Chromosome scale genomes of two invasive Adelges species identifies host immunity and developmental gene targets for selective pest control

AM Glendening[1]*, C Stephens[2]*, VS Vuruputoor[1]*, T Chaganti[2,6], MN Myles[1], DL Stern[3], SA Hogenhout[4], TC Mathers[5], N Pauloski[7,8], TA Cernak[2,+] & JL Wegrzyn[1,8,+] & KC Fetter[1,9,+]

*Co-first authors
+Co-corresponding authors

[1]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA 06269
[2]Department of Medicinal Chemistry, University of Michigan, Ann Arbor, MI, USA 48109
[3]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA, 20147
[4]Department of Crop Genetics, John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK
[5]Wellcome Sanger Institute: Cambridge, UK
[6]Canton High School, Canton, MI, USA 48187
[7]Department of Molecular and Cell Biology, University of Connecticut, Storrs, CT, USA 06269
[8]Institute for Systems Genomics, University of Connecticut, Storrs, CT, USA 06269
[9]Southeastern Fruit & Tree Nut Research Station, Agriculture Research Service, USDA, Byron GA, USA 31008

Adam Glendening: 0009-0002-0508-6547
Cole Stephens: 0009-0002-5150-3401
Vidya Vuruputoor: 0000-0002-5836-7054
Tesko Chaganti: 0009-0003-3197-6241
Meghan Myles: 0009-0006-9268-8990
David Stern: 0000-0002-1847-6483
Saskia Hogenhout: 0000-0003-1371-5606
Thomas Mathers: 0000-0002-8637-3515
Nicole Pauloski: 0009-0004-6135-7078
Tim Cernak: 0000-0001-5407-0643
Jill Wegrzyn: 0000-0001-5923-0888
Karl Fetter: 0000-0002-9234-9300

Keywords: adelgids, reference genome, forest tree health, juvenile hormone, aminopeptidases, invasive species, conservation

Abstract: Two invasive hemipteran adelgids are associated with widespread damage to several North American conifer species. *Adelges tsugae,* hemlock woolly adelgid, was introduced from East Asia and reproduces parthenogenetically in North America, where it has rapidly decimated *Tsuga canadensis* and *Tsuga caroliniana* (the eastern and Carolina hemlocks, respectively). *Adelges abietis*, pineapple gall adelgid, introduced from Europe, forms distinctive pineapple-shaped galls on several native spruce species. While not considered a major forest pest, it weakens trees and increases susceptibility to additional stressors. Broad-spectrum insecticides that are often used to control adelgid populations can have off-target impacts on beneficial insects. Whole genome sequencing was performed on both species to aid in development of targeted solutions that may minimize ecological impact. The assembled *A. tsugae* and *A. abietis* genomes, pooled from several parthenogenetic

females, are 220.75 Mbp and 253.16 Mbp, respectively. Each consists of eight autosomal chromosomes, as well as two sex chromosomes (X1/X2), supporting the XX-XO sex determination system. The genomes are over 96% complete based on BUSCO assessment. Genome annotation identified 11,424 and 12,060 protein-coding genes in *A. tsugae and A. abietis*, respectively. Comparative analysis across 29 Hemipteran species and 14 arthropod outgroups identified 31,666 putative gene families. Gene family analysis revealed contrasting gene family dynamics across a range of functional categories with significant expansions in immune-related aminopeptidases (*ERAP1*) and juvenile hormone binding proteins (*JHBP*), contractions in juvenile hormone acid methyltransferases (*JHAMT*) in developmental pathways, and conservation of nicotinic acetylcholine receptors (*nAchR*). Structural comparisons of proteins across species of key proteins within the aforementioned gene families showed that ERAP and nAChr were highly conserved among the subset of species assessed; however, JHAMT displayed species-specific variation.

Introduction

Adelgids (*Adelges*) include at least 65 species of host-specific, sap-feeding hemipterans within Aphidomorpha (Favret et al. 2015). These insects typically exhibit a life strategy similar to their aphid relatives, primarily feeding on plant sap from phloem tissue, yet they are uniquely associated with conifers (Sano and Ozaki 2012). A key distinction to their aphid relatives is the longer, more flexible stylet of adelgids which permits feeding on both vascular and nonvascular tissue (Dancewicz et al. 2021). Additionally, adelgids do not exhibit viviparity, a reproductive trait common in aphids (Havill and Foottit 2007; Chakrabarti 2018). The introduction of non-native adelgid species into North America has resulted in conservation issues for some conifer species. Notable examples include the balsam woolly adelgid (*Adelges piceae*) and hemlock woolly adelgid (*Adelges tsugae*) which have devastated fir and hemlock populations, respectively (Ellison et al. 2018; Fei et al. 2019; Zilahi-Balogh et al. 2017).

The pineapple gall adelgid or eastern spruce gall adelgid (*A. abietis*; Figure 1A) is anholocyclic on spruce (*Picea*) species, consisting of two parthenogenetic generations per year, with fundatrices (asexual females) overwintering at the base of buds. A large number of galls can occur on a single tree causing damage to its crown (Carter 1971) (Figure 1B). Dispersal of *A. abietis* occurs through movement of both winged gallicolae and crawlers. In North America, *A. abietis* has been observed on black spruce (*P. nigra*), white spruce (*P. glauca*), red spruce (*P. rubens*), as well as non-native European spruces, like the Norway spruce (*P. abies*) (Cook et al. 2012) (Figure 1C). Considered non-native to North America (Plumb 1952), the evolutionary and invasion history of this species remains unclear.

Evolving in eastern Asia, *A. tsugae* was introduced into eastern North America during the early 20th century and first collected in Virginia in 1951 (Figure 1D) (Havill et al. 2006). In its native range, *A. tsugae* alternates hosts between hemlock and spruce (Havill et al 2016). The invasive *A. tsugae* lineages reproduce parthenogenetically in North America, with endemic spruce species unsuited for sexual reproduction (McClure 1989). HWA disperse as first instar-nymphs (i.e. crawlers) and are moved by birds (Russo et al. 2019), humans, or wildlife (McClure 1990). Winged morphs are mobile, but do not contribute to the next generation, as this sexual phase is sterile in eastern North America (Russo et al. 2019). HWA crawlers synthesize white, woolly sacs (containing an average of 400 to 700 eggs) on the underside of hemlock branches near the base of the petiole where they feed and remain settled for the remainder of their life (Montgomery et al. 2009) (Figure 1E). The rapid population growth of *A. tsugae* within a single year, combined with its capacity to hitchhike on vertebrates, has enabled it to become a major pest of native eastern North American hemlocks (*Tsuga canadensis* and *Tsuga caroliniana*; Figure 1F), leading to significant ecological and economic threats to forest ecosystems (Orwig and Foster 1998)

Genomic resources in *Adelges* are growing, although the vast majority of adelgids lack a reference genome or other genetic resources (Dial et al. 2023). Mitochondrial sequences from *A. tsugae* and other species determined the invasion history of *A. tsugae*, pinpointing southern Japan as the source of the invasive eastern North American population (Havill et al. 2006a). A study using microsatellites and a mitochondrial marker demonstrated that the HWA populations in western North America are closely related to HWA populations in southern Japan, and likely arrived to western North America through vicariance of the circum-boreal forest during the Plio-Pleiostocene, or possibly via migratory birds (Havill et al. 2016). The fully sequenced *A. tsugae* mitochondrial genome enhanced the phylogenetic resolution of the species and resolved relationships among Aphiodoidea (Havill et al. 2006b; Yeh, Ko, and Wu 2020). Metagenomic sequencing of *A. tsugae* eggs enabled full characterization of the endosymbiont genomes, enabling a better understanding of their contributions to adelgid physiology (Weglarz et al. 2018).

This study contributes reference genomes for two species of *Adelges*. A combination of sequencing technologies and assemblers were used to generate two high-quality annotated genomes. The evolutionary histories of the genomes, and gene family expansions and contractions across a phylogenetically diverse set of insects, was investigated. The evolution of enzymes selected from key gene families that underwent significant expansion in their history was also investigated. Finally, protein modeling was used to examine rapidly evolving gene families as targets for selective pest control.

Methods & Materials:

DNA Sampling, Extraction, Preparation, and Sequencing

A. abietis were collected from a single gall growing on *Picea rubens* at Timberline Mountain in Davis, WV, USA on 21 July 2019. Genomic DNA for a 10X Chromium Linked-Read library was extracted from a pooled sample of approximately 30 individuals (detailed methods in File S1). Isolated DNA was quantified on a Qubit 2.0 fluorometer and DNA fragment size was checked by running on a 1.0% agarose gel. The Linked-Read library was prepared following the manufacturer's protocol and sequenced on an Illumina Nextseq 550. Hi-C Illumina short reads were generated for scaffolding from a single individual collected from a single gall growing on the same *Picea rubens* accession on 16 August 2021. This insect was frozen in liquid nitrogen in an Eppendorf tube, ground with a plastic pestle in 1mL 1% formaldehyde for 20 minutes. Formaldehyde was quenched by adding 110 uL 1.25M Glycine for 15 min. Sample was spun at top speed for 15 min in tabletop centrifuge, liquid was removed and replaced with Phosphate Buffered Saline, spun again, the supernatant was removed and sample was stored at -80°C prior to shipping to Phase Genomics (Seattle, WA) for Hi-C library preparation and sequencing.

A. tsugae were collected from a single accession of *T. canadensis* growing at the Mountain Research Station in Waynesville, NC (35.487 °N, –82.967 °W) in October 2023. Approximately 4 feet of one branch was collected and stored at 4 °C for 48 hours. Individual adelgids were collected into Macherey-Nagel Lysis Buffer T1 using forceps and a dissection microscope. A total of 400 sistens (adelgid nymphs) were collected in total to provide sufficient material for DNA extraction. DNA was extracted from the pooled sample using the Qiagen MagAttract high molecular weight (HMW) DNA Kit, with modifications based on (Mao et al. 2023) and further adjustments to timings, temperatures, and agitation intensities (File S1). Isolated DNA quality was measured with Thermo Scientific NanoDrop One spectrophotometry, Thermo Scientific Qubit 4 DNA fluorometry, and Agilent 4200 TapeStation Genomic DNA electrophoresis. Oxford Nanopore Technologies Ligation Sequencing Kit v14 (SQK-LSK114) was used to prepare the library. DNA repair was performed at 20 °C for 20 min followed by 10 min at 65 °C to inactivate the enzymes. DNA was eluted at 37 °C. The final library was quantified using a Qubit 4 fluorometer. The flow cell was loaded three times each with 18 fmol of library.

RNA Sampling, Extraction, Preparation, and Sequencing

Individual *A. tsugae* were collected from branches of *T. canadensis* growing in Leesburg, VA, USA, and *A. abietis* were collected from galls collected in Davis, WV. Salivary glands were dissected by placing insects in phosphate buffered saline and then grabbing insects with forceps at the pronotum and anterior abdomen and pulling apart. The salivary glands remain attached to the brain on the anterior portion. The glands were isolated from other tissue using Minutien insect pins, with their tips bent approximately 30 degrees, mounted in micro dissecting needle holders. Separated glands and the remaining carcass were collected separately into 100 uL of Arcturus PicoPure Extraction buffer. Total RNA was prepared using the Arcturus PicoPure RNA Isolation kit including the optional DNAse step. Barcoded RNASeq libraries were prepared for Illumina NextSeq 550 sequencing (150bp PE) with a method described previously (Cembrowski et al. 2018). Libraries with fewer than 15 million reads or mean read length below 70 bp were excluded as evidence for structural annotation (Table S1).

Genome assembly

A. abietis: A total of 1.19 Gb of 10X linked reads were assembled into a draft genome with Supernova with default parameters. Hi-C reads were aligned to the *Supernova* draft assembly with Juicer v1.6.2 (Durand, Shamim, et al. 2016). The 3D-DNA pipeline (Dudchenko et al. 2017) was applied with default settings in “haploid mode” to correct mis-assemblies and generate chromosome-scale super scaffolds. Manual review of the 3D-DNA assembly was conducted with Juicebox Assembly Tools (Durand, Robinson, et al. 2016). The scaffolded assembly was screened for contamination with BlobTools v1.0.1 (Laetsch and Blaxter 2017; Sujai Kumar et al. 2013) using aligned 10X Genomics linked reads and taxonomy

information from BLASTN v2.2.31 (Camacho et al. 2009) searches against the National Center for Biotechnology Information (NCBI) nucleotide database (nt, downloaded October 13, 2017). To calculate per-scaffold coverage, 10X Genomics linked reads were debarcoded with process_10xReads.py from the proc10xG package (https://github.com/ucdavis-bioinformatics/proc10xG) and aligned to the reviewed 3D-DNA assembly with BWA mem v0.7.7 (Vasimuddin et al. 2019). Finally, the decontaminated assembly was sorted by chromosome length, and assessed for completeness with BUSCO v5.4.5 (Manni et al. 2021), QUAST v5.2.0 (Gurevich et al. 2013), and by comparing K-mer content of the debarcoded 10X Genomics linked reads to the assembly with KAT comp v2.3.1(Mapleson et al. 2017; Shen, Sipos, and Zhao 2024). The genome was assessed for accuracy with Merqury v1.3 (Rhie et al. 2020).

