Genomes of two invasive *Adelges* species (hemlock woolly adelgid and pineapple gall adelgid*)* enable characterization of nicotinic acetylcholine receptors

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

*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

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

Keywords: adelgids, reference genome, Cys ligand-gated Ion channel receptors (Cys-LGIC), forest tree health, conifers, invasive species, insecticides, conservation

Abstract: Two invasive hemipteran adelgids cause widespread damage to North American conifers. *Adelges tsugae* (the hemlock woolly adelgid) has decimated *Tsuga canadensis* and *Tsuga caroliniana* (the Eastern and Carolina hemlocks, respectively). *A. tsugae* was introduced from East Asia and reproduces parthenogenetically in North America, where it can kill trees rapidly. *A. abietis,* introduced from Europe, makes “pineapple” galls on several North American spruce species, and weakens trees, increasing their susceptibility to other stresses. Broad-spectrum insecticides that are often used to control adelgid populations can have off-target impacts on beneficial insects and the development of more selective chemical treatments could improve control methods and minimize ecological damage. Whole genome sequencing was performed on both species to aid in development of targeted pest control solutions and improve species conservation. The assembled *A. tsugae* and *A. abietis* genomes are 220.75 Mbp and 253.16 Mbp, respectively, each consisting of nine chromosomes and both genomes are over 96%

complete based on BUSCO assessment. Genome annotation identified 11,424 and 14,118 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 expansions in *A. abietis* included ABC transporters and carboxypeptidases involved in carbohydrate metabolism, while both species showed contractions in core histone families and oxidoreductase pathways. Gene family expansions in *A. tsugae* highlighted families associated with the regulation of cell differentiation and development (*survival motor protein*, SMN; *juvenile hormone acid methyltransferase* JHAMT) as well as those that may be involved in the suppression of plant immunity (*clip domain serine protease-D*, CLIPD; *Endoplasmic reticulum aminopeptidase 1*, ERAP1). Among the analyzed gene families, Nicotinic acetylcholine receptors (nAChRs) maintained consistent copy numbers and structural features across species, a finding particularly relevant given their role as targets for current forestry management insecticides. Detailed phylogenetic analysis of nAChR subunits across adelgids and other ecologically important insects revealed remarkable conservation in both sequence composition and predicted structural features, providing crucial insights for the development of more selective pest control strategies.

Introduction

The introduction of several adelgid species into North America has had a devastating impact on forest health, in some instances leading to rapid tree mortality in large tracts of coniferous forests (Ellison et al. 2018; Fei et al. 2019). The adelgids include over 60 species of small sap-sucking hemipteran insects within the infraorder 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 can also be observed in the stylet of adelgids, which are longer and more flexible to allow for 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).

Adelges abietis, commonly referred to as the pineapple gall adelgid or eastern spruce gall adelgid, primarily targets Norway spruce (*Picea abies*) (Figure 1A). This species is characterized by its gall-inducing behavior, forming structures on host trees that disrupt normal growth patterns, leading to decreased vigor and increased susceptibility of host trees to other pests and pathogens (Cornelissen et al. 2008). It is believed to have arrived in North America in the early 20th century via the nursery trade. In North America, it targets non-native Norway spruce as well as native species, such as black spruce (*Picea nigra*), white spruce (*Picea glauca*), and red spruce (*Picea rubens*) (Pilichowski and Giertych 2018). The life cycle of *A. abietis* is completed entirely on spruce, and consists of two all-female generations per year, with fundatrices overwintering at the base of buds. Most individuals remain on the same tree where they were born, resulting in a clustered distribution that can exacerbate damage to host trees (Carter 1971).

The hemlock woolly adelgid, *Adelges tsugae*, is an invasive pest that infests Eastern hemlock (*Tsuga canadensis*) and Carolina hemlock (*Tsuga caroliniana*) throughout Eastern North America (Figure 1D). Originally from East Asia, *A. tsugae* was introduced to North America in the early 1950s, with genetic analyses indicating multiple independent introduction events (Mayfield et al. 2023). Among adelgids, *A. tsugae* is unique in using hemlocks as a secondary host. Tigertail spruce (*Picea polita*) is the primary host for *A. tsugae*, but is endemic to Asia. The invasive *A. tsugae* lineages reproduce parthenogenetically in North America. Thus, all *A. tsugae* in North America are female and spread in the crawler phase or as minuscule eggs (McClure and Cheah 1999). A characteristic feature of *A. tsugae* is the presence of white, woolly sacs on the underside of hemlock branches (Figure 1E). Each wool sac is capable of producing over 400 eggs, and combined with its parthenogenic nature, and ability to readily hitchhike on forest dwellers, *A. tsugae* have contributed to a significant decline in Eastern hemlock forests (Orwig and Foster 1998), Figure 1F).

Despite the ecological significance of adelgids, there has been little genetic research on these species. Recent studies have provided insight into adelgid mitochondrial genomes (Havill et al. 2006; Yeh et al. 2020), the relationships between adelgids and their bacterial endosymbionts (Weglarz et al. 2018; Szabó et al. 2022), early detection of adelgid invasion by environmental DNA (Sanders et al. 2023), and the evolutionary history of the clade within Aphididae (Mayfield et al. 2023). To date, among adelgids, only the Cooley spruce gall adelgid (*Adelges cooleyi*) genome has been sequenced, highlighting a significant gap in genomic resources for adelgids (Dial et al. 2023). The genome of *A. tsugae* and *A. abietis* will be invaluable for ongoing forest health studies utilizing environmental DNA collection and barcoding, as well as for efforts to track trait evolution across invasions and illuminate targets for insecticides.

Pest control strategies for adelgids include chemical treatments, biological control, and silvicultural management (Limbu et al. 2018). Integrated pest management (IPM) combines chemical applications with biological controls like non-native predatory insects (Havill and Foottit 2007). Breeding resistant *Tsuga* hybrids (Montgomery et al. 2009) or identifying lingering cultivars (Kinahan et al. 2020) are potential long-term strategies. Chemical treatments include horticultural oil, insecticidal soap, and systemic neonicotinoids, which suppress adelgid populations (Joseph et al. 2011). Neonicotinoids, acting as nicotinic acetylcholine receptor (nAChR) agonists, are effective but nonselective, risking off-target mortality in native insects (Suchail et al. 2001). To mitigate this, neonicotinoids are applied to individual trees via basal bark application, soil drench, or trunk injection. These precision approaches reduce, but do not eliminate off-target toxicity. Developing selective insecticides could improve forest management. We aimed to identify evolutionary traits unique to Adelgidae, focusing on nAChRs, the target of current insecticides.

Methods & Materials:

DNA Sampling, Extraction, Preparation, and Sequencing

A. abietis were collected from a single gall growing on *Picea rubens* collected near 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. HMW DNA extraction is described in detail 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-reads library was prepared following the manufacturer's protocol and the library was sequenced on an Illumina Nextseq 550.

Scaffolding reads in the form of Hi-C Illumina short reads were generated from a single individual collected from a single gall growing on the same *Picea rubens* accession on 16 August 2021. DNA extraction on the adelgid is described in File S1. The extracted DNA was sent to Phase Genomics (Seattle, WA) for Hi-C library preparation and sequencing.

A. tsugae were collected from a single accession of *Tsuga canadensis* growing at the Mountain Research Station in Waynesville, NC (35.487 °N, –82.967 °W) (Figure 1E). Approximately 4 feet of one branch was collected in October 2023 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 were collected in total to provide sufficient material for DNA extraction.

