated with one of these corrections was also excised and subsequently removed from the assembly, along with one other scaffold that was <200 bp in length, leaving the final scaffold count at 142.

The final curated assembly has a total length of 216.56 Mb with a scaffold N50 of 20.54 Mb (Fig. 1b; Table 1) and is about 75 Mb shorter than the flow cytometric estimate at 292.3 ± 1.1 Mb ($n = 2$). Over 99% of the genome is anchored to 10 chromosome-scale pseudomolecules between 16 and 37 Mb in length (Table S2, Supplementary Material online). This is consistent with the haploid chromosome number of other adelgids (Gavrilov-Zimin et al. 2015). All scaffolds had significant blastn and diamond blastx hits (e-value < 1e-25) to insect homologs, with matches to Adelgidae and Aphididae accounting for 58 scaffolds and a combined length of 216.03 Mb (Fig. S2, Supplementary Material online). BUSCO analysis of the final assembly yielded 98.6% complete BUSCO's (97.7% single copy and 0.9% duplicated) (Table 1; Table S3, Supplementary Material online). Mapping rates, *k*-mer completeness, and assembly consensus quality (QV) all indicated a highly complete and error-free assembly (Table 1). Furthermore, a spectra-cn plot showed low *k*-mer duplication and a large diploid peak at a multiplicity of 140 with a nearly undetectable haploid peak at 75 (Fig. S3, Supplementary Material online). This suggests strong homozygosity, as would be expected for the invasive eastern North American *A. tsugae* lineage which reproduces only by parthenogenesis and experienced a strong genetic bottleneck when introduced (Havill et al. 2016).

Approximately 52.73% (114.19 Mb) of the genome was masked as repetitive (244,333 repeats). However, most repeats were unclassifiable since only 48,617 repetitive elements could be annotated. Most of these (91.6%) were classified as transposons or long terminal repeats (Table 1). Furthermore, our attempt to build a taxon-specific repeat model revealed that over 84% of 924 identified repeat families could not be classified. Repeats containing the ancestral insect TTAGG telomere motif were observed in abundance across the genome, but only in excess of 40 repeats/10 Kb near the ends of three chromosomes (Fig. 1c). Nevertheless, loss of the ancestral telomere motif is well documented across the insecta, including the Hemiptera (Lukhtanov and Pazhenkova 2023), and therefore its absence here should not discount the chromosomal nature of the HWA genome. The genome is composed of 11,800 coding genes (120.29 Mb)