A. tsugae: Nanopore reads were basecalled from a single PromethION R10.4.1 flow cell running MinKnow v23.07.12. Initial statistics on the read count, length, and base quality were assessed using NanoPlot v1.33.0 (De Coster and Rademakers 2023). Reads passing an initial quality cutoff of >Q10 were screened for DNA contamination against RefSeq genomes (release 221) of archaea, bacteria, fungi, plants, and viruses using Centrifuge v1.0.4-beta and reassessed using NanoPlot after removal of contaminant-classified reads (Kim et al. 2016). The estimated genome size was obtained via kmer count using kmerfreq v4.0 (H. Wang et al. 2020) with GCE v1.0.2 (H. Wang et al. 2020). Filtered reads were assembled using both Flye v2.9.1 (Kolmogorov et al. 2019) and Canu v2.2 (Koren et al. 2017). Initial draft assemblies and each subsequent iteration were assessed using BUSCO hemiptera_odb10 for completeness, Merqury v1.3 (Rhie et al. 2020) for quality, and QUAST for contiguity. The Canu assembly was selected for downstream analysis. Heterozygosity was reduced for draft assemblies using Purge Haplotigs v1.1.2 (Roach, Schmidt, and Borneman 2018). The contig-level assemblies were scaffolded against the chromosome-level *A. abietis* genome using RagTag v2.1.0 (Alonge et al. 2022), reordering contigs into putative chromosomes, in descending order by length. The final assembly was re-assessed for quality metrics.

The *A. tsugae* and *A. abietis* final assemblies were screened for contaminant and adaptor sequences using NCBI Foreign Contamination Screen (FCS-GX and FCS-adaptor) v0.5.4 (Astashyn et al. 2024).

Structural and functional annotation

The repeat library for each species was generated *de novo* via RepeatModeler v2.0.4 (Flynn et al. 2020). This library was used to softmask the assembled references with RepeatMasker v4.1.4 (Smit 2015). The softmasked reference genomes were provided to EASEL (https://gitlab.com/PlantGenomicsLab/easel) v1.5 (regressor set to 70) to facilitate the structural and functional identification of protein-coding genes (Hart et al. 2020). Transcriptomic evidence provided gene

prediction support within EASEL through the alignment of RNASeq reads to their corresponding genome. The *A. tsugae* annotation was informed by four salivary gland libraries, four carcass libraries (salivary glands removed) (SRR30936309), and an adult overwintering sistens library (SRR1198669). The *A. abietis* annotation was informed by two salivary gland libraries, four carcass libraries, and a whole body library (SRR30936310). Additional protein model support for annotation was provided by NCBI's RefSeq protein database (release 208) (Goldfarb et al. 2025) and OrthoDB v11 (Kuznetsov et al. 2023).

Genome synteny was assessed using protein-protein comparisons of the *A. tsugae*, *A. abietis*, *A. cooleyi* (using the top 10 scaffolds, following the approach of (Li et al. 2023), and *Daktulosphaira vitifoliae* annotations using MCScanX (Wang et al. 2012), and was visual with SynVisio (Bandi and Gutwin 2020).

Comparative genomics among hemipteran

All Hemipteran species with genome annotations available on NCBI (45 total) were considered for comparative analysis, and 14 related species were used as outgroups (Table S2) (Bailey et al. 2022; Panfilio et al. 2019; Ma et al. 2021). All protein sets had a BUSCO completeness above 80% and no obvious anomalies in regards to total gene count. Orthofinder v2.5.5 was run with 45 species to classify potential orthologous genes and orthogroups (Emms and Kelly 2019). The longest sequence from each orthogroup was used as a representative to assign a function with EnTAP v1.2.1 (Hart et al. 2020) that integrated EggNOG v5.0.2 (Huerta-Cepas et al. 2019), Refseq Invertebrate release 224, and Swiss-Prot (release 2024_03).

Significant orthogroup expansions were identified with Cafe v5.1 (Mendes et al. 2021) based on birth and death process models. Gene turnover was estimated using the maximum likelihood inference method. The distances in the rooted tree obtained from the single-copy genes (from OrthoFinder) were transformed into ultrametric units (Table S3). Orthogroups with representation in only one species (CAFE5 clade_filter), overrepresentation in one species (one species ≥50 orthologs and all others ≤3), and extreme differentials between minimum and maximum counts (63 or more orthologs), were removed. Gene family expansion or contraction was considered significant at $P < 0.01$.

Assessing significantly expanded/contracted gene families

Hidden Markov models (HMMs) for Juvenile hormone (JH) acid methyltransferase (*JHAMT*; PTHR43464), Survival of motor neuron (SMN; PTHR39267), Endoplasmic reticulum aminopeptidase protein 1 (*ERAP1*; PTHR11533), Juvenile hormone binding protein (*JHBP*; PF06585), and nicotinic acetylcholine receptor (nAChR) alpha-7 subunit were retrieved from PantherDB (v17.0; (Thomas et al. 2022)) and Pfam (v35.0; (Finn et al. 2014)), cross-referenced through InterProScan (v5.59-91.0; (Jones et al. 2014)) to confirm domain architectures. HMM profiles were processed

into searchable databases using hmmpress (HMMER v3.3.2; (Z. Zhang and Wood 2003)) with default parameters. Subsequent batch CD-Search analysis (NCBI CDD v3.20; (Marchler-Bauer et al. 2015)) validated domain presence across targets using RPS-BLAST (E-value <1e−10, coverage >70%). Validated sequences were aligned with MAFFT v7.526 (L-INS-i algorithm; BLOSUM62 matrix, gap penalty=1.53; (Katoh and Toh 2008)) via the Computational Biology Research Consortium (CBRC) web server, followed by refinement with MaxAlign v1.3 (Gouveia-Oliveira, Sackett, and Pedersen 2007) to retain sequences covering ≥75% of consensus length and columns with ≤40% gaps. Genes will be referred to in italics (i.e, *ERAP1, JHAMT*), and protein names by the non-italicized version (i.e, ERAP1, JHAMT).

Homology modeling and structural comparison

Structures of insect proteins were predicted using AlphaFold3 (AF3) (Abramson et al. 2024). ERAP and JHAMT proteins were each folded as monomers. As nAChR proteins are pentameric, their structure prediction was simplified by folding all nAChR proteins as homopentamers, although only some nAChR subtypes are known to homopentamerize in vivo (Barrantes 2023). Models were evaluated for quality using their accompanying predicted template modeling (pTM) confidence scores. Proteins whose resulting models had a pTM score below 0.75 were excluded from the analysis in order to prevent the data from being skewed by outlier sequences of abnormally low homology. TM scores were calculated using Zhang et al's TM align algorithm and the scores normalized to the average length of the pair of compared proteins (Zhang and Skolnick 2004; Xu and Zhang 2010). Scores were then sorted by in-group (adelgids; *A. tsugae*, *A. abietis*, *A. cooleyi*) and outgroup (other insects; *R. maidis*, *A. mellifera*, *D. plexippus*, *D. melanogaster*) and arrayed combinatorially for visualization via heatmap.

Sequence similarity analysis

Protein sequences (i.e. amino acids) were aligned in MEGA v11.0.13 (Sudhir Kumar et al. 2018) with MUSCLE v3.8.31 (Edgar 2004). Amino acid sequence similarities were calculated with the "Sequence Manipulation Suite" for JavaScript using the "Ident and Sim" tool, with similar amino acid groups set to GAVLI, FYW, CM, ST, KRH, DENQ, P (Stothard 2000). Proteins resulting in models with low pTM confidence scores were excluded as mentioned above. Percent similarity scores were then sorted by in-group (adelgids; *A. tsugae*, *A. abietis*, *A. cooleyi*) and outgroup (other insects; *R. maidis*, *A. mellifera*, *D. plexippus*, *D. melanogaster*) and visualized via a kernel density estimate (KDE) plot (Table S4).

Binding site analysis

To identify insect protein binding site residues, full protein sequences for all selected insects were compared to the sequences of homologous proteins from other species with crystal structure data. nAChR sequences were aligned and searched for the conserved active site loop regions which have been previously reported (Gharpure,

Noviello, and Hibbs 2020). JHAMT sequences were aligned to those from three species of moth (*B*. *mori*, *M*. *sexta*, and *O*. *furnacalis*) and identified residues matched to the substrate contacting residues and cofactor contacting residues from the *B*. *mori* crystal structure (Protein Data Bank (PDB) ID: 7EC0, (Guo et al. 2021)) . ERAP1 sequences were aligned with a human ERAP1 sequence with a known crystal structure (PDB ID: 6M8P, (Maben et al. 2021)). Active site residues were identified by alignment in the crystal structure. C-terminal selectivity region residues were identified by alignment with residues implicated in substrate selectivity (Giastas et al. 2019). The resulting sequences were then compiled and visualized as logo plots (Crooks et al. 2004). Full sequences for ERAP1*,* JHAMT*,* and nAChR can be found in Table S4. Isolated binding site sequences for ERAP1, JHAMT, and nAChR can be found in Table S5, Table S6, and Table S7, respectively. Isolated active site sequences for ERAP*,* JHAMT*,* and nAChR can be found in Table S8.

Results and Discussion:

<u>Genome sequencing and assembly:</u>

Adelges abietis 10X sequencing generated 119M Illumina short reads (Table S9). The Supernova assembled genome produced a reference of 290.32 Mb in 9,607 contigs with an N50 of 4.77 Mb. Hi-C scaffolding of the *A. abietis* genome initially produced an assembly 253.17 Mb in length with 2998.36 N's per 100 Kb, an N50 of 27.91 Mb, and an accuracy value (QV) of 50.87 (Table S10). However, manual curation revealed structural inconsistencies within Chromosome 1 (signature of translocations in the two haplotypes). Manual adjustments, supported by HiC reads, resolved the assembly into ten chromosomes (Figure S1). Post manual corrections, the assembly was 253.17 Mb in length, with 2884.06 N's per 100 Kb, an N50 of 21.97 Mb, and the QV remained at 50.87. A total of 87.17% of sequences were contained within the ten chromosomes (eight autosomal and two sex chromosomes). BUSCO assessment found 2471 of 2510 expected complete (C:97.2%) single-copy orthologs from the Hemiptera_odb10 database and 47 duplicated (D:1.3%), leaving only 8 fragmented (F:0.3%) and 31 missing (M:1.2%) (Table 1).
Contaminant screening of *A. abietis* post assembly removed 61 (2,740,304 bp) unplaced scaffolds, of which 54 were classified as *Ca*. *Vallotia tarda* (33), *Ca. Vallotia lariciata* (20), *Ca. Profftia tarda* (1), or *Ca*. *Arsenophonus triatominarum* (0.04%) (Table S11).

A. tsugae sequencing generated 100.78 Gb of long reads (ONT) with an 11.02 Kb N50, of which 89.94 Gb passed an initial Q10 quality score threshold and 77.71 Gb passed contaminant screening (File S2; Table S9). Among the 4,037,933 (13.60%) reads classified as contaminants, the most abundant species was *Candidatus Pseudomonas adelgestsugas*, one of *A. tsugae*'s two dual-obligate endosymbiotic bacteria, with just over 3M unique reads representing 75.7% of those removed. The other paired endosymbiont, *Ca. Annandia adelgestsuga*, was not specifically

identified. However, *Ca. Annandia pinicola* (4.43%) and 19 intraspecific variants of *Buchnera aphidicola* (together 0.59%) were detected. Given the similarity of the *Buchnera* and *Annandia* genomes, as well as the similarity-based metagenomic classification technique used, these reads may represent *Ca. Annandia adelgestsuga (Weglarz et al. 2018)*. The second most abundant classification was *Serratia symbiotica* (9.05%), an endosymbiont unique to eastern North America and *Tsuga sieboldii*-specific *A. tsugae* lineages in Japan (C. D. von Dohlen et al. 2013). This was followed by *Porphyrobacter* sp. GA68 (0.66%), *Cellulophaga* sp. HaHaR_3_176 (0.18%), and *Rhizoctonia solani* (0.14%), an indoor air bacterium, a marine bacterium, and a plant-pathogenic fungus, respectively (Table S12).

The genome size of *A. tsugae* was estimated at 267.97 Mb yielding 290x read coverage. The Canu assembled draft genome contained 2,131 contigs with a 7 Mb N50. Reducing heterozygosity yielded 377 contigs with an N50 of 9 Mb. Scaffolding against the chromosome scale *A. abietis* reference yielded 345 contigs with a 21 Mb N50. This final assembly is 221.26 Mb in length, 95.49% of which is contained within ten putative chromosomes (eight autosomes and two sex chromosomes). This reference reports 1.45 N's per 100 kb, an N50 of 21.14 Mb, reflecting exceptional continuity. The Merqury QV score of 42.88 indicates a base-level accuracy of 99.99%, further underscoring the high quality of the assembly. BUSCO assessment found 2464 of 2510 expected complete (C:99.3%) single-copy orthologs from the Hemiptera_odb10 database and 28 duplicated (D:1.1%), leaving only 10 fragmented (F:0.4%) and 8 missing (M:0.3%) (Table 1).