Genomic DNA for nanopore sequencing was extracted from the pooled sample using the Qiagen MagAttract HMW DNA Kit, with modifications based on (Mao et al. 2023) and further adaptations outlined in 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 on the extracted *A. tsugae* DNA. 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

A. tsugae and *A. abietis*: Individual *A. tsugae* were collected from branches of *Tsuga canadensis* growing in Leesburg, VA, USA while *A. abietis* were collected from galls at the same location as the DNA sampling. For both adelgids, salivary glands and carcasses were dissected and collected separately into Arcturus PicoPure Extraction buffer. Total RNA was prepared using the Arcturus PicoPure RNA Isolation kit including the optional DNAse step. Barcoded RNA-seq libraries were prepared for Illumina NextSeq 550 sequencing (150bp PE) with a method described previously (Cembrowski et al. 2018). For *A. tsugae*, 8 libraries passed quality control with read lengths of 141-144 bp and 18.1-25.3 million bases per library (GC content 39.5-43.8%). For *A. abietis*, 20 libraries passed quality control with read lengths of 127-145 bp and 15.3-24.4 million bases per library (GC content 33.7-50.2%). Libraries with fewer than 15 million reads or mean read length below 70 bp were excluded from analysis.

Genome assembly

A. abietis: 10X linked reads were assembled into a draft genome with Supernova using 1.19G linked reads and 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 (Kumar et al. 2013; Laetsch and

Blaxter 2017) 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) with the options “-outfmt '6 qseqid staxids bitscore std sscinames sskingdoms stitle' -culling_limit 5 -evalue 1e-25”. 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 scaffold length with SeqKit v2.9.1 and assessed for with BUSCO, QUAST, 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 et al. 2024).

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 (Wang et al. 2020) with GCE v1.0.2 (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 v5.4.5 (Manni et al. 2021) hemiptera_odb10 for completeness, Merqury v1.3 (Rhie et al. 2020) for quality, and QUAST v5.2.0 (Gurevich et al. 2013) for contiguity. The Canu assembly was selected for downstream analysis. Heterozygosity was reduced for draft assemblies using Purge Haplotigs v1.1.2 (Roach et al. 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 nine chromosomes in descending order by length. The final assembly was re-assessed for quality metrics.

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, AFA, Hubley, R & Green, P. 2013-2015). 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 expression support within EASEL by alignment to their corresponding genomes, with *A. tsugae* informed by 4 salivary gland libraries, 4 carcass libraries (salivary glands removed) (SRR30936309), and an adult overwintering sistens library (SRR1198669) and *A. abietis* informed by 2 salivary

gland libraries, 4 carcass libraries, and a whole body library (SRR30936310). NCBI's RefSeq protein database (release 208) and OrthoDB v11 (Kuznetsov et al. 2023), with the addition of *A. cooleyi* proteins, provided external protein level support in the pipeline.

Genome synteny in *Adelges*

Genome synteny was assessed using protein-protein comparisons of the *A. tsugae* and *A. abietis* annotations using the GENESPACE pipeline v1.3.1 (Lovell et al. 2022). GENESPACE was run with default parameters, OrthoFinder v2.5.5, and MCScanX (Wang et al. 2012; Emms and Kelly 2019).

Comparative genomics among hemipteran

All Hemipteran species with genome annotations available on NCBI (38 total) were considered for comparative analysis, in addition to 14 related species established as outgroups of Hemiptera. Annotations with BUSCO protein completeness below 80% or gene counts greatly exceeding the average range were removed. Annotations that passed filtering were analyzed with Orthofinder v2.5.5 and 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. Orthogroups with extreme single-species representation (one species ≥50 orthologs and all others ≤3), representing only two species *(cladeandsizefilter.py)*, and groups with the largest differential (63 or more orthologs) between the minimum and maximum representation, were removed. Gene family expansions and contractions for those genes that are expanded or contracted with $P < 0.01$ were reported.

Comparative analysis of nAChR:

A total of 454 reference nAChR alpha-7 genes, representing 1056 transcripts, were obtained from NCBI's RefSeq protein database from 269 species across Insecta. A motif file was constructed via hmmer v3.3.2 (Zhang and Wood 2003) to scan all proteins from all species assessed to identify nAChRs without exclusive reliance on existing functional annotations. The resulting proteins were aligned with MAFFT v7.511 (Katoh and Toh 2008).

Phylogenetic relationships among sequences were inferred from the multiple

sequence alignment with IQ-TREE v1.6.12 (Nguyen et al. 2015). The Model Finder function was employed to identify the best-fit evolutionary model for the data-Q.insect+R9 (Kalyaanamoorthy et al. 2017). The resulting maximum likelihood tree was visualized and annotated using the Interactive Tree of Life (iTOL) web server v6.9 (Letunic and Bork 2024). Seven motifs for each of the transcripts and seed sequences were generated with MEME v0.10.1 (Bailey et al. 2015) based on seven functional nAChR domains and 501 proteins with p-value < 1e-30 were added to the gene trees.

A subset comprised of the probable nAChR genes generated by MEME analysis for *A. tsugae*, *A. abietis*, *A. cooleyi*, *Rhopalosiphum maidis*, *Apis mellifera*, *Danaus plexippus*, and *Drosophila melanogaster* was selected for further analysis. These were selected based on relatedness to Adelgidae (*R. maidis*, corn leaf aphid), as non-target pollinators of interest (*A. mellifera*, Western honey bee; *D. plexippus*, monarch butterfly), or as models with well characterized nAChRs (*D. melanogaster*, common fruit fly). Protein sequences were aligned in MEGA v11.0.13 (Kumar et al. 2018) with MUSCLE v3.8.31 (Edgar 2004). For the aligned sequences, 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).

Maximum likelihood gene trees were produced for both the full set of 501 genes and the selected species subset using MAFFT and IQ-TREE 2 (Minh et al. 2020). ModelFinder chose substitution model “Q.insect+R9” and rate heterogeneity model “FreeRate with 9 categories" for the full set and “VT+I+R5” and “Invar+FreeRate with 5 categories” for the subset (Figure S1) (Kalyaanamoorthy et al. 2017).

<u>nAChR Homology Modeling and Structural Comparison</u>

Structures of insect nAChRs proteins from the selected subset of species were predicted using AlphaFold3 (Abramson et al. 2024). Since these proteins are pentameric, structure estimation was simplified by folding all nAChR proteins as homopentamers, although only some nAChR subtypes are known to homopentamerize *in vivo.* Proteins were folded in duplicate to ensure the reproducibility of AlphaFold outputs, with arbitrary random seeds of 1 and 7000 between duplicate foldings. The batch using seed 7000 was selected for the final analysis due to improper folding which was noticed in some *A. tsugae* nAChrR AF3 models generated in the seed 1 batch. Improper folding was identified by abnormally high template modeling (TM) scores when compared to other nAChRs from the same batch. These scores were calculated in C++ 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). Full data and confidence values for all AF3 models are available in supplemental file S3.

nAChR Active Site Analysis

To identify insect nAChR active site residues, sequences were compared to known human nAChR active site residues (Gharpure et al. 2020). Insect nAChRs were aligned in MEGA with MUSCLE and visualized as logo plots (Crooks et al. 2004). Areas with high homology were then compared with human active site sequences, leading to the identification of likely active sites. AlphaFold3 generated protein structures were then used to check the locations of the identified residues against ligand-bound nAChR and acetylcholine-binding protein crystal structures (PDB: 5FJV, 2ZJU).