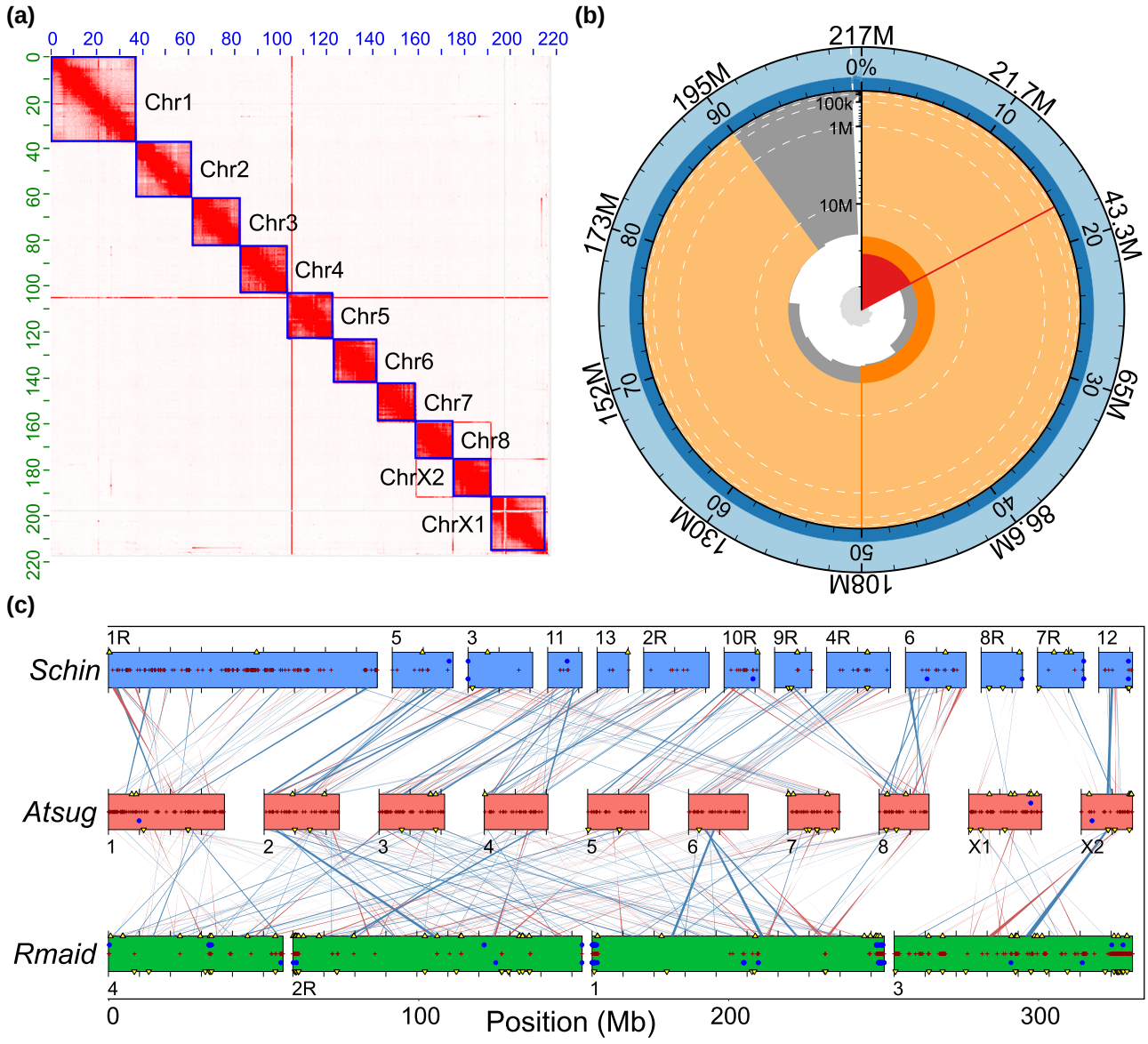

Fig. 1. Genome assembly of *Adelges tsugae*, ihAdeTsug: metrics and chromosome synteny. a) Hi-C contact map of *A. tsugae* assembly visualized using JuiceBox Assembly Tools. Autosomes are shown in order of size from top-left to bottom-right. Sex chromosomes X1 and X2 are shown in reverse order due to unresolved linkages between X1 and unplaced scaffolds visible at the lower right. b) BlobToolKit snail plot showing *A. tsugae* assembly composition. The main plot is divided into 1,000 size-ordered bins around the circumference, with each bin representing 0.1% of the 216,556,762 bp assembly. The distribution of chromosome lengths is shown in dark gray, and the plot radius scaled to the longest chromosome present in the assembly (37,413,131 bp) is shown in red. Orange and pale-orange arcs show the N50 and N90 chromosome lengths (20,533,966 and 16,550,355 bp, respectively). The pale gray spiral shows the cumulative chromosome count on a log scale with white scale lines showing successive orders of magnitude. The blue and pale-blue area around the outside of the plot displays the distribution of GC (blue), AT (pale blue), and N (white) percentages, using the same bins as the inner plot. c) Three-way ChromSyn plot showing connections between syntenic regions greater than 50 Kb between the chromosomes of *Adelges tsugae* (Atsug), *Schlechtendalia chinensis* (Schin), and *Rhopalosiphum maidis* (Rmaid). Showing only 13 (of 208) scaffolds of the *S. chinensis* genome (280.43 Mb, N50 = 20.58 Mb) and 4 (of 220) scaffolds of the *R. maidis* genome (326.0 Mb, N50 = 93.3 Mb). Red lines are inverted regions. Plus symbols within chromosomes denote assembly gaps. Yellow triangles are predicted positions of BUSCO orthologs that were not found in each assembly assuming conserved synteny. Blue dots are locations of telomeric repeat regions (TTAGG) with greater than 40 repeats/10 Kb window. The location of triangles and dots above or below the chromosome center line indicates the strand. Reversed chromosomes are labeled with an "R" suffix. Chromosomes 7, 8, and 12 and chromosome 3 are the sex chromosomes of *S. chinensis* and *R. maidis*, respectively.