Of the assembled Aphidomorpha genomes, Adelgidae and Phylloxeridae tend to have smaller genomes than the Aphididae. (Table S2) While Aphididae genomes are variable in size, ranging between 300-600 Mb (Huang et al. 2025), the average genome size of the characterized Aphididae is 369.4 Mb. In contrast, all of the assessed Adelgidae and Phylloxeridae genomes are smaller than 300 Mb. *A. tsugae* is the smallest, with a genome size of 220.75 Mb, and *A. abietis*, *A. cooleyi*, and *Daktulosphaira vitifoliae* have estimated genome sizes of 253.16 Mb, 270.2 Mb, and 282.6 Mb, respectively (Rispe et al. 2020; Dial et al. 2023).

RNA sequencing and quality control

For *Adelges tsugae*, eight RNASeq libraries passed quality control (150bp PE), ranging from 18.1 to 25.3M reads per library. For *A. abietis*, seven RNASeq libraries passed quality control (150bp PE), ranging from 15.5 to 24.4M reads per library (Table S1).

Structural and functional annotation:

A total of 17.66% of the *A. tsugae* genome and 28.31% in *A. abietis* were softmasked and identified as repetitive DNA (Table 1; Table S13; Figure 2A, 2B). The overall repeat content masked in both *A. abietis* and *A. tsugae* are lower than the

repeat content of *A. cooleyi* (43.6% of genome masked; (Dial et al. 2023)) and the phylloxeran *Daktulosphaira vitifoliae* (53.26% of the genome masked; (Li et al. 2023)), on the other hand, the range of repeats masked in the adelgid genomes presented here fall within that of aphids, *Myzus persicae* (22% of genome masked), and *Acyrthosiphon pisum* (29% of genome masked; (Mathers et al. 2021)). LINEs are the most abundant elements in both genomes (2.22% *A. abietis*, 2.75% *A. tsugae*) while LTRs (1.04% *A. abietis*, 0.23% *A. tsugae*) and SINEs (0.00% *A. abietis*, 0.01% *A. tsugae*) are less common, a pattern common among some Hemipterans including aphids and whiteflies (Petersen et al. 2019). *A. tsugae* differs from the pattern in DNA transposon content with only 0.46% of sequence compared to *A. abietis*'s 2.14%, though this and the overall low TE abundance may be explained by the large portion of unclassified TEs (*A. abietis* 20.59%, *A. tsugae* 11.68%). Regions near chromosome ends tend towards a higher density of repeats, with some exceptions (Figure 2C, 2D). The repeat age distribution of *A. tsugae* and *A. abietis* reveals a continuous pattern, marked by distinct proliferation phases (Figure 2A, 2B). The most recent proliferation phase in *A. tsugae* is characterized predominantly by DNA transposons and LINE elements, with minor contributions from LTR retrotransposons. Earlier activity show comparable profiles dominated by unclassified elements, though LINEs and DNA transposons maintain consistent representation across all phases. Notably, while LTR retrotransposons constitute a smaller fraction of each burst, they persist through all proliferation phases. The TE profile in *A. tsugae* differs from *A. abietis* and other aphids like *Myzus persicae* and *Metopolophium dirhodum*, which display recent expansions dominated by DNA transposons (Figure 2A, 2B; Table S13) (Baril et al. 2023).

Identification of protein-coding regions in *A. tsugae* resulted in 11,424 protein-coding genes with BUSCO completeness of C:98.0% [S:96.7% ,D:1.3%], F:0.2%, M:1.8%. A total of 12,060 *A. abietis* protein-coding genes were identified with a BUSCO completeness of C:95.9% [S:94.7%, D:1.2%], F:0.2%, M:3.9% (Table 1;Table S14; Table S15). Functional annotations, including both sequence similarity searches and alignments to EggNOG gene families, were available for 94.32% of *A. tsugae* protein-coding genes and 90.69% of *A. abietis* genes (Table S15).

Synteny analysis revealed strong conservation of chromosomal architecture across adelgid genomes, with *A. cooleyi*, *A. tsugae* and *A. abietis* supporting a structure of eight autosomes and two sex chromosomes (X1 and X2) (Figure 2E, Table S16). Adelgids, phylloxera, and aphids largely share an XX-XO sex chromosome system. Despite this, their chromosomes exhibit different evolutionary patterns. Recent comparative studies by Li et al (2023) observed frequent inter-chromosomal autosomal rearrangements (and frequent intra-X-chromosomal rearrangements) among *D. vitifoliae*, *A. cooleyi*, and *A. pisum*, and suggested that the elevated rate of autosomal rearrangements is a lineage-specific feature of aphids. However, Huang et al (2025), with a broader sampling, noted more variation among aphid lineages,

and specifically that those less speciose had fewer inter-autosomal rearrangements. Among the three genomes representing adelgids, there appears to be fewer autosomal and X chromosome rearrangements. In aphids, Li et al (2020) observed that the X chromosomes were characterized by TE expansions, and genes with low levels of expression and under weak purifying selection. Among the assembled adelgids, X1 of *A. abietis* was both larger and more repeat rich (45.9 Mb, 2,505 genes, 21% of the gene space) compared to *A. tsugae* (26.9 Mb, 1,504 genes, 13% of the gene space) (Table S17).

Comparative genomics among hemipteran

Comparative genomic analysis across Hemiptera identified 32,340 orthogroups comprising 93.69% of total proteins examined, of which 85.61% had functional annotations (Table S3). Of the 1,908 orthogroups shared across all 45 species, only 59 were single-copy orthologs. While most orthogroups (68.93%) lacked adelgid representation, 862 were adelgid-specific, with 205 shared across all three *Adelges* species. A total of 33,189, and 123 groups were unique to *A. tsugae*, *A. abietis*, and *A. cooleyi*, respectively (Figure 3A).

Phylogenetic analysis of 45 high-quality genomes found 244 gene families with significant changes in size (Table S18; Figure 3B). Comparison of gene family dynamics identified 235 significantly divergent orthogroups after filtering transposable elements (Table S19). Of these, 32 orthogroups were absent across all *Adelges*, with 31 likely lost in the adelgid-phylloxeran common ancestor (Table S20). Among lost gene families, signal transduction mechanisms were most abundant, including G-protein receptors, immunoglobulins, and PET domain proteins.

At the Adelges most recent common ancestor (MRCA) node, 24 orthogroups are expanding and 34 contracting. Within this lineage, *A. tsugae* has 19 expanding and 62 contracting orthogroups, *A. cooleyi* has 42 expanding and 27 contracting, and *A. abietis* has 29 expanding and 46 contracting. The largest expansions in *A. abietis* were associated with carbohydrate transport and metabolism, particularly ABC transporters and carboxypeptidases (Table S21). This suggests adaptation to host plant metabolism, a pattern commonly observed in specialized herbivores coping with host plant defenses (Wu et al. 2024; Wu et al. 2019). While phylloxera evolved genes for host interaction, adelgids expanded core metabolic functions, focusing on energy production (Rispe et al. 2020). This pattern aligns with highly specialized herbivores, showcasing distinct adaptive strategies between closely related species (von Dohlen and Moran 2000).

While the comparative analysis revealed broad patterns of gene family evolution in Adelgidae, a closer examination of specific gene families provides insights into potentially adaptive mechanisms. Genes associated with the suppression of host

recognition, development, and signal transduction are of particular interest.

Effector gene families

Adelgids, like other hemipterans, use effector proteins to manipulate host resistances for successful feeding. Analysis of effector gene families revealed significant expansions in *A. tsugae* and *A. cooleyi*, particularly in Endoplasmic reticulum aminopeptidase 1 (*ERAP1*). ERAP proteins belong to the M1 family of zinc-dependent aminopeptidases, characterized by specific conserved structural motifs essential for their function (Figure 4A). The structural analysis of ERAP1 reveals the presence of critical motifs, including the H-E-X-X-H-(X)18-E zinc binding motif and the G-A-M-E-N motif, which are characteristic of the gluzincin protease family (Kochan et al. 2011; Tsujimoto et al. 2020). These conserved sequences are essential for the catalytic activity of the enzyme, with the zinc-binding motif directly participating in peptide hydrolysis and the G-A-M-E-N motif contributing to exopeptidase specificity by interacting with the N-terminal amino acid of the substrate (Kochan et al. 2011). The C-terminal selectivity region of ERAP1 allows it to function as a “molecular ruler,” preferentially processing peptides of specific lengths to generate optimal substrates for downstream biological processes, as seen in insects (Kanaseki et al. 2006; Evnouchidou et al. 2014). *ERAP1* copy numbers are elevated in *Adelges tsugae* (12 copies), *A. cooleyi* (12) (Figure 2C, 2D), and *R. maidis* (14) compared to non-hemipteran insects like *A. mellifera* (8) and *D. plexippus* (10) (Table S22). This pattern mirrors findings in *Acyrthosiphon pisum*, where *ERAP1*-like C-terminal domains were enriched in salivary effectors, implicating these proteins in disrupting plant immune signaling through peptide hydrolysis of host defense molecules (Boulain et al. 2018; Chen et al. 2019).

In order to better understand the diversity of ERAP1 structures across insects, structural analysis was conducted. Similar studies were conducted for two other targets, JHAMT and nAChR, which are discussed in the following sections. ERAP1 exhibited low interspecies variability. ERAP1 sequence similarity was consistently moderate, with an average value of 43% across species (Figure 4B). ERAP1 also demonstrated high structural similarity, with an average TM-score of 0.87, indicating that its overall 3D structure is well-conserved (Figure 4C). The catalytic site of ERAP1 was highly conserved across all species (Figure 4D). In contrast, the region of the binding pocket known for substrate C-terminal binding, which influences substrate selectivity, was highly variable. This suggests that while the overall structures are largely similar, substrate selectivity among these proteins may differ (Giastas et al. 2019). The overall predicted structures of insect ERAP1 overlapped remarkably well with human ERAP1 crystal structures, further indicating a high level of structural conservation throughout animals (PDB ID: 6RQX, Figure S2).

Developmental gene families

Survival of Motor Neuron (SMN) proteins are evolutionary conserved components of the spliceosomal machinery (Sm), essential for small nuclear ribonucleoprotein (snRNP) assembly and RNA processing (Chaytow et al. 2018). They contain three domains: Gemin2-binding complex which facilitates binding to gemin proteins required for the snRNP assembly (Singh et al. 2017), Tudor domain which binds symmetrically dimethylated arginine residues in Sm proteins (Selenko et al. 2001); and YG-Box domain that mediates oligomerization of *SMN* subunits, essential for functional *SMN* complexes (Praveen et al. 2014). *SMN* orchestrates snRNP biogenesis by chaperoning Sm proteins to spliceosomal RNAs, a depletion of which causes impairment to neuronal development in insects like *D. melanogaster* (Grice and Liu 2011). The *SMN* family was expanding in *A. tsugae* (1 copy), neutral in *A. cooleyi* (12 copies), and contracting in *A. abietis* (no copies), whereas *D. plexippus* had 2 copies, and *D. melanogaster* had one copy (Table S22).

Core histone families (H2A/H2B/H3/H4) showed contractions in the adelgid MRCA, *A. tsugae,* as well as *A. abieitis*, contrasting with aphids which show expanded chromatin, remodeling protein diversity (Mathers et al. 2017; Rider, Srinivasan, and Hilgarth 2010). Interestingly, core histone families are also expanded in *A. cooleyi* (Table S22).

Juvenile hormones (JH), critical sesquiterpenoid regulators of development and reproduction in insects (Shinoda and Itoyama 2003), show contrasting evolutionary trajectories in *A. tsugae* gene families. While *juvenile hormone acid methyltransferase* (*JHAMT*), a ~33 kDa enzyme catalyzing the final JH biosynthesis step via SAM-dependent methylation of JH acids (Figure 4E) underwent contraction in *A. tsugae* (5 copies). In contrast, haemolymph juvenile hormone binding protein (JHBP) expanded markedly in *A. tsugae* (9 copies). This expansion pattern is consistent with *A. cooleyi* (*JHAMT*: 6 copies; *JHBP*: 9 copies) but contrasts sharply with *A. abietis*, which retained ancestral JHAMT levels (5 copies), with JHBP expanded to 20 copies (Table S21).

JHAMT has been implicated as a viable target for insect control, with RNAi feeding studies disrupting development in *B. dorsalis* (Zhou et al. 2022). As JH signaling is generally unique to insects, JHAMT represents an opportunity for the development of chemical insecticides with little risk of off-target effects. JHAMT was the least conserved among the three proteins studied, with an average sequence similarity of 20% compared to the other adelgids and insects (Figure 4F). Within adelgids, JHAMT proteins showed high similarity, but only low to moderate similarity when compared to the other insects. Structural similarity mirrored this pattern, displaying occasional high-scoring matches but generally exhibiting considerable structural diversity (Figure 4G). The substrate binding region was also quite variable, with

known binding residues from previous studies aligning poorly with the sequences observed in adelgids (Figure 4H) (Guo et al. 2021). The predicted structures of adelgid JHAMT enzymes differ greatly from known crystal structures of JHAMT, suggesting again that these enzyme's structures are highly variable (PDB ID: 7EC0, Figure S1). Given the rapid expansion of *JHAMT*-encoding genes in adelgids, these structural variations may result from such expansions, potentially leading to changes in selectivity or function.