Results and Discussion:

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

Adelges abietis 10X sequencing generated 119M Illumina short reads (Table S1). The Supernova assembled genome produced a reference of 290.32 Mb in 9,607 contigs with an N50 of 4.77 Mb. Scaffolding conducted via Hi-C produced an assembly that is 253.16 Mb in length, 87.2% of which is contained within 9 (2*n*=18) chromosomes (Steffan 1968; Favret et al. 2015). This reference reports 2998.36 N's per 100 Kb, and an N50 of 27.91 Mb. 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).

Adelges 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 S1, Table S2). Among the 4,037,933 (13.60%) reads classified as contaminants, the most abundant species was *Candidatus Pseudomonas adelgestsugas*, one of *A. abietis*'s two dual-obligate endosymbiotic bacteria, with just over 3 M unique reads representing 75.7% of those removed. The other paired endosymbiont, *Ca. Annandia adelgestsuga*, was not 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 (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 S2).

Genome size was estimated at 267.97 Mb yielding 290x read coverage. The Canu assembled draft genome contained 2,131 contigs with a 7 Mb N50. Heterozygosity reduction 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 220.75 Mb in length, 99.95% of which is contained within nine putative chromosomes. This reference reports 1.45 N's per 100 kb, an N50 of 24.57 Mb, and a Merqury QV of 42.88. 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 to date, the Phylloxeroidea tend to have

smaller genomes than the Aphidoidea. The average genome size of Aphidoidea is 369.4 Mb, and none are smaller than 300 Mb. In contrast, all of the Phylloxeridae genomes are smaller than 300 Mb. *A. tsugae* is the smallest, with an estimated genome size of 220.75 Mb, and *A. abieties*, *A. cooleyi*, and *Daktulosphaira vitifoliae* have estimated genome sizes of 253.16 Mb, 270.2 Mb, and 282.6 Mb, respectively (Table S3).

Structural and functional annotation:

For structural annotation,17.90% of the *A. tsugae* genome and 24.91% in *A. abietis* were softmasked and identified as repetitive DNA (Table 1). Transposable element (TE) annotation classified 15.13% and 25.98% of the *A. abietis* and *A. tsugae* genomes, respectively (Figure 2A, 2B). 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. tusgae*) and SINEs (0.00% *A. abietis*, 0.01%) 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 landscapes of *A. tsugae* shows a continuous pattern of TE accumulation across Kimura substitution levels, with four distinct bursts of activity, whereas *A. abietis* shows two distinct peaks. The most recent burst reflects predominately DNA transposon and LINE element activity, with a smaller contribution from LTR retrotransposons. The three older bursts in *A. tsugae* shows similar profiles, dominated by unclassified elements but with consistent contributions from LINEs and DNA transposons. While LTRs make up a smaller proportion of each burst, they maintain presence across all bursts. The recovery of four distinct peaks of TE activity in *A. tsugae*, differs from that of *A. abietis*, and in aphids, which typically show evidence of more recent TE activity, with large families of DNA transposons dominating this activity, as observed in species like *M. persicae* and *M. dirhodum* (Figure 2A, 2B) (Baril et al. 2023).

Genome annotation identified 11,424 *A. tsugae* 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 14,118 *A. abietis* protein-coding genes were identified with a BUSCO completeness of C:96.5% [S:94.9%, D:1.6%], F:0.3%, M:3.2% (Table 1; Table S4). The gene count of *A. tsugae* is lower than the estimates for the *A. abietis* and *A. cooleyi* (13,556 protein-coding genes), and is the lowest estimate among available Hemiptera genomes with the exception of the 12,266 *Aphis gossypii* protein-coding genes (Table S3, Table S4) (Zhang et al. 2022). 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 S4). Assessment of synteny identified 196 syntenic blocks between *A.*

tsugae and *A. abietis* genomes, with high collinearity within each chromosome pair. While the overall chromosomal structure is conserved, internal rearrangements are particularly visible in chromosomes 1 and 2 (Figure 2E). This level of chromosomal conservation contrasts markedly with patterns seen in aphids, where chromosome homology cannot be determined between tribes that diverged over 30 million years ago, but aligns more closely with the conserved synteny observed in blood-feeding hemipterans in the *Rhodnius* and *Triatoma* genus (Mathers et al. 2021).

Comparative genomics among hemipteran

Across the Hemiptera, 31,666 orthogroups were identified, including 643,982 (93.69%) of 687,323 total proteins examined, 85.61% of which were assigned functional descriptions based on their longest representative sequences (Table S6). A total of 1908 orthogroups were shared by all 45 species, 59 of which were estimated to be orthologous and single-copy. 22,254 orthogroups (68.93%) had no representation in *Adelges*, conversely, 862 orthogroups (2.67%) were represented exclusively by one or more adelgids and 205 (0.63%) by all three. A total of 33, 189, and 123 groups were unique to *A. tsugae*, *A. abietis*, and *A. cooleyi*, respectively (Figure 3A).

The two adelgids described here were included in a set of 29 Hemipteran relatives and across families (Aphididae, Adelgidae, Aleyrodidae, Aphalaridae, Cicadellidae, Cimicidae, Coccidae, Delphacidae, Miridae, Pediculidae, Pentatomidae, Phylloxeridae, Pseudococcidae, Psyllidae) and 14 additional arthropods serving as outgroups (Thysanoptera, Blattodea, Coleoptera, Diptera, Hymenoptera, Lepidoptera, Arachnida, Branchiopoda) for a total of 45 species curated by availability of high quality genome resources (Table S3). Phylogenetic analysis of orthologous gene families support Phylloxeridae (represented by *Daktulosphaira vitifoliae*) and Adelgidae as sister families composing superfamily Phylloxeroidea, sister group to Aphidoidea (Dial et al. 2023) (Figure 3B).

Within this phylogenetic framework, of the 244 orthogroups showing significantly different evolutionary rates between parent and child nodes, 235 remained after filtering out TE families. Of these, 32 orthogroups are completely absent across all three *Adelges* species. While N- acetylegalactosaminyltransferase-9 was expanded in *D. vitifoliae*, the remaining 31 orthogroups were likely lost in a common ancestor before the divergence of adelgids and phylloxerans (Table S6). *Signal transduction mechanisms* were most abundant in the lost families, and included G-protein receptors, immunoglobulins, diaglycerol binding proteins, and PET domain proteins involved in cell differentiation. Conversely, among the 22 orthogroups present in at least one adelgid and absent in *D. vitifoliae*, 27% are assigned to *energy production and conversion* functions - primarily related to the mitochondrial respiratory chain, especially NADH pathways.The differential retention of these functional categories may reflect the specialized feeding ecology of adeligids on Pinaceae hosts. While

phylloxera have evolved many unique genes that are active during feeding and interacting with their hosts (Rispe et al. 2020), adelgids have expanded their core metabolic functions, with a focus on energy production, reflecting the patterns of highly specialized herbivores (Dohlen and Moran 2000).

At the *Adelges* most recent common ancestor (MRCA) node, 24 orthogroups are expanding and 34 contracting. Within this lineage, *A. tsugae* has a total of 19 expanding and 62 contracting orthogroups, while *A. abietis* has 29 expanding and 46 contracting (Figure 3B; Table S7). The largest expansions in *A. abietis* were associated with *carbohydrate transport and metabolism*, particularly ABC transporters and carboxypeptidases, suggesting adaptation to host plant metabolism, a pattern commonly observed in specialized herbivores such as aphids where metabolic gene families undergo adaptive evolution to cope with host plant defenses (Wu et al. 2019; Wu et al. 2024).