Table 1 Summary of genome assembly and annotation statistics for *A. tsugae* (ihAdeTsug)

Elements	Value
Genome assembly	
Assembly size (Mb)	216.56
Number of scaffolds/contigs	142/528
Longest scaffold/contig (Mb)	37.41/35.00
N50 scaffold/contig length (Mb)	20.54/0.900
GC (%)	30.93
Gaps (%)	1.849
% chromosomal	99.31
Short-/linked-read mapping rates (%)	93.3/78.2
k-mer completeness (%)	87.23
Assembly consensus quality (QV)	44.1
BUSCO completeness (%)[a]	C:98.6% [S:97.7%, D:0.9%], F:0.6%, M:0.8%, *n*:2510 (Hemiptera_odb10)
Repeat annotations	
Type I transposons/SINE	124 (8629 bp)
Type I transposons/LINE	11,927 (2,599,150 bp)
Type II transposons/DNA	28,460 (2,806,703 bp)
LTR	4,793 (529,289 bp)
Other	3,313 (547,264 bp)
Gene annotations	
Protein-coding/noncoding genes	11,800 (120,290,929 bp)/1,930 (7,508,078 bp)
mRNA transcripts	20,403 (211,610,416 bp)
5′ UTR/3′ UTR	21,514 (3,163,593 bp)/15,757 (7,426,794 bp)
Small/long noncoding genes	424 (36,336 bp)/1,667 (9,025,655 bp)
BUSCO completeness (%)	C:98.6% [S:97.6%, D:1.0%], F:0.7%, M:0.7%, *n*:2510, E:2.6% (Hemiptera_odb10)

[a]Metaeuk as gene predictor.

with an average length 10,195.15 bp, 1930 noncoding genes, and 20,403 mRNA transcripts (Table 1). BUSCO analysis of the proteome identified 98.6% Hemiptera orthologs as complete (97.6% single copy and 1.0% duplicated) (Table 1).

Synteny analysis revealed strong conservation of sex chromosome composition and structure between HWA, *Schlechtendalia chinensis* (Bell, 1851) (Aphididae: Eriosomatinae) and *Rhopalosiphum maidis* (Fitch, 1856) (Aphididae: Aphidinae) (Fig. 1c). This was previously demonstrated for the Aphidomorpha, but without comparisons that included a chromosome-scale adelgid assembly (Wei et al. 2022; Mathers et al. 202l; Li et al. 2023; Huang et al. 2025). Two scaffolds (X1 and X2) were syntenic with the sex chromosome(s) of the two other species (Wei et al. 2022), providing confirmation of a $2(X_1X_2)/X_1X_20$ (♀/♂) sex determination system for HWA. Strong concordance between the eight remaining autosomes and those of *S. chinensis* was also revealed, suggesting that elevated autosomal rearrangement within the Aphididae (Li et al. 2023) may be limited to aphid subfamilies that have undergone more recent diversification.

The mitochondrial genome was also assembled into a completely circular contig 25,980 bp in length (Fig. S4, Supplementary Material online). The arrangement of protein-coding genes (PCGs), transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs) agrees with the existing HWA mitogenome reference (MT263947) (Yeh et al. 2020). The mitogenome consists of 13 PCGs, 22 tRNAs, 2 rRNAs, and large control (4655 bp) and repeat (5755 bp) regions characterized by AT-rich tandem repeats. The control region contains an imperfect compound repeat of 232 bases (2.9 copies) plus 23 bases (2 copies) with a 3 bp overlap repeated four times between tRNA-Ile and rRNA-S. The repeat region is composed of an imperfect repeat of 207 bases with 24.9 copies between nucleotide positions 420 and 6175. The two HWA mitogenomes have approximately 94% pairwise sequence identity among all PCGs and both rRNAs, as expected based on prior characterizations of this species' global genetic diversity (Havill et al. 2016).