Signal transduction gene families

As members of the cysteine (Cys)-Loop ligand-gated ion channel (CysLIGC) receptor family, *nAChR*s are highly conserved across insects and most animals (Marcovich et al. 2020). nAChRs are composed of five subunits, being either homopentameric or heteropentameric, and are membrane-bound, with extracellular, transmembrane, and intracellular regions (Figure 4I). Insect nAChR subunits can be separated into two distinct classes, α and β, which are primarily differentiated by the presence or absence of a disulfide linkage between two adjacent cysteine (Cys) residues in the ligand binding site. The ligand binding site is located in the extracellular region at the interface between two subunits, and is composed of six loops, with three being provided by each subunit in the interface. In order to bind a ligand, at least one of the subunits at the given interface must be an α subunit, with the other being either α or β (Rosenthal and Yuan 2021). Since *nAChR*s are the target of the most commonly used insect control methods for adelgids (McCarty and Addesso 2019), a better understanding of *nAChR* diversity within Adelgidae and between Adelgidae and Insecta was sought. While *nAChR* composition and expression is not the only factor determining susceptibility to chemical controls, expression of different subunits and mutations within subunits can impact susceptibility to nicotinic compounds (Zhang et al. 2018; Cartereau et al. 2020).

Many of the most commonly used chemical insecticides target insect nAChRs (Millar and Denholm 2007). One such insecticide, imidacloprid, is currently the standard of care for trees infested with *A. tsugae*, but its use is complicated by its broad-spectrum action against non-target insects (Eisenback et al. 2010). A better understanding of nAChR structural diversity could potentially help guide the discovery of precision insecticides that avoid these off target effects. Adelgid nAChR sequence similarities are typically moderate, with an average of 46% across different insect origins, and with occasional higher similarity pairs (>70%) that likely indicate nAChRs of the same or similar subtypes, and some low similarity pairs as outliers (Figure 4J). The 3D structural similarity also follows a moderate to high pattern (>0.65), with scores more evenly distributed and no notable outliers (Figure 4K). nAChR ligand binding sites, distinguished by loop identity, were highly conserved across all species studied, displaying minimal variability in key binding residues (Figure 4L). Overall, predicted adelgid nAChR structures generally align well with

known human crystal structures, indicating a high level of conservation which extends across animals and indicates a vital conservation of function (PDB ID: 5FJV, Figure S1).

Conclusion: The high-quality genome assemblies presented here for *A. tsugae* and *A. abietis* reveal conservation in genome architecture, from chromosome-level synteny to shared patterns of repeat evolution and gene family dynamics. Gene family expansions in *A. abietis* included metabolic gene families, involved in carbohydrate transport, while expansions in *A. tsugae* are for genes involved in cell differentiation and development, and suppression of plant immunity. Genes associated with nicotinic acetylcholine receptors (*nAChR*s) exhibited strong conservation across species, reflecting their conserved role in neural signaling. The *Endoplasmic reticulum aminopeptidase 1* (*ERAP1*) gene family exhibited expansion, and *juvenile hormone acid methyltransferase* (*JHAMT*) displayed contraction relative to ancestral states. While nAChR and ERAP1 demonstrated conservation in protein structure across insects, JHAMT exhibited notable variability in key regions, pointing to functional divergence and an opportunity to potentially disrupt juvenile hormone signaling pathways.

Table 1. Assembly and Annotation Statistics for *A. tsugae* and *A. abietis*

Assembly	**Total Chromosomes**	**N50 (Mb)**	**Length (bp)**	**Sequence in Chromosome**	**% Gaps**	**% Repeat**
A. tsugae	8+2	21.14 Mb	221,267,597	95.49%	2.402%	17.90%
A. abietis	8+2	21.97 Mb	253,164,600	87.2%	2.884%	24.91%
BUSCO Genome, Hemiptera						
	Complete	**Single**	**Duplicate**	**Fragment**	**Missing**	***n* Searched**
A. tsugae	99.3%	98.2%	1.1%	0.4%	0.3%	2510
A. abietis	98.5%	97.2%	1.3%	0.3%	1.2%	2510
BUSCO Annotation, Hemiptera						
	Complete	**Single**	**Duplicate**	**Fragment**	**Missing**	***n* Searched**
A. tsugae	98.0%	96.7%	1.3%	0.2%	1.8%	2510
A. abietis	95.9%	94.7%	1.2%	0.2%	3.9%	2510

Figures:

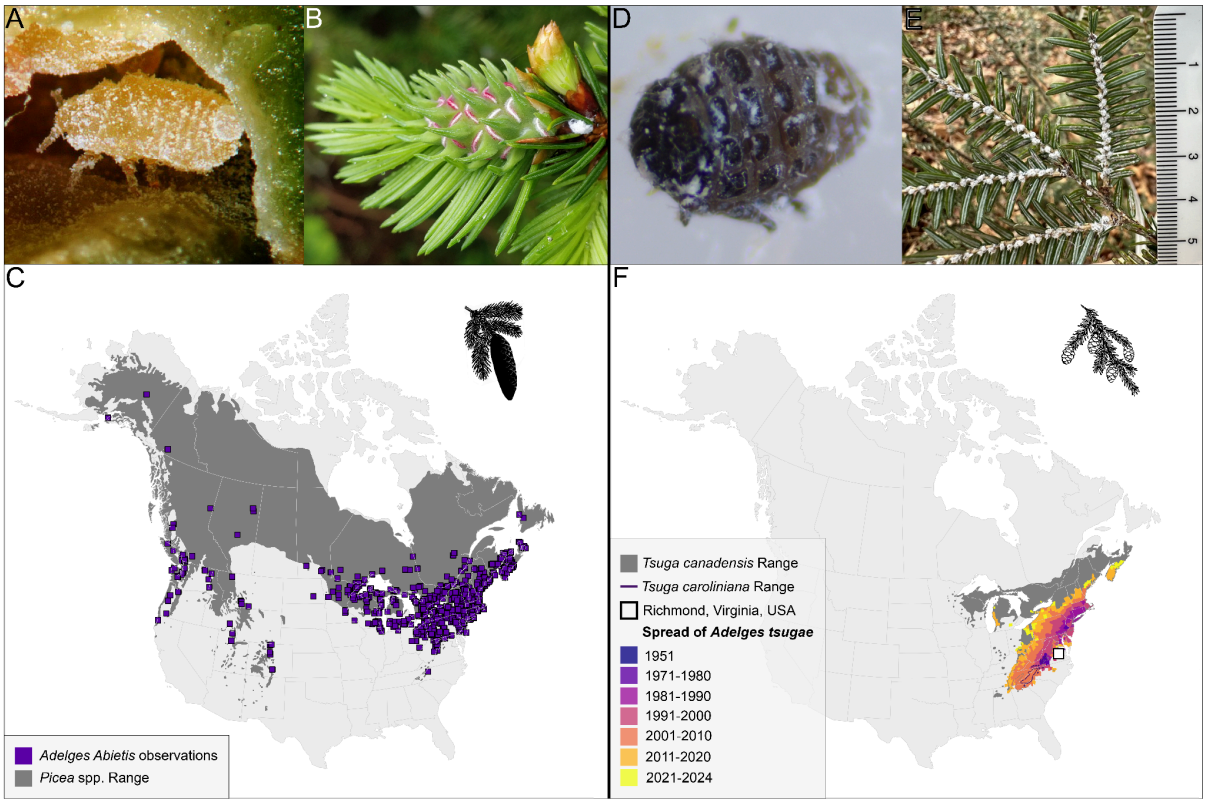


Figure 1. (A) Immature gallicolae of *A. abietis* within a gall on *Picea abies* (source: influentialpoints.com). (B) *A. abietis* gall in early development, with stem mother visible. (C) Distribution of North American *Picea* species and observations of *Adelges abietis* (source: GBIF, iNaturalist) - Inset: silhouette representing a spruce. (D) Hemlock woolly adelgid adult- Inset: silhouette representing a hemlock. (E) Eastern hemlock branch infested with *A. tsugae* (scale bar in cm). (F) Range of *Tsuga canadensis* and *Tsuga caroliniana* (E. Little and Luther 1971) and the dispersal of *Adelges tsugae* since 1951 (source: 1951-2020 Data derived from USDA FS-FHP, 2021-2024 observations compiled from infestation reports).

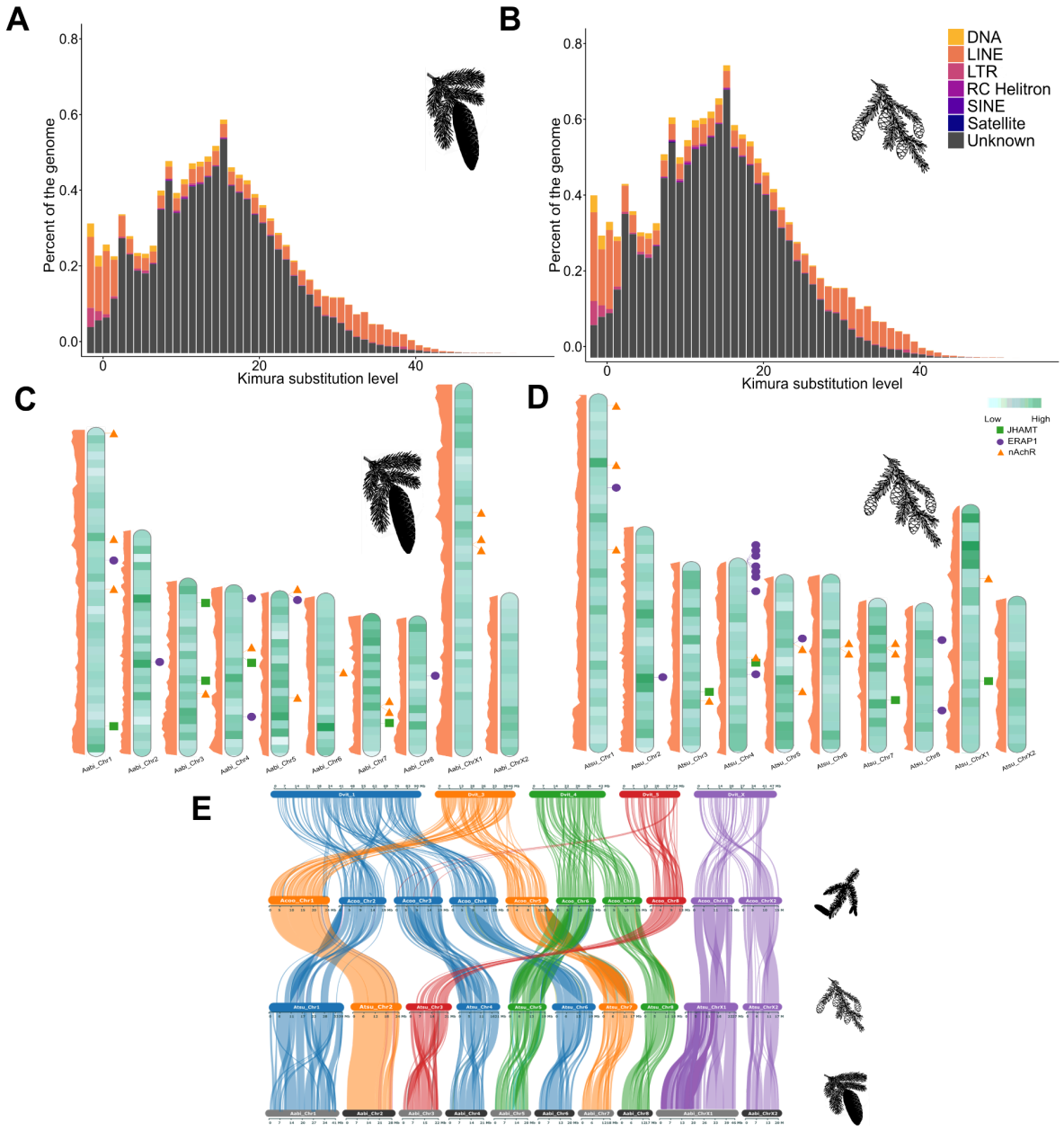
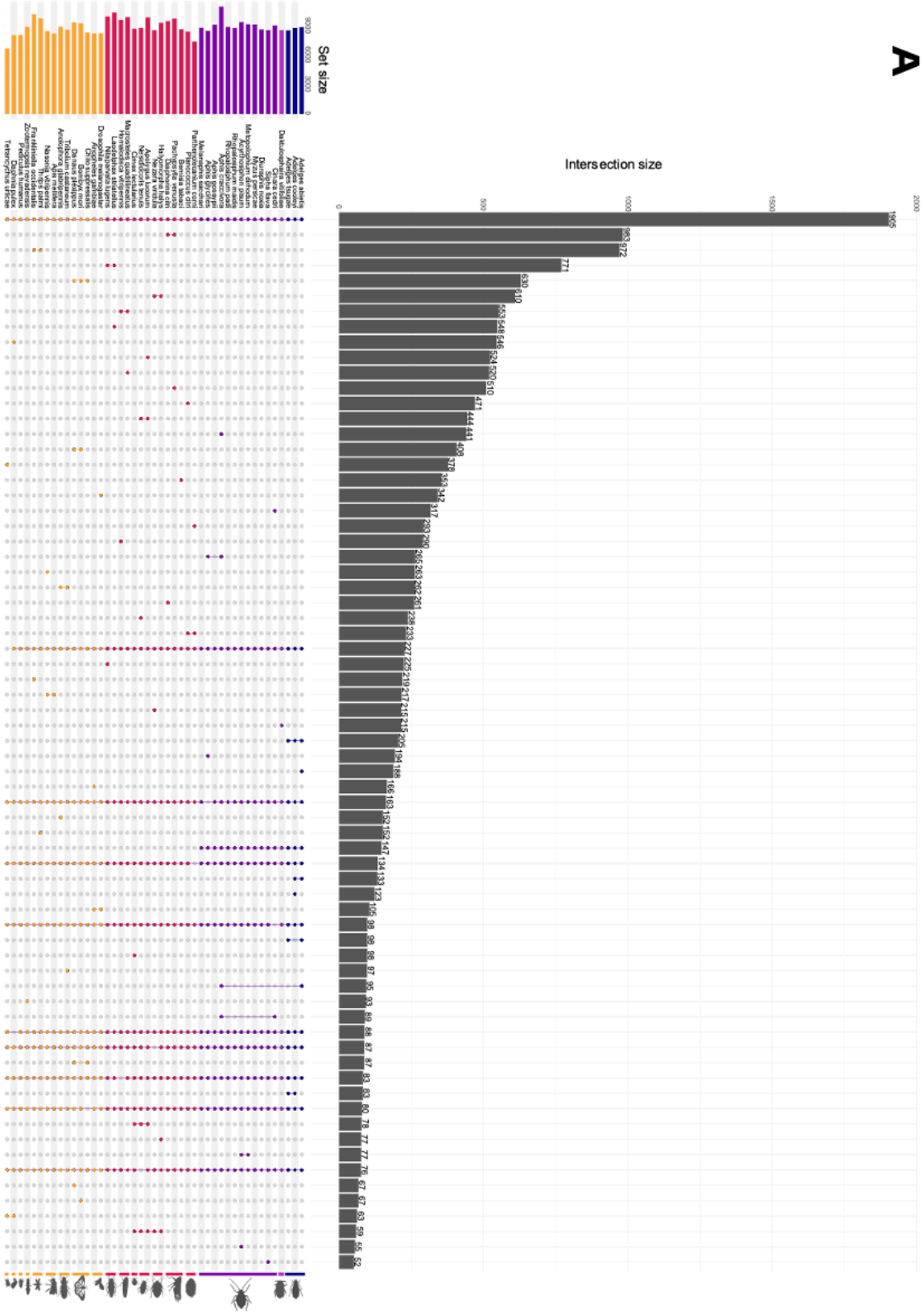