Specific expansions were observed in *A. tsugae* for Endoplasmic reticulum aminopeptidase 1(ERAP1)—a domain commonly associated with the salivary effector protein family, aminopeptidase-N (apN), where transcriptomic and proteomic analyses have shown that ERAP-1 proteins undergo lineage specific expansions, and shows evidence of positive selection in in *Acyrthosiphon pisum (Boulain et al. 2018)*. Although its specific function in this context remains unclear, proteins secreted from plant pests have been implicated in the negative regulation of plant immunity (Chen et al. 2019). In addition, clip domain serine protease- D (CLIPD) , another rapidly evolving and expanding gene family in *A. tsugae,* is associated with regulating immune response in insects (Ruzzante et al. 2022). The survival motor neuron protein (SMN) family was significantly expanded in *A. tsugae* and is a well-studied target in *Drosophila*, which serves as a model for examining this key gene associated with muscular dystrophy in humans (Grice and Liu 2011). In insects, SMN has been shown to interact with several gemins and is critical for RNA processing and cell differentiation (Maccallini et al. 2020). Loss-of-function studies of SMN in *Drosophila* have revealed defective cell divisions and disruptions to the developmental cycle (Rajendra et al. 2007). The expansion of SMN in *A. tsugae,* contrasting with its contraction in *A. abietis*, suggests potentially different requirements for RNA processing and cellular differentiation between these closely related species.

While the MRCA of *A. abietis* and *A. tsugae* exhibits contractions in core histone families H2A/H2B/H3/H4, aphids demonstrate expanded diversity in chromatin-remodeling proteins compared to other arthropods, a pattern possibly linked to their developmental plasticity (Rider et al. 2010; Mathers et al. 2017). This contrast in chromatin-associated gene families may reflect different evolutionary strategies; aphids require greater regulatory flexibility for their complex life cycles and host switching, while the adelgids rely on specific conifer hosts (Havill et al.

2007).

Juvenile hormone is a crucial sesquiterpenoid hormone that regulates a wide range of essential physiological processes, including development and reproduction (Shinoda and Itoyama 2003). While the *juvenile hormone acid methyltransferase* (JHAMT) shows contraction in the adelgid MRCA, *A. tsugae* shows expansion of the haemolymph juvenile hormone binding protein (JHP). Studies in aphids have shown that silencing *juvenile hormone acid methyltransferase* (JHAMT), a rapidly expanding gene family in *A. tsugae*, leads to increased mortality and delayed development (Zhang et al. 2024).

Proteins associated with phosphorylation and dephosphorylation pathways showed variable patterns of evolution, with serine/threonine protein kinases expanding in *A. abietis* while phosphatases contracted in the adelgid MRCA (Table S7). These regulatory proteins are of particular interest as they modulate cholinergic receptor function through post-translational modifications (Wiesner and Fuhrer 2006; Makhnovskii et al. 2013; Jiménez-Pompa et al. 2023). Within this group, proteins involved in phosphorylation-dependent regulation interact with nicotinic acetylcholine receptors (nAChRs), which are cholinergic ligand-gated ion channels mediating fast synaptic transmission in insect nervous systems and serving as common targets of insecticides (Matsuda et al. 2001). While these regulatory proteins show dynamic evolution, the nAChR genes themselves maintain relatively conserved distributions across both adelgid genomes.

Comparative analysis of nAChRs:

Since nAChRs are the target of the most commonly used insect control methods for adelgids (P McCarty and Addesso 2019), a better understanding of nAChR diversity within Adelgidae and between Adelgidae and Insecta was sought. While nAChR composition and expression are not the only factors 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).

Gene family analysis placed the 501 nAChR curated genes into 14 orthogroups . Six of these were exclusively annotated as nAChR genes. The other eight included 168 additional genes, which had a range of predicted functions ranging from diverse cholinergic receptors, or other ion channel related proteins. One group was associated with *inorganic ion transport and metabolism* while all other groups were associated with *signal transduction mechanisms.* None of the 14 orthogroups were significantly expanding, and estimates on the rate of evolution could only be estimated for eight families, none meeting the threshold of a p-value < 0.01.

To facilitate a comparative analysis of nAChRs within Adelgidae and throughout

Insecta, a group of seven relevant insects were chosen. This group included *A. tsugae* (HWA), *A. abietis* (PGA), *A. cooleyi*, *Rhopalosiphum maidis*, *Apis mellifera*, *Danaus plexippus*, and *Drosophila melanogaster*. Additionally, the 12 identified *A. tsugae* nAChRs were each assigned an individual name (HWA1–HWA12) to facilitate discussion (Figure 2C, 2D; Table S8).

When clustered phylogenetically by protein sequence in an unrooted maximum likelihood gene tree (Figure 4A), insect nAChRs tended to cluster by subtype, reflecting their structural and functional conservation. The designations of these clusters are apparent from the inclusion of proteins with known designations from common reference organisms, such as *D. melanogaster* and *A. mellifera*, with clear clusters forming for subtypes α1, α2, α3, α4, β1, and β2. This clustering pattern reflects the fundamental structural organization of nAChRs as members of the Cysteine (Cys)-Loop ligand-gated ion channel (CysLIGC) receptor family, which are highly conserved throughout Insecta and most eukaryotes.

These nAChRs are composed of five subunits, being either homopentameric or heteropentameric, and are membrane-bound, with extracellular, transmembrane, and intracellular regions (Figure 4B and 4C). The subtypes are primarily differentiated by the presence or absence of a disulfide linkage between two adjacent cysteine (Cys) residues in the ligand binding site, separating them into α and β classes. 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). Within this structural framework, subtypes α5 and α7 did not partition into distinct lineages from one another, and β3 did not separate well in clustering, while HWA1, HWA3, HWA4, HWA9, and HWA10 could not be assigned to clear clusters.

High sequence homology between nAChR sequences from *A. tsugae* and other selected species was observed (Figure 5A and Figure S2). HWA2, HWA4, HWA9, and HWA10 stand out as being the most unique of the observed nAChRs, each having consistently lower sequence similarity with other nAChRs. HWA4 and HWA10 each had two high similarity matches, one with an *A. abietis* nAChR and one with an *A. cooleyi* nAChR, suggesting that these nAChRs may be rather unique to Adelgidae. HWA9 had the lowest observed overall similarity within the set, making it the most unique of the studied nAChRs. Similar trends emerge when the nAChRs are compared via template modeling (TM scores) of 3D AlphaFold 3 (AF3) models (Figure 5B and Figure S3). Overall, the observed nAChRs showed a high level of structural homology, with occasional exceptions such as HWA2 and HWA6, both of which have on average lower TM scores when compared to the other nAChRs (0.5-0.6 as opposed to 0.7-0.8). Generally, the largest structural differences are seen

in the transmembrane domain, with insertions of transmembrane domains observed in some proteins.

Both the structures and sequences of the ligand binding sites were analyzed. In order to probe ligand binding site conservation across species, a representative *A. tsugae* nAChrR AF3 structure (Figure 5C, HWA5) was compared with a relatively distantly related human nAChR crystal structure (Figure 5D, PDB: 5FJV). Overall, the structures are very similar. Key aromatic tyrosine (Tyr) and tryptophan (Trp) residues, known to be important in ligand binding, are conserved (Barington et al. 2016). The key geminal disulfide bond can be seen in the human crystal structure, and while its linkage was not successfully predicted by AF3, both Cys residues can be seen in the expected positions. Notably, the human nAChR binding site is in a closed conformation, because the PDB structure is ligand bound (ligand hidden for clarity; Figure 5D), while the AF3 generated HWA5 nAChR structure is in an open conformation, thus explaining the difference in relative positions of the disulfide moieties. Aside from this one difference, both binding site structures exhibit near total structural overlap. To assess if these ligand binding sites differ between Adelgidae and the rest of Insecta, the aligned ligand binding sites for each group were visualized as logo plots (Figure 5E). The nAChR ligand binding site sequences are highly conserved across all observed species. Loop C contains a variable number of residues, with some nAChRs having one or two amino acid inserts within the loop. This can be seen as high variability at positions 42, 46, and 47 in the logo plots (Figure 5E). It is unlikely that these inserts would affect ligand binding, since they are predominantly present in β-type sequences which lack the geminal Cys-Cys linkage and are therefore incapable of ligand binding. Loop C has been extensively investigated for its role in receptor function, particularly in ligand binding and channel gating. Key mutations in Loop C, such as the P242E mutation in Drosophila and E219P in vertebrate receptors, significantly affect neonicotinoid sensitivity, demonstrating its importance in species-selective insecticide action (Toshima et al. 2009; Shimada et al. 2020).