Materials and Methods

Sample Collection and Extraction

Whole adults, nymphs, and eggs of HWA were collected by hand on five occasions between 2013 and 2021. Samples were reared for up to 1 week at room temperature to accumulate specimens before transferring to 95% ethanol and storing at −80 °C until further use. Four samples were taken from previously identified HWA-infested eastern hemlock (*Tsuga canadensis*) at the USDA Forest Service Northern Research Station in Hamden, New Haven County, Connecticut, and one from another location

4 km away (Table S1, Supplementary Material online). A slide-mounted adult voucher specimen of sample ihAdeTsug4 was retained and deposited in the Yale Peabody Museum (YPM; Connecticut, USA) with accession number YPM#ENT595371. Isolation of DNA from each sample was performed on pooled specimens using commercially available kits, including the Roche Diagnostics High Pure PCR template kit (ihAdeTsug1; Weglarz et al. 2018), Qiagen Blood and Cell Culture DNA Mini kit (ihAdeTsug2), and MagAttract HMW DNA kit (ihAdeTsug3 and ihAdeTsug4).

Library Construction and Sequencing

A Pacific Biosciences long-insert library was prepared by combining high molecular DNA for samples ihAdeTsug3 and ihAdeTsug4 and selecting for fragments of 15 Kb in size without shearing at the University of Maryland School of Medicine's Institute for Genome Sciences, Genome Resource Centre (Maryland, USA). Continuous long-read sequencing was performed in a single SMRT Cell 8 M run on a PacBio Sequel II over a 15 h movie collection time. A 10× Genomics Chromium linked-read library (2 × 150 bp) was prepared using sample ihAdeTsug2 and subsequently sequenced on an Illumina HiSeqX at the McGill University and Génome Québec Innovation Centre (Québec, Canada). Sample ihAdeTsug5 was used to prepare a single Arima Hi-C sequencing library (2 × 150 bp), which was then sequenced in two runs on an Illumina MiSeq v2 at the McGill Genome Centre (Québec, Canada). The Illumina libraries (2 × 150 bp with 500 bp inserts) for sample ihAdeTsug1 noted above were sequenced on a HiSeq2500 at the Yale Centre for Genome Analysis (Connecticut, USA) (Weglarz et al. 2018).

The genome size was determined through flow cytometric estimation using genomic DNA isolated from two adult females following standard protocols for insects (Johnston et al. 2019). *Drosophila virilis* (1C = 328 Mb) was used as the standard.

Genome Assembly

Raw PacBio subreads were input to Falcon-Unzip (Chin et al. 2016) using the PacBio Assembly Tool Suite (pb-assembly) (Table S2, Supplementary Material online). Longstitch (Coombe et al. 2021) was used to correct misassembled contigs and then scaffold and gap-fill using preassembly reads (preads) from the Falcon output. ARCS (Yeo et al. 2018) was used to identify and correct misassembled scaffolds using 10× Chromium linked-reads. Linked-reads were first barcode-processed using Long Ranger (Marks et al. 2019) "basic."

Short-read polishing and gap-filling were completed using the ntEdit + Sealer Assembly Finishing Protocol (Li et al. 2022). Illumina short-reads and 10× linked-reads were trimmed and filtered of low-quality and contaminant sequences using RQCFilter2 in BBmap (Bushnell 2014) and then error-corrected using Lighter (Song et al. 2014) before being input into ntEdit + Sealer. Lighter was run with automated estimation of alpha using a *k*-mer length of 17 and an approximated genome size of 300 Mb. Prior to trimming and filtering, the 10× linked-reads were first barcode-processed using process_10xReads.py of proc10xG (https://github.com/ucdavis-bioinformatics/proc10xG). The Bloom filter size used in ntEdit + Sealer was estimated at 20 GB using ntCard (Mohamadi et al. 2017).