Figure 2. (A, B) Repeat landscapes of known transposable element (TE) families (DNA transposons, LINEs, and LTR retrotransposons) and unclassified repeat families in *A. tsugae* and *A. abietis*, respectively. (C&D) Chromosome models of *A. tsugae* (left) and *A. abietis* (right) with gene density from light to dark shown in green (*A. tsugae* 0-56 genes per 500,000 bp, *A. abietis* 1-49). The focal gene family members, including: *JHAMT* (green boxes), *ERAP* (purple circle), and *nAChR* (orange triangle) are shown in their respective positions on the far right. The repeat density track is on the left side in orange. (E) Synteny analysis between *Daktulosphaira vitifoliae*, *Adelges cooleyi* (longest ten scaffolds), and the ten pseudo-chromosomes of *A. abietis* and *A. tsugae*. Adelgids exhibit strong synteny in both sex chromosomes and autosomes.


A
Intersection size
Set size
Addipse alberta
Acropolis cooeyi
Acropolis tabae
Dactylosphaeria villosae
Chaite ecoli
Sipha maydis
Myzus persicae
Dysaphis nosa
Metopolophium dirhodum
Acyrthosiphon pisum
Rhopalosiphum padi
Aphis craccivora
Aphis gossypii
Aphis glycines
Melanaphis sacchari
Pentatrichopsis corni
Pentatrichopsis sp1
Brevicoryne brassicae
Pachypsylla venusta
Scolypanema cibi
Hylyncorpha halys
Kalotermes flavicollis
Apolygus lucorum
Cimex lectularius
Cimex hemiptera
Macrosteles quadripunctatus
Thyanta pallidovirens
Leptoglossus stigmatus
Arphnites gallicae
Drosophila melanogaster
Drosophila simulans
Bactrocera dorsalis
Bactrocera minor
Ceratitis capitata
Tribolium castaneum
Anoplophora glabripennis
Nasonia vitripennis
Vespula vulgaris
Fox silinella occidentalis
Zoraptera pedicularius
Dermatophagoides farinae
Daphnia pulex
Tetranychus urticae
1995
963
972
771
630
610
555
548
546
524
520
510
471
444
441
409
378
358
342
317
293
290
265
263
262
261
238
233
227
225
219
217
215
215
205
194
188
166
163
162
152
147
134
133
123
105
98
98
96
97
95
93
90
88
87
87
83
83
80
78
77
77
76
67
67
63
59
55
52

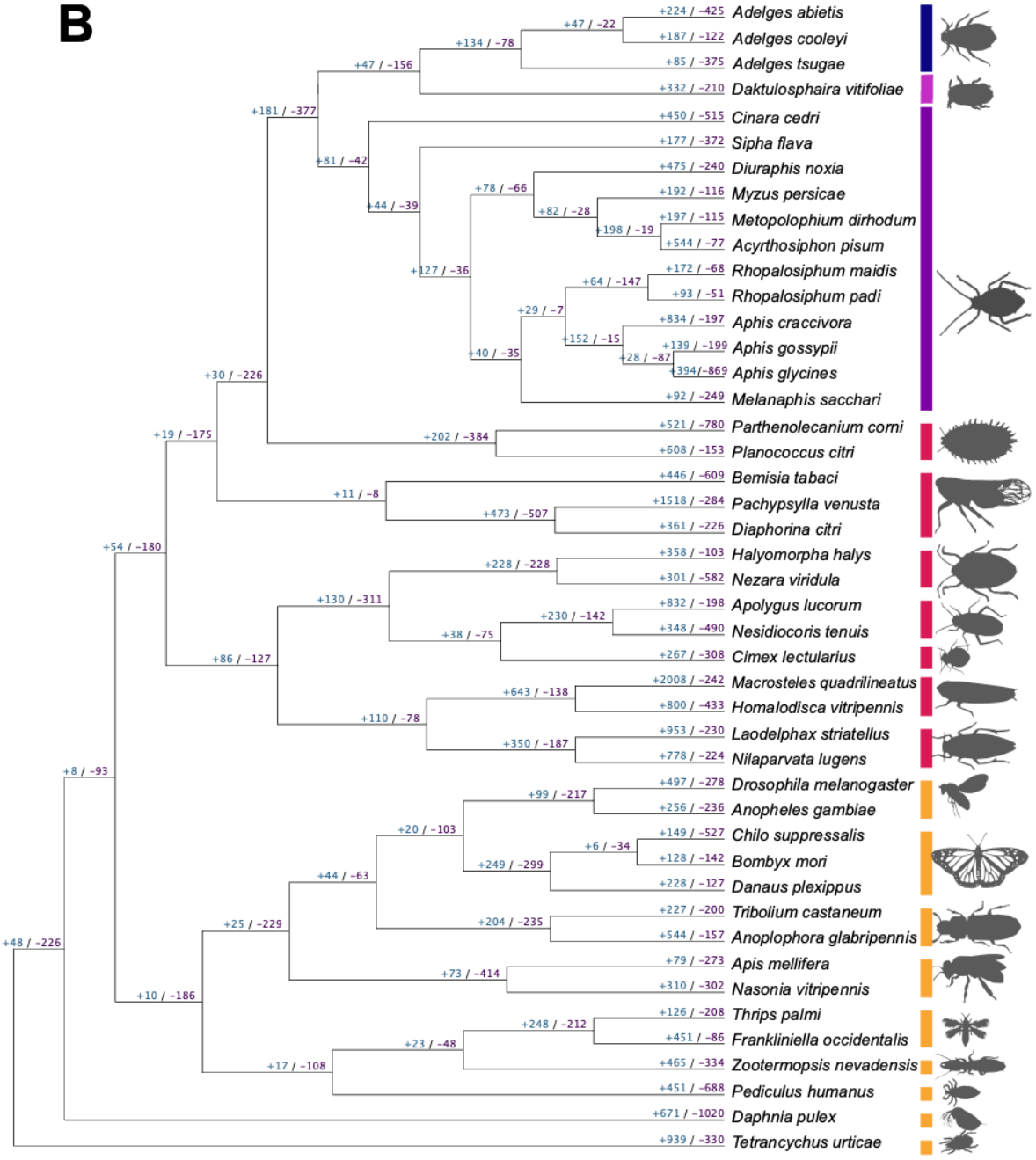

B
+224 / -425 Adelges abietis
+187 / -122 Adelges cooleyi
+85 / -375 Adelges tsugae
+332 / -210 Daktulosphaira vitifoliae
+450 / -515 Cinara cedri
+177 / -372 Sipha flava
+475 / -240 Diuraphis noxia
+192 / -116 Myzus persicae
+197 / -115 Metopolophium dirhodum
+544 / -77 Acyrthosiphon pisum
+172 / -68 Rhopalosiphum maidis
+93 / -51 Rhopalosiphum padi
+834 / -197 Aphis craccivora
+139 / -199 Aphis gossypii
+394/-869 Aphis glycines
+92 / -249 Melanaphis sacchari
+521 / -780 Parthenolecanium corni
+608 / -153 Planococcus citri
+446 / -609 Bemisia tabaci
+1518 / -284 Pachypsylla venusta
+361 / -226 Diaphorina citri
+358 / -103 Halyomorpha halys
+301 / -582 Nezara viridula
+832 / -198 Apolygus lucorum
+348 / -490 Nesidiocoris tenuis
+267 / -308 Cimex lectularius
+2008 / -242 Macrosteles quadrilineatus
+800 / -433 Homalodisca vitripennis
+953 / -230 Laodelphax striatellus
+778 / -224 Nilaparvata lugens
+497 / -278 Drosophila melanogaster
+256 / -236 Anopheles gambiae
+149 / -527 Chilo suppressalis
+128 / -142 Bombyx mori
+228 / -127 Danaus plexippus
+227 / -200 Tribolium castaneum
+544 / -157 Anoplophora glabripennis
+79 / -273 Apis mellifera
+310 / -302 Nasonia vitripennis
+126 / -208 Thrips palmi
+451 / -86 Frankliniella occidentalis
+465 / -334 Zootermopsis nevadensis
+451 / -688 Pediculus humanus
+671 / -1020 Daphnia pulex
+939 / -330 Tetrancychus urticae
+47 / -22
+134 / -78
+47 / -156
+181 / -377
+81 / -42
+44 / -39
+78 / -66
+82 / -28
+198 / -19
+127 / -36
+64 / -147
+29 / -7
+152 / -15
+28 / -87
+40 / -35
+30 / -226
+19 / -175
+202 / -384
+11 / -8
+473 / -507
+54 / -180
+228 / -228
+130 / -311
+230 / -142
+38 / -75
+86 / -127
+643 / -138
+110 / -78
+350 / -187
+8 / -93
+99 / -217
+20 / -103
+6 / -34
+249 / -299
+44 / -63
+25 / -229
+204 / -235
+48 / -226
+73 / -414
+10 / -186
+248 / -212
+23 / -48
+17 / -108

Figure 3. (A) Orthologous gene group interactions between 45 species. Minimum interaction size displayed is 50. Rightmost vertical bars represent lineages, and are grouped from top to bottom and darkest to lightest: Adelgidae, Phylloxeridae, Aphididae, other Hemiptera, and non-Hemipteran arthropods. (B) Phylogenetic tree generated from the gene family analysis with the number of significantly expanding and contracting genes indicated at nodes. Silhouette images from bugwood.org.