Recent research has expanded beyond nAChRs to examine the entire CysLGIC superfamily, especially in the context of protecting ecologically beneficial species, as these channels serve as targets for various insecticides including avermectins, fipronil, and dieldrin (Meng et al. 2015; Wang et al. 2024). This broader investigation is crucial because insecticides target multiple members of this superfamily, including γ-aminobutyric acid gated chloride channels, glutamate-gated chloride channels, histamine-gated chloride channels, and pH-sensitive chloride channels. Broadening the study of nAChR conservation across Insecta is important to future goals of selective insecticide development, since nAChR similarities exceed 60% between species and reach up to 98% in some cases (Figure 5A; Figure S2). (Jones et al. 2007).

Conclusion: The high-quality genome assemblies presented here for *A. tsugae* and *A. abietis* reveal high 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 involved in cell differentiation and development, and suppression of plant immunity. These genomic resources, combined with the detailed characterization of nAChR diversity and distribution, provide a foundation for understanding the evolution of insecticide targets in adelgids.

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

Assembly	**Total Chromosomes**	**N50 (Mb)**	**Length (bp)**	**Sequence in Chroms**	**% Gaps**	**% Repeat**
A. tsugae	9	24.56 Mb	220,755,323	95.27%	0.000%	17.90%
A. abietis	9	27.91 Mb	253,165,100	87.2%	2.998%	24.91%
BUSCO Genome, Hemiptera						
	Complete	**Single**	**Duplicate**	**Fragment**	**Missing**	***n* Searched**
A. tsugae	99.3%	98.1%	1.2%	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	96.5%	94.9%	1.6%	0.3%	3.2%	2510

Figures:

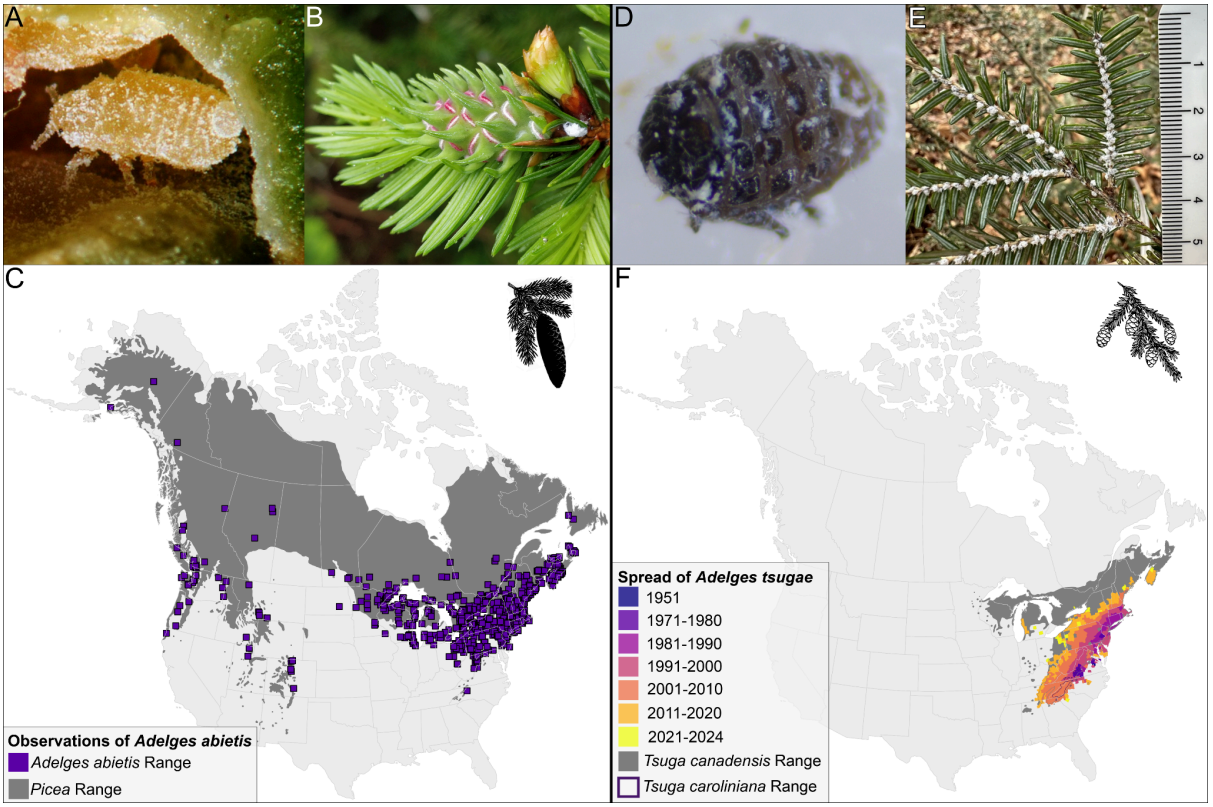


Figure 1. (A) *A. abietis* immature gallicolae in gall on *Picea abies* (source: influentialpoints.com). (B) *A. abietis* gall in early development, with stem mother visible. (C) Distribution of North American *Picea* and observations of *Adelges abietis* (source: GBIF, iNaturalist). (D) Hemlock woolly adelgid adult (Collected July, 2024, Ann Arbor, MI). (E) Eastern hemlock branch infested with *A. tsugae*, with centimeter ruler for scale. (F) Range of *Tsuga canadensis* and *Tsuga caroliniana* (Little and Luther 1971) and the spread of *Adelges tsugae* since 1951 (source: 1951-2020 Data derived from USDA FS-FHP, 2021-2024 observations compiled from infestation reports).