Redundant and junk scaffolds were then identified and removed from the assembly using Purge Haplotigs (Roach et al. 2018) following alignment of the raw PacBio subreads to the assembly. Arima Hi-C reads were mapped to the purged assembly and filtered on the basis of chimerism and mapping quality using the Arima mapping pipeline (https://github.com/ArimaGenomics/mapping_pipeline). Technical replicates were also merged before removal of duplicate reads. Hi-C scaffolding was performed with YaHS (Zhou et al. 2023). Juicer Tools (Durand et al. 2016) was used to generate a Hi-C contact matrix for visualization in Juicebox. Manual curation of the assembly was performed using Juicebox Assembly Tools (Dudchenko et al. 2018).

The curated assembly was checked for contamination using Blobtools2 from the Blobtoolkit (Challis et al. 2020). First, each scaffold was partitioned into 100 Kb sequences with 500 bp overlap using the shred.sh script from BBmap, and then each partition was queried against the NCBI nt database (released 05-July-2024) using blastn (Altschul et al. 1990). Similarly, Diamond blastx (Buchfink et al. 2021) searches were performed against the UniProt reference proteomes (release 2024_06) database (Bateman et al. 2023). Coverage of the raw PacBio subreads was then assessed by mapping them to the genome assembly using minimap2 (Li 2021). BUSCO (Manni et al. 2021) was used to assess the completeness of the assembly iteratively after each step using the Hemiptera_odb10 dataset. The assembly was also assessed by calculating mapping rates of the error-corrected short- and 10× linked-reads using minimap2, as well as *k*-mer completeness and QV with Merqury (Rhie et al. 2020) using the 10× linked-reads following preparation of a meryl *k*-mer database (k = 19).

Structural Annotation and Synteny Analysis

Annotation was completed by the European Bioinformatics Institute. Masking of repeats was performed using REpeatDetector (REd) (Girgis 2015). Repetitive DNA elements were annotated using RepeatMasker (Smit et al. 2015) and RepBase (release 20170127) (Bao et al. 2015). A de novo transposable element reference library was also generated using RepeatModeller2 (Flynn et al. 2020)

but was ultimately not used for repeat annotation. Gene annotation was performed using the Ensembl Gene Annotation system (Aken et al. 2016; https://beta.ensembl.org/help/articles/non-vertebrate-genome-annotation) and publicly available RNA-seq data on NCBI (PRJNA242203). ChromSyn (Edwards et al. 2022) was used to examine chromosome synteny between HWA, *S. chinensis* (GCA_019022885.1) (Wei et al. 2022) and *R. maidis* (GCF_003676215.2) (Chen et al. 2019), by identifying syntenic regions that share complete single-copy orthologs from the BUSCO Hemiptera_odb10 dataset.

Mitogenome Assembly and Annotation

The mitogenome was assembled separately using the Falcon preassembly reads noted above. These were mapped to a publicly available mitogenome reference of HWA collected in Taiwan (MT263947) (Yeh et al. 2020) using minimap2. Mapped reads were assembled into contigs using Flye (Kolmogorov et al. 2019), and then contigs were used as input for MitoHiFi (Uliano-Silva et al. 2023) where they were circularized and annotated. Tandem Repeats Finder (Benson 1999) was used to identify tandem repeats signaling the putative control and repeat regions in noncoding locations of the mitogenome. The mitogenome was then visualized using MITOZ (Meng et al. 2019).