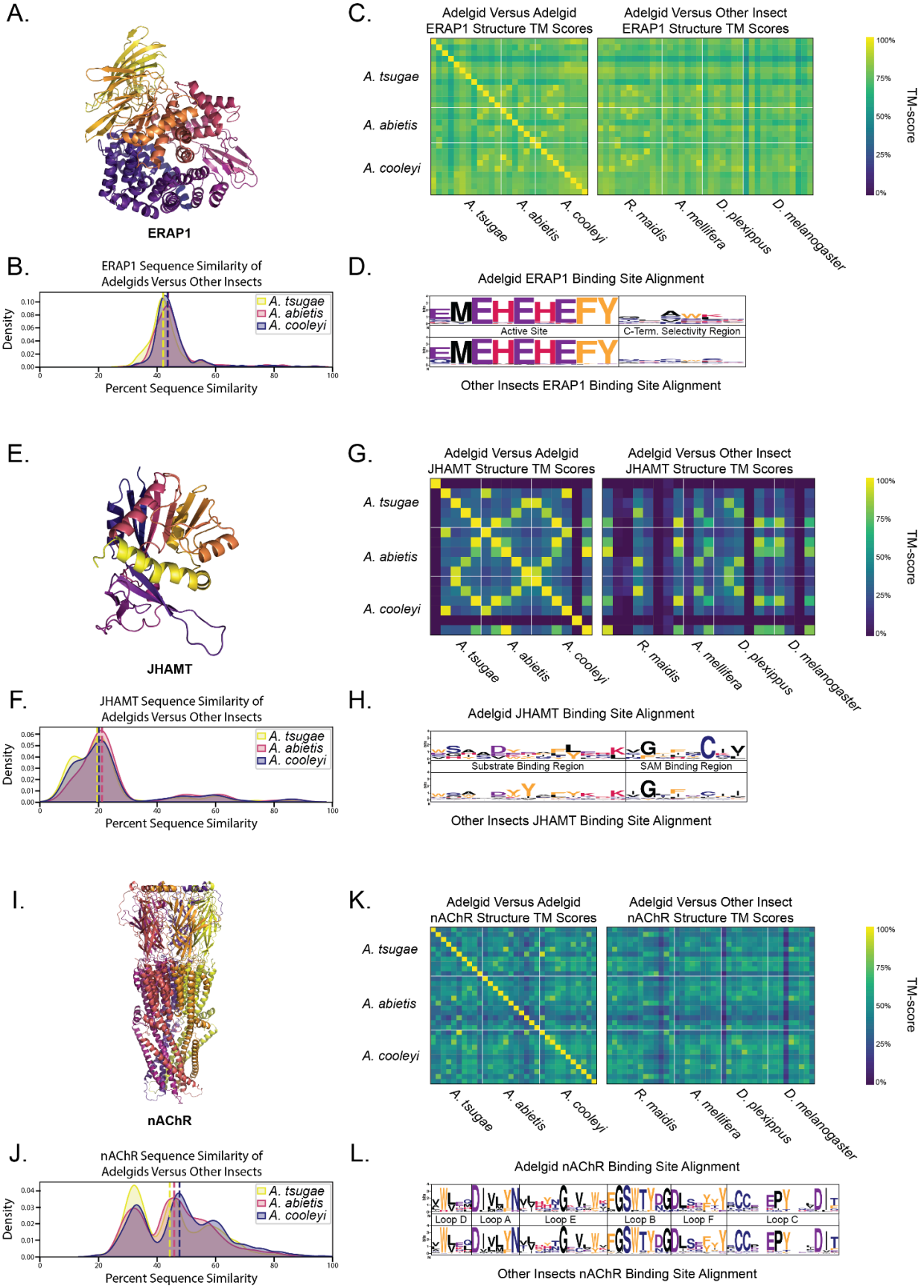

A.
ERAP1
B.
ERAP1 Sequence Similarity of Adelgids Versus Other Insects
Density
Percent Sequence Similarity
A. tsugae
A. abietis
A. cooleyi
C.
Adelgid Versus Adelgid ERAP1 Structure TM Scores
Adelgid Versus Other Insect ERAP1 Structure TM Scores
A. tsugae
A. abietis
A. cooleyi
R. maidis
A. mellifera
D. plexippus
D. melanogaster
TM-score
D.
Adelgid ERAP1 Binding Site Alignment
Active Site
C-Term. Selectivity Region
Other Insects ERAP1 Binding Site Alignment
E.
JHAMT
F.
JHAMT Sequence Similarity of Adelgids Versus Other Insects
Density
Percent Sequence Similarity
A. tsugae
A. abietis
A. cooleyi
G.
Adelgid Versus Adelgid JHAMT Structure TM Scores
Adelgid Versus Other Insect JHAMT Structure TM Scores
A. tsugae
A. abietis
A. cooleyi
R. maidis
A. mellifera
D. plexippus
D. melanogaster
TM-score
H.
Adelgid JHAMT Binding Site Alignment
Substrate Binding Region
SAM Binding Region
Other Insects JHAMT Binding Site Alignment
I.
nAChR
J.
nAChR Sequence Similarity of Adelgids Versus Other Insects
Density
Percent Sequence Similarity
A. tsugae
A. abietis
A. cooleyi
K.
Adelgid Versus Adelgid nAChR Structure TM Scores
Adelgid Versus Other Insect nAChR Structure TM Scores
A. tsugae
A. abietis
A. cooleyi
R. maidis
A. mellifera
D. plexippus
D. melanogaster
TM-score
L.
Adelgid nAChR Binding Site Alignment
Loop D
Loop A
Loop E
Loop B
Loop F
Loop C
Other Insects nAChR Binding Site Alignment

Figure 4. (A) AlphaFold3 structure of an *A. tsugae* ERAP1. (B) Kernel density estimate (KDE) plot of *A. tsugae, A. abietis,* and *A. cooleyi* percent sequence similarity compared to other insect (*R. maidis*, *A. mellifera*, *D. plexippus*, and *D. melanogaster*) for ERAP1 proteins. (C) Structure similarity heatmap comparing adelgid and non-adelgid ERAP1 proteins. The number of proteins per heat map varies with the copy number. (D) Logo plot of the consensus sequences of ERAP1 ligand binding sites, labeled by binding loop identity. (E) AlphaFold3 structure of an *A. tsugae* JHAMT. (F) KDE plot of *A. tsugae*, *A. abietis*, and *A. cooleyi* percent sequence similarity compared to other insect JHAMT proteins. (G) Structure similarity heatmap comparing adelgid and non-adelgid JHAMT proteins. (H) Logo plot of the consensus sequences of JHAMT ligand binding sites, separated into substrate binding site and S-Adenosyl methionine (SAM) binding site. (I) AlphaFold3 structure of an *A. tsugae* nAChR. (J) KDE plot of *A. tsugae*, *A. abietis*, and *A. cooleyi* percent sequence similarity compared to other insect nAChR proteins. (K) Structure similarity heatmap comparing adelgid and non-adelgid nAChR proteins. (L) Logo plot of the consensus sequences of nAChR ligand binding sites, separated into active site residues and C-terminal selectivity region residues.

Acknowledgments
The authors acknowledge the contributions of the HPC resources from the Computational Biology Core and support for sequencing through the Center for Genome Innovation, both within the Institute for Systems Genomics at the University of Connecticut. We would also like to thank Dr. Gaelen Burke and Dr. Dustin Dial at the University of Georgia for providing laboratory space as well as Dr. Ben Smith at the Mountain Research Station in North Carolina. Dr. Nathan Havill of the USDA Forest Service and Dr. Bryan Brunet of Agriculture and Agri-Food Canada are thanked for providing preliminary nAChR sequences.

Conflict of Interest
No conflicts of interest

Funder Information
A.M.G. is a RaMP (Research and Mentoring for Postbaccalaureates) fellow at the University of Connecticut supported with an award from the National Science Foundation (DBI-2217100 to J.L.W). K.C.F received support from the USDA Postdoctoral Fellowship Program (2023-67012-40000). K.C.F and V.S.V are partially supported through the Trees In Peril award to J.L.W. from The Nature Conservancy. V.S.V. was additionally supported by NSF DBI-1943371 awarded to J.L.W. T.A.C gratefully acknowledges support from the Alfred P. Sloan Foundation. C.S. was supported by the NIH T32 training program (2T32GM132046-06). SAH and TCM were funded from the UK Research and Innovation (UKRI) Biotechnology and Biological Sciences Research Council (BBSRC) grants (BB/V008544/1 and BB/R009481/1). Additional Support is provided by the BBSRC Institute Strategy Programmes (BBS/E/J/000PR9797 and BBS/E/JI/230001B) awarded to the John Innes Centre (JIC). The JIC is grant-aided by the John Innes Foundation.

Data availability: All scripts and data are described in https://gitlab.com/PlantGenomicsLab/adelges-tsugae-genomics. NCBI BioProject ID PRJNA1163707 contains the raw sequencing reads for *A. tsugae* and *A. abietis*. RNA reads (SRR30936311 ONT long reads for *A. tsugae*, SRR31285667 HiC reads for *A. abietis*, SRR31285820-3 10x linked reads for *A. abietis*, SRR30936309 and SRR30936310 short RNA reads for *A. tsugae*), as well as the whole genome assemblies SAMN44664291 (*A. tsugae*), and SAMN44664292 (*A. abietis*). The structural and functional annotations are also hosted in the Gitlab.

Reviewer link:
https://dataview.ncbi.nlm.nih.gov/object/PRJNA1163707?reviewer=34uchqgm147jpb9vticjicj84r

Literature Cited:

Abramson, Josh, Jonas Adler, Jack Dunger, Richard Evans, Tim Green, Alexander Pritzel, Olaf Ronneberger, et al. 2024. "Accurate Structure Prediction of Biomolecular Interactions with AlphaFold 3." *Nature* 630 (8016): 493–500.

Alonge, Michael, Ludivine Lebeigle, Melanie Kirsche, Katie Jenike, Shujun Ou, Sergey Aganezov, Xingang Wang, Zachary B. Lippman, Michael C. Schatz, and Sebastian Soyk. 2022. "Automated Assembly Scaffolding Using RagTag Elevates a New Tomato System for High-Throughput Genome Editing." *Genome Biology* 23 (1): 258.

Astashyn, Alexander, Eric S. Tvedte, Deacon Sweeney, Victor Sapojnikov, Nathan Bouk, Victor Joukov, Eyal Mozes, et al. 2024. "Rapid and Sensitive Detection of Genome Contamination at Scale with FCS-GX." *Genome Biology* 25 (1): 60.

Bailey, Emma, Linda Field, Christopher Rawlings, Rob King, Fady Mohareb, Keywan-Hassani Pak, David Hughes, et al. 2022. "A Scaffold-Level Genome Assembly of a Minute Pirate Bug, Orius Laevigatus (Hemiptera: Anthocoridae), and a Comparative Analysis of Insecticide Resistance-Related Gene Families with Hemipteran Crop Pests." *BMC Genomics* 23 (1): 45.

Bandi, Venkat, and Carl Gutwin. 2020. "Interactive Exploration of Genomic Conservation." In *Proceedings of the 46th Graphics Interface Conference on Proceedings of Graphics Interface 2020 (GI'20)*. Waterloo, CAN: Canadian Human-Computer Communications Society.

Baril, Tobias, Adam Pym, Chris Bass, and Alex Hayward. 2023. "Transposon Accumulation at Xenobiotic Gene Family Loci in Aphids." *Genome Research* 33 (10): 1718–33.

Barrantes, Francisco J. 2023. "Structure and Function Meet at the Nicotinic Acetylcholine Receptor-Lipid Interface." *Pharmacological Research: The Official Journal of the Italian Pharmacological Society* 190 (April):106729.

Boulain, Hélène, Fabrice Legeai, Endrick Guy, Stéphanie Morlière, Nadine E. Douglas, Jonghee Oh, Marimuthu Murugan, et al. 2018. "Fast Evolution and Lineage-Specific Gene Family Expansions of Aphid Salivary Effectors Driven by Interactions with Host-Plants." *Genome Biology and Evolution* 10 (6): 1554–72.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L. Madden. 2009. "BLAST+: Architecture and Applications." *BMC Bioinformatics* 10 (1): 421.

Carter, C. 1971. "Conifer Woolly Aphids (Adelgidae) in Britain." https://doi.org/10.5555/19716605345.

Cembrowski, Mark S., Matthew G. Phillips, Salvatore F. DiLisio, Brenda C. Shields, Johan Winnubst, Jayaram Chandrashekar, Erhan Bas, and Nelson Spruston. 2018. "Dissociable Structural and Functional Hippocampal Outputs via Distinct Subiculum Cell Classes." *Cell* 173 (5): 1280–92.e18.

Chakrabarti, S. 2018. "Aphids." In *Pests and Their Management*, 871–908. Singapore: Springer Singapore.

Chaytow, Helena, Yu-Ting Huang, Thomas H. Gillingwater, and Kiterie M. E. Faller. 2018. "The Role of Survival Motor Neuron Protein (SMN) in Protein Homeostasis." *Cellular and Molecular Life Sciences: CMLS* 75 (21): 3877–94.

Chen, Chun-Yu, Yao-Qian Liu, Wei-Meng Song, Dian-Yang Chen, Fang-Yan Chen,

Xue-Ying Chen, Zhi-Wen Chen, et al. 2019. “An Effector from Cotton Bollworm Oral Secretion Impairs Host Plant Defense Signaling.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (28): 14331–38.

Cook, Stephen P., Frank Merickel, Karen Ward, and Nathan Havill. 2012. “First Report of Adelges Abietis (Linneaus) (Hemiptera: Adelgidae) in Idaho.” *The Pan-Pacific Entomologist* 88 (4): 435–36.