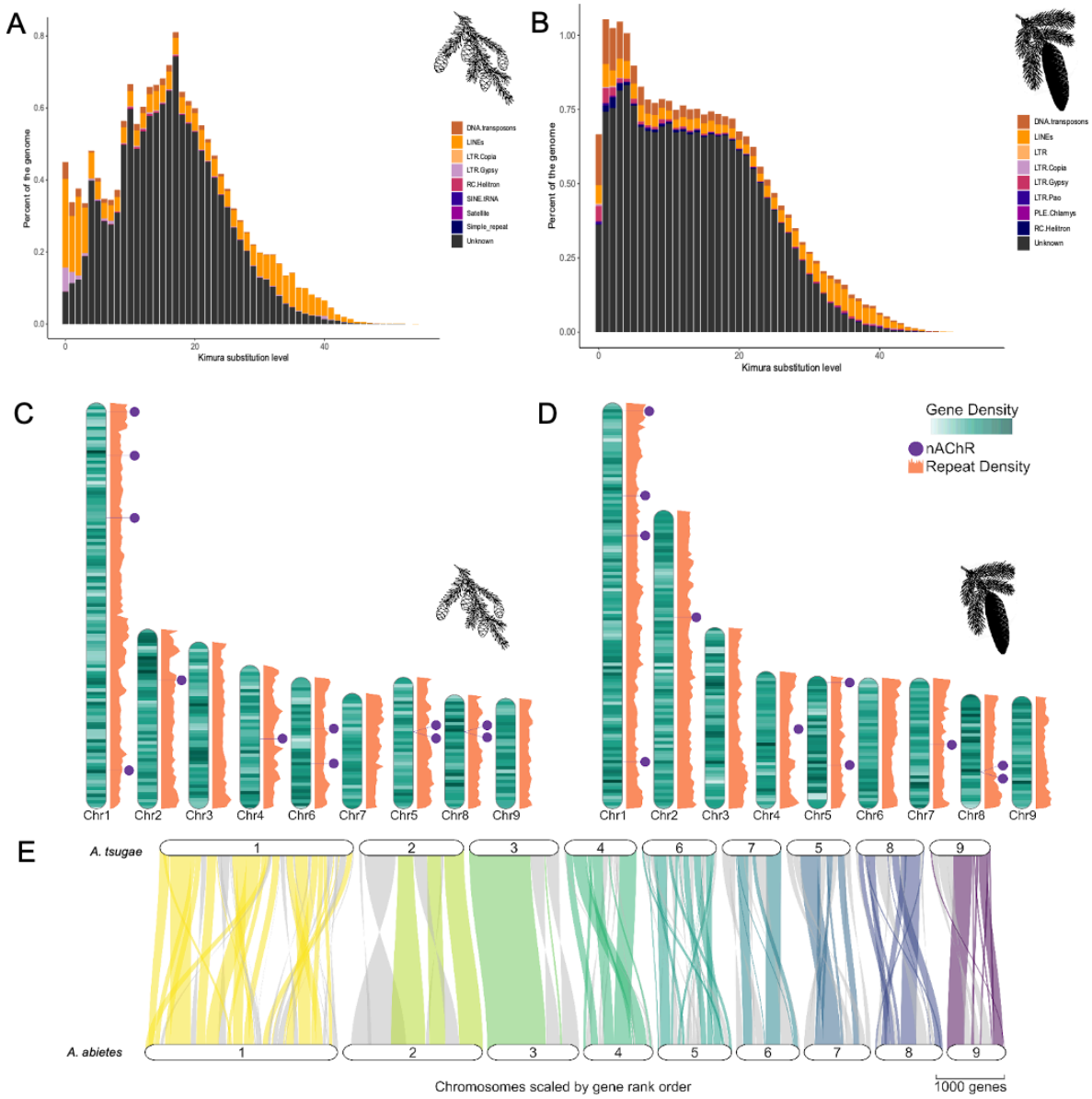
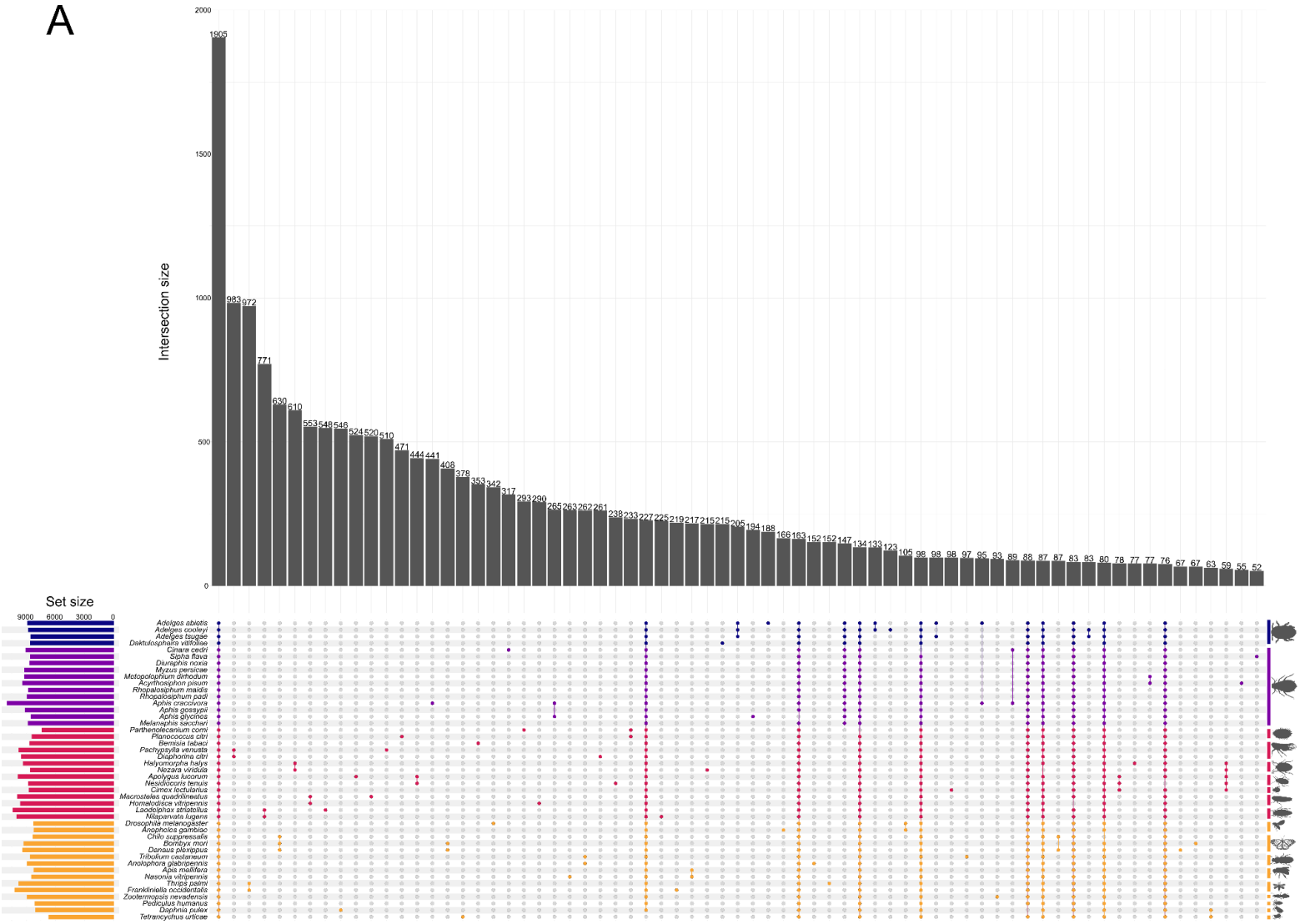


Figure 2. (A, B) Repeat landscape of known TE families in *A. tsugae* and *A. abietis*, respectively. (C&D) Chromosome models of *A. tsugae* (left) and *A. abietis* (right) with gene density from light to dark (*A. tsugae* 0-56 genes per 500,000 bp, *A. abietis* 1-49), repeat density tracks alongside chromosomes (*A. tsugae* 2-1231 repeats per 500,000 bp, *A. abietis* 224-1317), and nAChR positions marked by points. While numbered according to length, *A. tsugae* chromosomes are presented in the order produced by scaffolding to show homology. (E) Synteny between *A. abietis* and *A. tsugae*. Inverted regions are desaturated to differentiate them from their respective chromosomes.