Supplementary Material

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

Acknowledgments

Michael Montgomery collected sample# 19-124-03. Jackson Eyres coordinated bioinformatic expertise at the Ottawa Research and Development Centre. J. Spencer Johnston performed genome size estimation by flow cytometry at the Department of Entomology, Texas A&M University, College Station, Texas, USA. Swati Sinha and Anna Lazar completed the repeat annotation in coordination with L.H. S.W.S. would like to acknowledge the University of Toronto McLaughlin Centre and his endowed position as Northbridge Chair in Paediatric Research at The Hospital for Sick Children (Toronto, Canada).

Funding

This work was supported by Agriculture and Agri-Food Canada (project number J-002279 "Systematics of Invertebrate Pests'); the U.S. Department of Agriculture Forest Service, Northern Research Station; the National Science Foundation (grant numbers DEB-1655182 to C.D.V.D. and DEB-1655177 to G.R.B.); Genome Canada (grant number 18107 to the Canada BioGenome Project); the Utah Agricultural Experiment Station, Utah State University (paper number 9852); the Wellcome Trust (grant numbers 222155/Z/20/Z, 226458/Z/22/Z, and 226083/Z/22/Z); the Biotechnology and Biological Sciences Research Council (grant numbers BB/W019108/1, BB/T015608/1, and BB/X018695/1); the UK Medical Research Council (grant number MR/S000453/1); and the European Molecular Biology Laboratory (EMBL) core funding and the EMBL transversal research themes funding under the new scientific program. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

Data Availability

All genome sequencing data (long-reads, linked-reads, short-reads, and Hi-C reads) and the genome assembly (GCA_045999825.2) can be accessed under the NCBI BioProject PRJNA1187642. Genome annotations are available from https://ftp.ebi.ac.uk/pub/ensemblorganisms/Adelges_tsugae/GCA_045999825.2/. RepeatMasker annotations and the RepeatModeller2 library may be accessed at https://ftp.ebi.ac.uk/pub/databases/ensembl/projects/cbp/adelges_tsugae_gca045999825v2/repeat/ and https://ftp.ebi.ac.uk/pub/databases/ensembl/repeats/unfiltered_repeatmodeler/species/adelges_tsugae/, respectively.

Literature Cited

Aken BL, et al. The Ensembl gene annotation system. Database. 2016:2016:baw093. https://doi.org/10.1093/database/baw093.

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

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

Bateman A, et al. UniProt: the universal protein knowledgebase in 2023. Nucleic Acids Res. 2023:51:D523–D531. https://doi.org/10.1093/nar/gkac1052.

Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999:27:573–580. https://doi.org/10.1093/nar/27.2.573.

Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021:18:366–368. https://doi.org/10.1038/s41592-021-01101-x.

Bushnell B. BBMap: a fast, accurate, splice-aware aligner. LBNL-7065E. Lawrence Berkeley National Laboratory; 2014. https://escholarship.org/uc/item/1h3515gn (Accessed 25 April 2025).

Canadian Food Inspection Agency. 2024. Hemlock woolly adelgid—*Adelges tsugae* (Annand). https://inspection.canada.ca/en/plant-health/invasive-species/insects/hemlock-woolly-adelgid (Accessed 30 May 2024).

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit—interactive quality assessment of genome assemblies. G3 (Bethesda). 2020:10:1361–1374. https://doi.org/10.1534/g3.119.400908.