Crooks, Gavin E., Gary Hon, John-Marc Chandonia, and Steven E. Brenner. 2004. “WebLogo: A Sequence Logo Generator.” *Genome Research* 14 (6): 1188–90.

Dancewicz, Katarzyna, Beata Gabryś, Iwona Morkunas, and Sławomir Samardakiewicz. 2021. “Probing Behavior of Adelges Laricis Vallot (Hemiptera: Adelgidae) on Larix Decidua Mill: Description and Analysis of EPG Waveforms.” *PloS One* 16 (5): e0251663.

De Coster, Wouter, and Rosa Rademakers. 2023. “NanoPack2: Population-Scale Evaluation of Long-Read Sequencing Data.” *Bioinformatics* 39 (5). https://doi.org/10.1093/bioinformatics/btad311.

Dial, Dustin T., Kathryn M. Weglarz, Bryan M. T. Brunet, Nathan P. Havill, Carol D. von Dohlen, and Gaelen R. Burke. 2023. “Whole-Genome Sequence of the Cooley Spruce Gall Adelgid, Adelges Cooleyi (Hemiptera: Sternorrhyncha: Adelgidae).” *G3 (Bethesda, Md.)* 14 (1). https://doi.org/10.1093/g3journal/jkad224.

Dohlen, Carol D. V. O. N., and Nancy A. Moran. 2000. “Molecular Data Support a Rapid Radiation of Aphids in the Cretaceous and Multiple Origins of Host Alternation.” *Biological Journal of the Linnean Society. Linnean Society of London* 71 (4): 689–717.

Dohlen, Carol D. von, Usha Spaulding, Kathleen Shields, Nathan P. Havill, Cristina Rosa, and Kelli Hoover. 2013. “Diversity of Proteobacterial Endosymbionts in Hemlock Woolly Adelgid (Adelges Tsugae) (Hemiptera: Adelgidae) from Its Native and Introduced Range: Proteobacterial Endosymbionts of Hemlock Woolly Adelgid.” *Environmental Microbiology* 15 (7): 2043–62.

Dudchenko, Olga, Sanjit S. Batra, Arina D. Omer, Sarah K. Nyquist, Marie Hoeger, Neva C. Durand, Muhammad S. Shamim, et al. 2017. “De Novo Assembly of the Aedes Aegypti Genome Using Hi-C Yields Chromosome-Length Scaffolds.” *Science (New York, N.Y.)* 356 (6333): 92–95.

Durand, Neva C., James T. Robinson, Muhammad S. Shamim, Ido Machol, Jill P. Mesirov, Eric S. Lander, and Erez Lieberman Aiden. 2016. “Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom.” *Cell Systems* 3 (1): 99–101.

Durand, Neva C., Muhammad S. Shamim, Ido Machol, Suhas S. P. Rao, Miriam H. Huntley, Eric S. Lander, and Erez Lieberman Aiden. 2016. “Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments.” *Cell Systems* 3 (1): 95–98.

Edgar, Robert C. 2004. “MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput.” *Nucleic Acids Research* 32 (5): 1792–97.

Eisenback, Brian M., Scott M. Salom, Loke T. Kok, and Anthony F. Lagalante. 2010. “Lethal and Sublethal Effects of Imidacloprid on Hemlock Woolly Adelgid (Hemiptera: Adelgidae) and Two Introduced Predator Species.” *Journal of Economic Entomology* 103 (4): 1222–34.

Ellison, Aaron M., David A. Orwig, Matthew C. Fitzpatrick, and Evan L. Preisser.

2018. “The Past, Present, and Future of the Hemlock Woolly Adelgid (Adelges Tsugae) and Its Ecological Interactions with Eastern Hemlock (Tsuga Canadensis) Forests.” *Insects* 9 (4). https://doi.org/10.3390/insects9040172.

Emms, David M., and Steven Kelly. 2019. “OrthoFinder: Phylogenetic Orthology Inference for Comparative Genomics.” *Genome Biology* 20 (1): 238.

Evnouchidou, Irini, Mirjana Weimershaus, Loredana Saveanu, and Peter van Endert. 2014. “ERAP1-ERAP2 Dimerization Increases Peptide-Trimming Efficiency.” *The Journal of Immunology* 193 (2): 901–8.

Favret, Colin, Nathan P. Havill, Gary L. Miller, Masakazu Sano, and Benjamin Victor. 2015. “Catalog of the Adelgids of the World (Hemiptera, Adelgidae).” *ZooKeys* 534 (534): 35–54.

Fei, Songlin, Randall S. Morin, Christopher M. Oswalt, and Andrew M. Liebhold. 2019. “Biomass Losses Resulting from Insect and Disease Invasions in US Forests.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (35): 17371–76.

Finn, Robert D., Alex Bateman, Jody Clements, Penelope Coggill, Ruth Y. Eberhardt, Sean R. Eddy, Andreas Heger, et al. 2014. “Pfam: The Protein Families Database.” *Nucleic Acids Research* 42 (Database issue): D222–30.

Flynn, Jullien M., Robert Hubley, Clément Goubert, Jeb Rosen, Andrew G. Clark, Cédric Feschotte, and Arian F. Smit. 2020. “RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families.” *Proceedings of the National Academy of Sciences of the United States of America* 117 (17): 9451–57.

Gharpure, Anant, Colleen M. Noviello, and Ryan E. Hibbs. 2020. “Progress in Nicotinic Receptor Structural Biology.” *Neuropharmacology* 171 (108086): 108086.

Giastas, Petros, Anastasia Mpakali, Athanasios Papakyriakou, Aggelos Lelis, Paraskevi Kokkala, Margarete Neu, Paul Rowland, John Liddle, Dimitris Georgiadis, and Efstratios Stratikos. 2019. “Mechanism for Antigenic Peptide Selection by Endoplasmic Reticulum Aminopeptidase 1.” *Proceedings of the National Academy of Sciences of the United States of America* 116 (52): 26709–16.

Goldfarb, Tamara, Vamsi K. Kodali, Shashikant Pujar, Vyacheslav Brover, Barbara Robbertse, Catherine M. Farrell, Dong-Ha Oh, et al. 2025. “NCBI RefSeq: Reference Sequence Standards through 25 Years of Curation and Annotation.” *Nucleic Acids Research* 53 (D1): D243–57.

Gouveia-Oliveira, Rodrigo, Peter W. Sackett, and Anders G. Pedersen. 2007. “MaxAlign: Maximizing Usable Data in an Alignment.” *BMC Bioinformatics* 8 (1): 312.

Grice, Stuart J., and Ji-Long Liu. 2011. “Survival Motor Neuron Protein Regulates Stem Cell Division, Proliferation, and Differentiation in Drosophila.” *PLoS Genetics* 7 (4): e1002030.

Guo, Pengchao, Yunshi Zhang, Li Zhang, Haiyang Xu, Huan Zhang, Zhan Wang, Yongliang Jiang, David Molloy, Ping Zhao, and Qingyou Xia. 2021. “Structural Basis for Juvenile Hormone Biosynthesis by the Juvenile Hormone Acid Methyltransferase.” *The Journal of Biological Chemistry* 297 (5): 101234.

Gurevich, Alexey, Vladislav Saveliev, Nikolay Vyahhi, and Glenn Tesler. 2013. “QUAST: Quality Assessment Tool for Genome Assemblies.” *Bioinformatics* 29 (8): 1072–75.

Hart, Alexander J., Samuel Ginzburg, Muyang Sam Xu, Cera R. Fisher, Nasim Rahmatpour, Jeffry B. Mitton, Robin Paul, and Jill L. Wegrzyn. 2020. “EnTAP: Bringing Faster and Smarter Functional Annotation to Non-Model Eukaryotic Transcriptomes.” *Molecular Ecology Resources* 20 (2): 591–604.

Havill, Nathan P., and Robert G. Foottit. 2007. “Biology and Evolution of Adelgidae.” *Annual Review of Entomology* 52:325–49.

Havill, Nathan P., Michael E. Montgomery, Guoyue Yu, Shigehiko Shiyake, and Adalgisa Caccone. 2006a. “Mitochondrial DNA from Hemlock Woolly Adelgid (Hemiptera: Adelgidae) Suggests Cryptic Speciation and Pinpoints the Source of the Introduction to Eastern North America.” *Annals of the Entomological Society of America* 99 (2): 195–203.

———. 2006b. “Mitochondrial DNA from Hemlock Woolly Adelgid (Hemiptera: Adelgidae) Suggests Cryptic Speciation and Pinpoints the Source of the Introduction to Eastern North America.” *Annals of the Entomological Society of America* 99 (2): 195–203.

Huang, Chen, Bingru Ji, Zhaohui Shi, Jiangyue Wang, Jiaqing Yuan, Peng Yang, Xiao Xu, et al. 2025. “A Comparative Genomic Analysis at the Chromosomal-Level Reveals Evolutionary Patterns of Aphid Chromosomes.” *Communications Biology* 8 (1): 427.

Huerta-Cepas, Jaime, Damian Szklarczyk, Davide Heller, Ana Hernández-Plaza, Sofia K. Forslund, Helen Cook, Daniel R. Mende, et al. 2019. “eggNOG 5.0: A Hierarchical, Functionally and Phylogenetically Annotated Orthology Resource Based on 5090 Organisms and 2502 Viruses.” *Nucleic Acids Research* 47 (D1): D309–14.

Jones, Philip, David Binns, Hsin-Yu Chang, Matthew Fraser, Weizhong Li, Craig McAnulla, Hamish McWilliam, et al. 2014. “InterProScan 5: Genome-Scale Protein Function Classification.” *Bioinformatics* 30 (9): 1236–40.

Kanaseki, Takayuki, Nicolas Blanchard, Gianna Elena Hammer, Federico Gonzalez, and Nilabh Shastri. 2006. “ERAAP Synergizes with MHC Class I Molecules to Make the Final Cut in the Antigenic Peptide Precursors in the Endoplasmic Reticulum.” *Immunity* 25 (5): 795–806.

Katoh, Kazutaka, and Hiroyuki Toh. 2008. “Recent Developments in the MAFFT Multiple Sequence Alignment Program.” *Briefings in Bioinformatics* 9 (4): 286–98.

Kim, Daehwan, Li Song, Florian P. Breitwieser, and Steven L. Salzberg. 2016. “Centrifuge: Rapid and Sensitive Classification of Metagenomic Sequences.” *Genome Research* 26 (12): 1721–29.

Kochan, Grazyna, Tobias Krojer, David Harvey, Roman Fischer, Liye Chen, Melanie Vollmar, Frank von Delft, et al. 2011. “Crystal Structures of the Endoplasmic Reticulum Aminopeptidase-1 (ERAP1) Reveal the Molecular Basis for N-Terminal Peptide Trimming.” *Proceedings of the National Academy of Sciences of the United States of America* 108 (19): 7745–50.

Kolmogorov, Mikhail, Jeffrey Yuan, Yu Lin, and Pavel A. Pevzner. 2019. “Assembly of Long, Error-Prone Reads Using Repeat Graphs.” *Nature Biotechnology* 37 (5): 540–46.

Koren, Sergey, Brian P. Walenz, Konstantin Berlin, Jason R. Miller, Nicholas H. Bergman, and Adam M. Phillippy. 2017. “Canu: Scalable and Accurate Long-Read Assembly via Adaptive K-Mer Weighting and Repeat Separation.” *Genome Research* 27 (5): 722–36.

Kumar, Sudhir, Glen Stecher, Michael Li, Christina Knyaz, and Koichiro Tamura. 2018. “MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms.” *Molecular Biology and Evolution* 35 (6): 1547–49.

Kumar, Sujai, Martin Jones, Georgios Koutsovoulos, Michael Clarke, and Mark Blaxter. 2013. “Blobology: Exploring Raw Genome Data for Contaminants, Symbionts and Parasites Using Taxon-Annotated GC-Coverage Plots.” *Frontiers in Genetics* 4 (November):237.

Kuznetsov, Dmitry, Fredrik Tegenfeldt, Mosè Manni, Mathieu Seppey, Matthew Berkeley, Evgenia V. Kriventseva, and Evgeny M. Zdobnov. 2023. “OrthoDB v11: Annotation of Orthologs in the Widest Sampling of Organismal Diversity.” *Nucleic Acids Research* 51 (D1): D445–51.

Laetsch, Dominik R., and Mark L. Blaxter. 2017. “BlobTools: Interrogation of Genome Assemblies.” *F1000Research* 6 (1287): 1287.

Little, Elbert, and Leslie A. Luther. 1971. “Atlas of United States Trees:(no. 1146). Conifers and Important Hardwoods.” Edited by E. L. Little. *Jr* 1 (1146).

Li, Yiyuan, Bo Zhang, and Nancy A. Moran. 2020. “The Aphid X Chromosome Is a Dangerous Place for Functionally Important Genes: Diverse Evolution of Hemipteran Genomes Based on Chromosome-Level Assemblies.” *Molecular Biology and Evolution* 37 (8): 2357–68.