A


Intersection size
2000
1500
1000
500
0
1905
983
972
771
630
610
553
548
546
524
520
510
471
444
441
408
378
353
342
317
293
290
265
263
262
261
238
233
227
225
219
217
215
215
205
194
188
166
163
152
152
147
134
133
123
105
98
98
98
97
95
93
89
88
87
87
83
83
80
78
77
77
76
67
67
63
59
55
52
Set size
9000
6000
3000
0
Adelges abietis
Adelges cooleyi
Adelges tsugae
Daktulosphaira vitifoliae
Cinara cedri
Sipha flava
Diuraphis noxia
Myzus persicae
Metopolophium dirhodum
Acyrthosiphon pisum
Rhopalosiphum maidis
Rhopalosiphum padi
Aphis craccivora
Aphis gossypii
Aphis glycines
Melanaphis sacchari
Parthenolecanium corni
Planococcus citri
Bemisia tabaci
Pachypsylla venusta
Diaphorina citri
Halyomorpha halys
Nezara viridula
Apolygus lucorum
Nesidiocoris tenuis
Cimex lectularius
Macrosteles quadrilineatus
Homalodisca vitripennis
Laodelphax striatellus
Nilaparvata lugens
Drosophila melanogaster
Anopheles gambiae
Chilo suppressalis
Bombyx mori
Danaus plexippus
Tribolium castaneum
Anolophora glabripennis
Apis mellifera
Nasonia vitripennis
Thrips palmi
Frankliniella occidentalis
Zootermopsis nevadensis
Pediculus humanus
Daphnia pulex
Tetranychus urticae

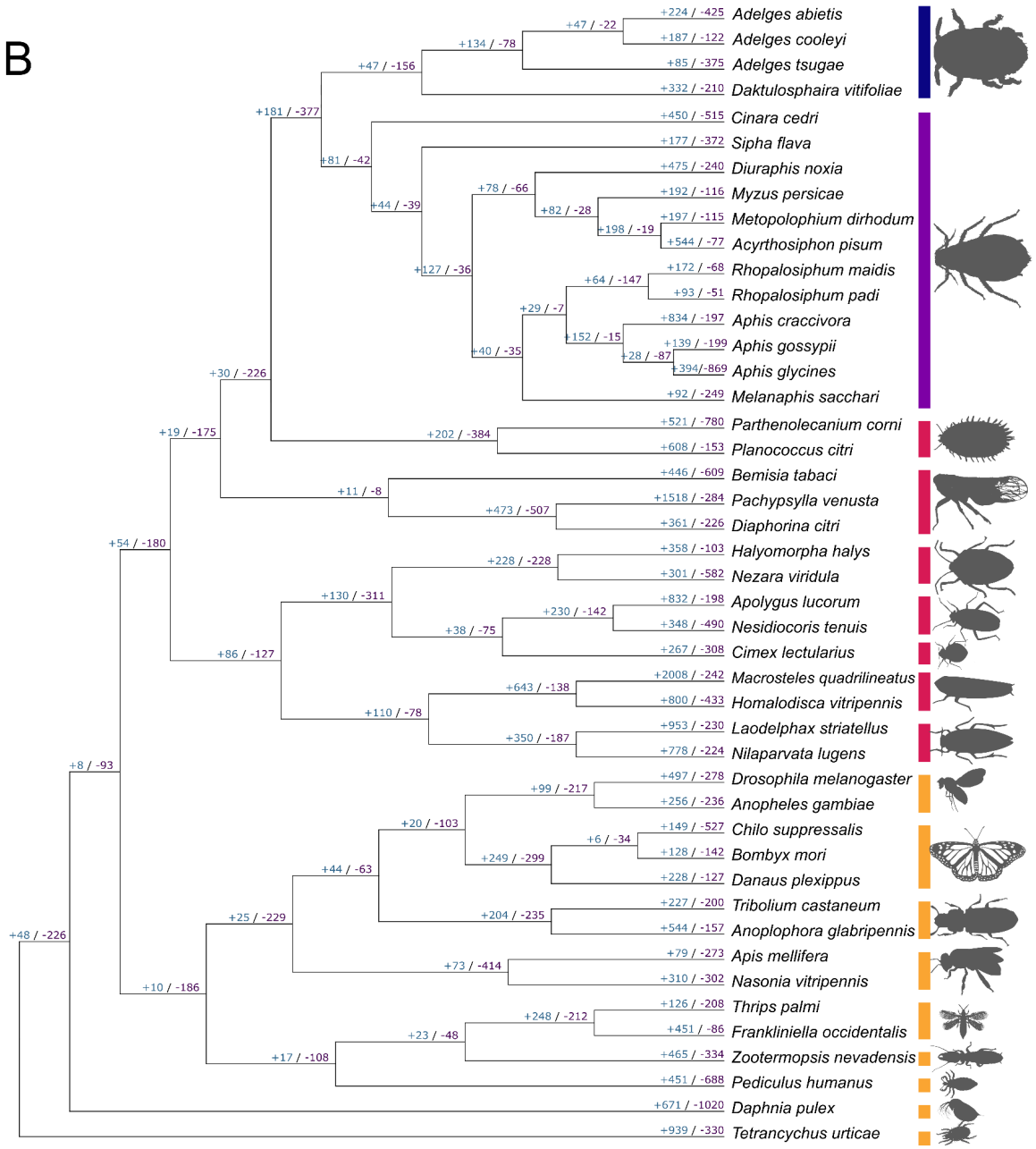


Figure 3. (A) Upset plot of orthologous gene group interactions between 45 species. Minimum interaction size displayed is 50. Rightmost vertical bars represent lineages, and are grouped from top to bottom and darkest to lightest: Phylloxeroidea, Aphididae, other Hemiptera, and non-Hemipteran arthropods.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