Chen W, et al. Genome sequence of the corn leaf aphid (*Rhopalosiphum maidis* Fitch). GigaScience. 2019:8:giz033. https://doi.org/10.1093/gigascience/giz033.
Chin C-S, et al. Phased diploid genome assembly with single-molecule real-time sequencing. Nat Methods. 2016:13:1050–1054. https://doi.org/10.1038/nmeth.4035.
Coombe L, et al. LongStitch: high-quality genome assembly correction and scaffolding using long reads. BMC Bioinform. 2021:22:534. https://doi.org/10.1186/s12859-021-04451-7.
Dial DT, et al. Whole-genome sequence of the Cooley spruce gall adelgid, *Adelges cooleyi* (Hemiptera: Sternorrhyncha: Adelgidae). G3 (Bethesda). 2023:14:jkad224. https://doi.org/10.1093/g3journal/jkad224.
Dudchenko O, et al. 2018. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under $1000 [preprint]. bioRxiv 254797. https://doi.org/10.1101/254797.
Durand NC, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016:3:95–98. https://doi.org/10.1016/j.cels.2016.07.002.
Edwards RJ, Dong C, Park RF, Tobias PA. 2022. A phased chromosome-level genome and full mitochondrial sequence for the dikaryotic myrtle rust pathogen, Austropuccinia psidii [preprint]. bioRxiv 489119. https://doi.org/10.1101/2022.04.22.489119.
Ellison AM, Orwig DA, Fitzpatrick MC, Preisser EL. The past, present, and future of the hemlock woolly adelgid (*Adelges tsugae*) and its ecological interactions with eastern hemlock (*Tsuga canadensis*) forests. Insects. 2018:9:172. https://doi.org/10.3390/insects9040172.
Emilson C, et al. 2018. Hemlock woolly adelgid management plan for Canada. Natural Resources Canada, Canadian Forest Service, Great Lakes Forestry Centre. Information Report Number GLC-X-21. p. 26.
Evans AM, Gregoire TG. A geographically variable model of hemlock woolly adelgid spread. Biol Invasions. 2007:9:369–382. https://doi.org/10.1007/s10530-006-9039-z.
Fidgen JG, Whitmore MC, Macquarrie CJK, Turgeon JJ. Detection of *Adelges tsugae* (Hemiptera: Adelgidae) wool using Velcro-covered balls. Can Entom. 2021:153:640–650. https://doi.org/10.4039/tce.2021.24.
Fitzpatrick MC, Preisser EL, Porter A, Elkinton J, Ellison AM. Modeling range dynamics in heterogeneous landscapes: invasion of the hemlock woolly adelgid in eastern North America. Ecol Appl. 2012:22:472–486. https://doi.org/10.1890/11-0009.1.
Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020:117:9451–9457. https://doi.org/10.1073/pnas.1921046117.
Gavrilov-Zimin I, Stekolshchikov A, Gautam DC. General trends of chromosomal evolution in aphidococca (Insecta, Homoptera, Aphidinea + Coccinea). Comp Cytogenet. 2015:9:335–422. https://doi.org/10.3897/CompCytogen.v9i3.4930.
Girgis HZ. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. BMC Bioinform. 2015:16. https://doi.org/10.1186/s12859-015-0654-5.
Havill NP, et al. Ancient and modern colonization of North America by hemlock woolly adelgid, *Adelges tsugae* (Hemiptera: Adelgidae), an invasive insect from East Asia. Mol Ecol. 2016:25:2065–2080. https://doi.org/10.1111/mec.13589.
Huang C, Ji B, Shi Z, Wang J, Yuan J, Yang P, Xu X, Jing H, Xu L, Fu J, et al. A comparative genomic analysis at the chromosomal-level reveals evolutionary patterns of aphid chromosomes. Commun Biol. 2025:8. https://doi.org/10.1038/s42003-025-07851-0.
Johnston JS, Bernardini A, Hjelmen CE. Genome size estimation and quantitative cytogenetics in insects. In: Brown SJ, Pfrender ME, editors. Insect genomics: methods and protocols. Springer; 2019. p. 15–26.
Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019:37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
Leung K, et al. Next-generation biological control: the need for integrating genetics and genomics. Biol Rev Camb Philos Soc. 2020:95:1838–1854. https://doi.org/10.1111/brv.12641.
Li F, et al. Insect genomes: progress and challenges. Insect Mol Biol. 2019:28:739–758. https://doi.org/10.1111/imb.12599.
Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021:37:4572–4574. https://doi.org/10.1093/bioinformatics/btab705.
Li JX, Coombe L, Wong J, Birol I, Warren RL. ntEdit + Sealer: efficient targeted error resolution and automated finishing of long-read genome assemblies. Curr Protoc. 2022:2:e442. https://doi.org/10.1002/cpz1.442.
Li Z, et al. Phylloxera and aphids show distinct features of genome evolution despite similar reproductive modes. Mol Biol Evol. 2023:40:msad271. https://doi.org/10.1093/molbev/msad271.
Limbu S, Keena MA, Whitmore MC. Hemlock woolly adelgid (Hemiptera: Adelgidae): a non-native pest of hemlocks in eastern North America. J Integr Pest Manage. 2018:9:27. https://doi.org/10.1093/jipm/pmy018.
Lukhtanov V, Pazhenkova EA. Diversity and evolution of telomeric motifs and telomera DNA organization in insects. Biol J Linn Soc. 2023:140:536–555. https://doi.org/10.1093/biolinnean/blad068.
Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol. 2021:38:4647–4654. https://doi.org/10.1093/molbev/msab199.
Marks P, et al. Resolving the full spectrum of human genome variation using Linked-Reads. Genome Res. 2019:29:635–645. https://doi.org/10.1101/gr.234443.118.
Mathers TC, et al. Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. Mol Biol Evol. 2021:38:856–875. https://doi.org/10.1093/molbev/msaa246.
McAvoy TJ, et al. Twenty-two-year study of the spread and impact of hemlock woolly adelgid and elongate hemlock scale in southwest Virginia. J For. 2025:123:103–131. https://doi.org/10.1007/s44392-024-00005-w.
McClure MS. Biological control of hemlock woolly adelgid in the eastern United States. FHTET-2000-08. USDA Forest Service, Forest Health Technology Enterprise Team; 2001. p. 19.
Meng G, Li Y, Yang C, Liu S. Mitoz: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Res. 2019:47:e63. https://doi.org/10.1093/nar/gkz173.
Mohamadi H, Khan H, Birol I. ntCard: a streaming algorithm for cardinality estimation in genomics data. Bioinformatics. 2017:33:1324–1330. https://doi.org/10.1093/bioinformatics/btw832.
Morin RS, Liebhold AM, Gottschalk KW. Anisotropic spread of hemlock woolly adelgid in the eastern United States. Biol Invasions. 2009:11:2341–2350. https://doi.org/10.1007/s10530-008-9420-1.
Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020:21. https://doi.org/10.1186/s13059-020-02134-9.
Roach MJ, Schmidt SA, Borneman AR. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinform. 2018:19:1–10. https://doi.org/10.1186/s12859-018-2485-7.
Smit AFA, Hubley R, Green P. 2015. RepeatMasker Open-4.0. 2013-2015. http://www.repeatmasker.org.