Li, Zheng, Allen Z. Xue, Gerald P. Maeda, Yiyuan Li, Paul D. Nabity, and Nancy A. Moran. 2023. “Phylloxera and Aphids Show Distinct Features of Genome Evolution despite Similar Reproductive Modes.” *Molecular Biology and Evolution* 40 (12). https://doi.org/10.1093/molbev/msad271.

Maben, Zachary, Richa Arya, Dimitris Georgiadis, Efstratios Stratikos, and Lawrence J. Stern. 2021. “Conformational Dynamics Linked to Domain Closure and Substrate Binding Explain the ERAP1 Allosteric Regulation Mechanism.” *Nature Communications* 12 (1): 5302.

Manni, Mosè, Matthew R. Berkeley, Mathieu Seppey, and Evgeny M. Zdobnov. 2021. “BUSCO: Assessing Genomic Data Quality and Beyond.” *Current Protocols* 1 (12): e323.

Mao, Meng, Tyler J. Simmonds, Corinne M. Stouthamer, Tara M. Kehoe, Scott M. Geib, and Gaelen R. Burke. 2023. “A Chromosome Scale Assembly of the Parasitoid Wasp Venturia Canescens Provides Insight into the Process of Virus Domestication.” *G3 (Bethesda, Md.)* 13 (10): jkad137.

Mapleson, Daniel, Gonzalo Garcia Accinelli, George Kettleborough, Jonathan Wright, and Bernardo J. Clavijo. 2017. “KAT: A K-Mer Analysis Toolkit to Quality Control NGS Datasets and Genome Assemblies.” *Bioinformatics (Oxford, England)* 33 (4): 574–76.

Marchler-Bauer, Aron, Myra K. Derbyshire, Noreen R. Gonzales, Shennan Lu, Farideh Chitsaz, Lewis Y. Geer, Renata C. Geer, et al. 2015. “CDD: NCBI’s Conserved Domain Database.” *Nucleic Acids Research* 43 (Database issue): D222–26.

Marcovich, Irina, Marcelo J. Moglie, Agustín E. Carpaneto Freixas, Anabella P. Trigila, Lucia F. Franchini, Paola V. Plazas, Marcela Lipovsek, and Ana Belén Elgoyhen. 2020. “Distinct Evolutionary Trajectories of Neuronal and Hair Cell Nicotinic Acetylcholine Receptors.” *Molecular Biology and Evolution* 37 (4): 1070–89.

Mathers, Thomas C., Yazhou Chen, Gemy Kaithakottil, Fabrice Legeai, Sam T. Mugford, Patrice Baa-Puyoulet, Anthony Bretaudeau, et al. 2017. “Rapid

Transcriptional Plasticity of Duplicated Gene Clusters Enables a Clonally Reproducing Aphid to Colonise Diverse Plant Species." *Genome Biology* 18 (1): 27.
Mathers, Thomas C., Roland H. M. Wouters, Sam T. Mugford, David Swarbreck, Cock van Oosterhout, and Saskia A. Hogenhout. 2021. "Chromosome-Scale Genome Assemblies of Aphids Reveal Extensively Rearranged Autosomes and Long-Term Conservation of the X Chromosome." *Molecular Biology and Evolution* 38 (3): 856–75.
Ma, Weihua, Le Xu, Hongxia Hua, Mengyao Chen, Mengjian Guo, Kang He, Jing Zhao, and Fei Li. 2021. "Chromosomal-Level Genomes of Three Rice Planthoppers Provide New Insights into Sex Chromosome Evolution." *Molecular Ecology Resources* 21 (1): 226–37.
McClure, Mark S. 1990. "Role of Wind, Birds, Deer, and Humans in the Dispersal of Hemlock Woolly Adelgid (Homoptera: Adelgidae)." *Environmental Entomology* 19 (1): 36–43.
Mendes, Fábio K., Dan Vanderpool, Ben Fulton, and Matthew W. Hahn. 2021. "CAFE 5 Models Variation in Evolutionary Rates among Gene Families." *Bioinformatics (Oxford, England)* 36 (22-23): 5516–18.
Millar, Neil S., and Ian Denholm. 2007. "Nicotinic Acetylcholine Receptors: Targets for Commercially Important Insecticides." *Invertebrate Neuroscience: IN* 7 (1): 53–66.
Panfilio, Kristen A., Iris M. Vargas Jentzsch, Joshua B. Benoit, Deniz Erezyilmaz, Yuichiro Suzuki, Stefano Colella, Hugh M. Robertson, et al. 2019. "Molecular Evolutionary Trends and Feeding Ecology Diversification in the Hemiptera, Anchored by the Milkweed Bug Genome." *Genome Biology* 20 (1): 64.
Petersen, Malte, David Armisén, Richard A. Gibbs, Lars Hering, Abderrahman Khila, Georg Mayer, Stephen Richards, Oliver Niehuis, and Bernhard Misof. 2019. "Diversity and Evolution of the Transposable Element Repertoire in Arthropods with Particular Reference to Insects." *BMC Evolutionary Biology* 19 (1): 11.
Plumb, George H. 1952. "The Formation and Development of the Norway Spruce Gall Caused by Adelges Abietis." https://doi.org/10.5555/19530603523.
Praveen, Kavita, Ying Wen, Kelsey M. Gray, John J. Noto, Akash R. Patlolla, Gregory D. Van Duyne, and A. Gregory Matera. 2014. "SMA-Causing Missense Mutations in Survival Motor Neuron (Smn) Display a Wide Range of Phenotypes When Modeled in Drosophila." *PLoS Genetics* 10 (8): e1004489.
Rhie, Arang, Brian P. Walenz, Sergey Koren, and Adam M. Phillippy. 2020. "Merqury: Reference-Free Quality, Completeness, and Phasing Assessment for Genome Assemblies." *Genome Biology* 21 (1): 245.
Rider, S. D., Jr, D. G. Srinivasan, and R. S. Hilgarth. 2010. "Chromatin-Remodelling Proteins of the Pea Aphid, Acyrthosiphon Pisum (Harris): Aphid Chromatin Genes." *Insect Molecular Biology* 19 Suppl 2 (0 2): 201–14.
Rispe, Claude, Fabrice Legeai, Paul D. Nabity, Rosa Fernández, Arinder K. Arora, Patrice Baa-Puyoulet, Celeste R. Banfill, et al. 2020. "The Genome Sequence of the Grape Phylloxera Provides Insights into the Evolution, Adaptation, and Invasion Routes of an Iconic Pest." *BMC Biology* 18 (1): 90.
Roach, Michael J., Simon A. Schmidt, and Anthony R. Borneman. 2018. "Purge Haplotigs: Allelic Contig Reassignment for Third-Gen Diploid Genome Assemblies." *BMC Bioinformatics* 19 (1): 460.
Rosenthal, Justin S., and Quan Yuan. 2021. "Constructing and Tuning Excitatory

Cholinergic Synapses: The Multifaceted Functions of Nicotinic Acetylcholine Receptors in Drosophila Neural Development and Physiology." *Frontiers in Cellular Neuroscience* 15 (September):720560.
Russo, Nicholas J., Chris S. Elphick, Nathan P. Havill, and Morgan W. Tingley. 2019. "Spring Bird Migration as a Dispersal Mechanism for the Hemlock Woolly Adelgid." *Biological Invasions* 21 (5): 1585–99.
Sano, Masakazu, and Kenichi Ozaki. 2012. "Variation and Evolution of the Complex Life Cycle in Adelgidae (Hemiptera): Complex Life Cycle of Adelgids." *Entomological Science* 15 (1): 13–22.
Selenko, P., R. Sprangers, G. Stier, D. Bühler, U. Fischer, and M. Sattler. 2001. "SMN Tudor Domain Structure and Its Interaction with the Sm Proteins." *Nature Structural Biology* 8 (1): 27–31.
Shen, Wei, Botond Sipos, and Liuyang Zhao. 2024. "SeqKit2: A Swiss Army Knife for Sequence and Alignment Processing." *iMeta* 3 (3): e191.
Shinoda, Tetsuro, and Kyo Itoyama. 2003. "Juvenile Hormone Acid Methyltransferase: A Key Regulatory Enzyme for Insect Metamorphosis." *Proceedings of the National Academy of Sciences of the United States of America* 100 (21): 11986–91.
Singh, Ravindra N., Matthew D. Howell, Eric W. Ottesen, and Natalia N. Singh. 2017. "Diverse Role of Survival Motor Neuron Protein." *Biochimica et Biophysica Acta. Gene Regulatory Mechanisms* 1860 (3): 299–315.
Smit, AFA, Hubley, R & Green, P. 2013-2015. "RepeatMasker Open-4.0." RepearMasker. 2013-2015. http://www.repeatmasker.org.
Stothard, P. 2000. "The Sequence Manipulation Suite: JavaScript Programs for Analyzing and Formatting Protein and DNA Sequences." *BioTechniques* 28 (6): 1102, 1104.
Thomas, Paul D., Dustin Ebert, Anushya Muruganujan, Tremayne Mushayahama, Laurent-Philippe Albou, and Huaiyu Mi. 2022. "PANTHER: Making Genome-Scale Phylogenetics Accessible to All." *Protein Science: A Publication of the Protein Society* 31 (1): 8–22.
Tsujimoto, Masafumi, Kazuma Aoki, Atsushi Ohnishi, and Yoshikuni Goto. 2020. "Endoplasmic Reticulum Aminopeptidase 1 beyond Antigenic Peptide-Processing Enzyme in the Endoplasmic Reticulum." *Biological & Pharmaceutical Bulletin* 43 (2): 207–14.
Vasimuddin, M., Sanchit Misra, Heng Li, and Srinivas Aluru. 2019. "Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems." *arXiv [cs.DC]*. arXiv. http://arxiv.org/abs/1907.12931.
Wang, Hengchao, Bo Liu, Yan Zhang, Fan Jiang, Yuwei Ren, Lijuan Yin, Hangwei Liu, Sen Wang, and Wei Fan. 2020. "Estimation of Genome Size Using K-Mer Frequencies from Corrected Long Reads." *arXiv [q-bio.GN]*. arXiv. http://arxiv.org/abs/2003.11817.
Wang, Yupeng, Haibao Tang, Jeremy D. Debarry, Xu Tan, Jingping Li, Xiyin Wang, Tae-Ho Lee, et al. 2012. "MCScanX: A Toolkit for Detection and Evolutionary Analysis of Gene Synteny and Collinearity." *Nucleic Acids Research* 40 (7): e49.
Weglarz, Kathryn M., Nathan P. Havill, Gaelen R. Burke, and Carol D. von Dohlen. 2018. "Partnering with a Pest: Genomes of Hemlock Woolly Adelgid Symbionts Reveal Atypical Nutritional Provisioning Patterns in Dual-Obligate Bacteria." *Genome Biology and Evolution* 10 (6): 1607–21.
Wu, Chao, Swapan Chakrabarty, Minghui Jin, Kaiyu Liu, and Yutao Xiao. 2019.

"Insect ATP-Binding Cassette (ABC) Transporters: Roles in Xenobiotic Detoxification and Bt Insecticidal Activity." *International Journal of Molecular Sciences* 20 (11): 2829.

Wu, Yi-Ting, Rui Ma, Jiang-Wen Wei, Li-Wen Song, Youssef Dewer, Sen-Shan Wang, Lei Liu, and Jing-Jiang Zhou. 2024. "ApCarE4 and ApPOD3 Participate in the Adaptation of Pea Aphids to Different Alfalfa Varieties." *Scientific Reports* 14 (1): 25444.

Xu, Jinrui, and Yang Zhang. 2010. "How Significant Is a Protein Structure Similarity with TM-Score = 0.5?" *Bioinformatics (Oxford, England)* 26 (7): 889–95.

Yeh, Hsin-Ting, Chiun-Cheng Ko, and Li-Wei Wu. 2020. "The First Complete Mitochondrial Genome of Adelges Tsugae Annand (Hemiptera: Adelgidae)." *Mitochondrial DNA. Part B, Resources* 5 (3): 2288–90.

Zhang, Yang, and Jeffrey Skolnick. 2004. "Scoring Function for Automated Assessment of Protein Structure Template Quality." *Proteins* 57 (4): 702–10.

Zhang, Zemin, and William I. Wood. 2003. "A Profile Hidden Markov Model for Signal Peptides Generated by HMMER." *Bioinformatics* 19 (2): 307–8.

Zhou, Qi-Hao, Qiang Zhang, Rui-Lin Yang, Guo-Rui Yuan, Jin-Jun Wang, and Wei Dou. 2022. "RNAi-Mediated Knockdown of Juvenile Hormone Acid O-Methyltransferase Disrupts Larval Development in the Oriental Fruit Fly, Bactrocera Dorsalis (Hendel)." *Pesticide Biochemistry and Physiology* 188 (105285): 105285.

Zilahi-Balogh, G., L. Humble, R. Foottit, J. Burleigh, and A. Stock. 2017. "History of the Balsam Woolly Adelgid, Adelges Piceae (Ratzeburg), in British Columbia with Notes on a Recent Range Expansion." *Journal of the Entomological Society of British Columbia* 113 (March):21–38.