Song L, Florea L, Langmead B. Lighter: fast and memory-efficient sequencing error correction without counting. Genome Biol. 2014:15:509. https://doi.org/10.1186/s13059-014-0509-9.
Uliano-Silva M, et al. Mitohifi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. BMC Bioinform. 2023:24:288. https://doi.org/10.1186/s12859-023-05385-y.
Weglarz KM, Havill NP, Burke GR, von Dohlen CD. Partnering with a pest: genomes of hemlock woolly adelgid symbionts reveal atypical nutritional provisioning patterns in dual-obligate bacteria. Genome Biol Evol. 2018:10:1607–1621. https://doi.org/10.1093/gbe/evy114.
Wei HY, et al. Chromosome-level genome assembly for the horned-gall aphid provides insights into interactions between gall-making insect and its host plant. Ecol Evol. 2022:12:e8815. https://doi.org/10.1002/ece3.8815.
Yeh H-T, Ko C-C, Wu L-W. The first complete mitochondrial genome of *Adelges tsugae* Annand (Hemiptera: Adelgidae). Mitochondrial DNA B. 2020:5:2288–2290. https://doi.org/10.1080/23802359.2020.1772682.
Yeo S, Coombe L, Warren RL, Chu J, Birol I. ARCS: scaffolding genome drafts with linked reads. Bioinformatics. 2018:34:725–731. https://doi.org/10.1093/bioinformatics/btx675.
Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. Bioinformatics. 2023:39:btac808. doi:10.1093/bioinformatics/btac808.

Associate editor: Christopher Wheat