Figure 4

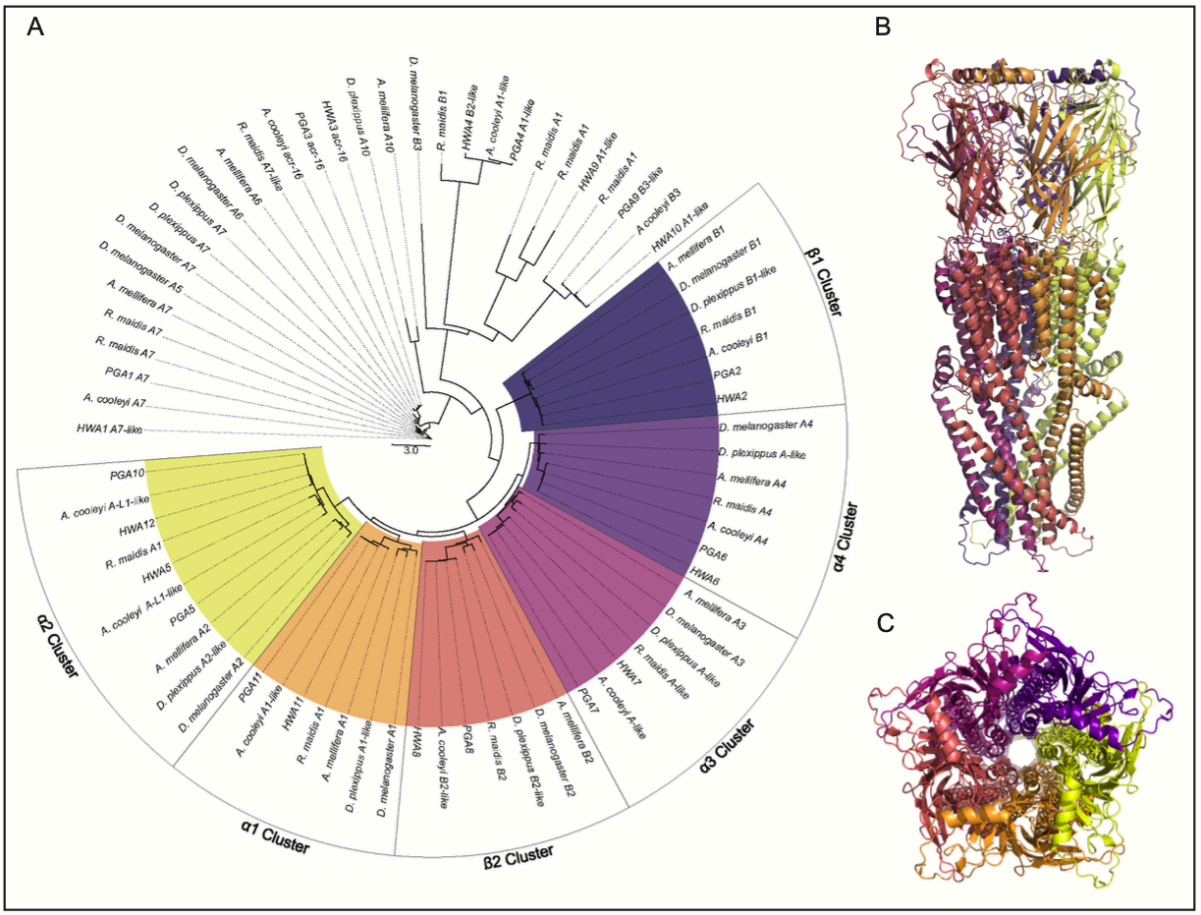


Figure 4. Protein phylogeny of nAChR subtypes and representative AF3 nAChR models. (A) Protein phylogeny of insect nAChRs from selected species. Areas where there is clear clustering by nAChR subtype are highlighted and labeled with their subtype designation. Trees with full gene names and full set clustering available in Figure S1a-b. (B) Side view of AF3 modeled HWA3 nAChR. (C) Top-down view of AF3 modeled HWA3 nAChR.

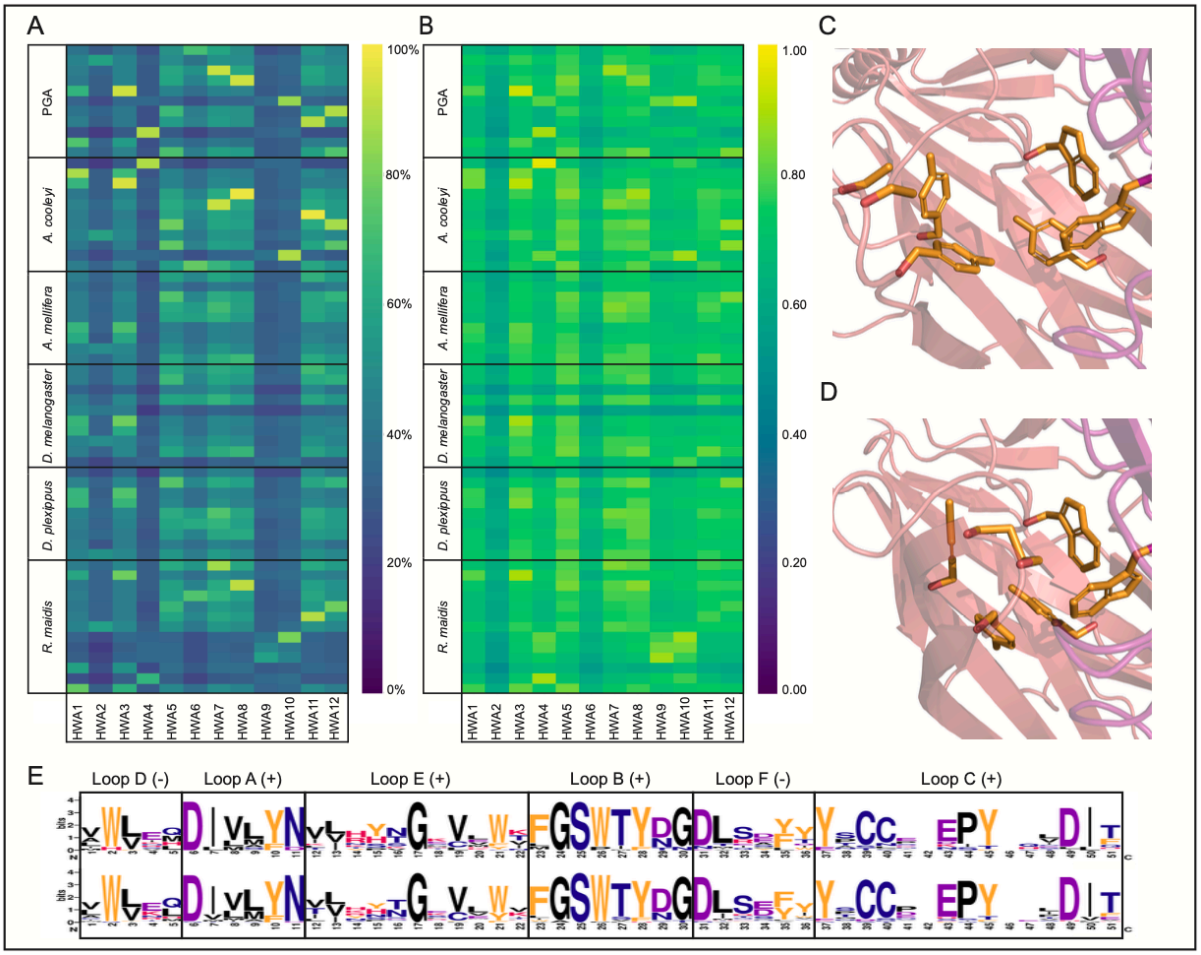


Figure 5. Sequence and 3D comparisons of full nAChR sequences and ligand binding sites. (A) Percent sequence similarity heat map of *A. tsugae* (HWA) nAChR protein sequences versus other selected insect nAChRs. (B) TM score heat map of AlphaFold3 generated *A. tsugae* nAChR protein structures versus other selected AlphaFold3 generated insect nAChRs. (C) AlphaFold3 generated ligand binding site of HWA5 nAChR pentamer, open configuration. (D) Ligand binding site of pentameric Human alpha 2 nAChR crystal structure, closed configuration (PDB: 5FJV). (E) Logo plot visualizations of ligand binding site sequences. Ligand binding site consensus sequences for identified nAChRs from *Adelges tsugae* (HWA)*, Adelges abietis* (PGA), and *Adelges cooleyi* are shown in one logo plot (top), while the ligand binding site consensus sequences of nAChRs from other selected species (*Rhopalosiphum maidis, Apis mellifera, Danaus plexippus*, and *Drosophila melanogaster*) are shown in another logo plot (bottom). The corresponding loops are labeled along with their presence on the principal (+) binding face or complementary (–) binding face.

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 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. Nathan Havill of the US Forest Service and Brian 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 J, Adler J, Dunger J, Evans R, Green T, Pritzel A, Ronneberger O, Willmore L, Ballard AJ, Bambrick J, et al. 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature. 630(8016):493–500.

Alonge M, Lebeigle L, Kirsche M, Jenike K, Ou S, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. 2022. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. Genome Biol. 23(1):258.

Bailey TL, Johnson J, Grant CE, Noble WS. 2015. The MEME suite. Nucleic Acids Res. 43(W1):W39–49.

Baril T, Pym A, Bass C, Hayward A. 2023. Transposon accumulation at xenobiotic gene family loci in aphids. Genome Res. 33(10):1718–1733.

Barington L, Rummel PC, Lückmann M, Pihl H, Larsen O, Daugvilaite V, Johnsen AH, Frimurer TM, Karlshøj S, Rosenkilde MM. 2016. Role of conserved disulfide bridges and aromatic residues in extracellular loop 2 of chemokine receptor CCR8 for chemokine and small molecule binding. J Biol Chem. 291(31):16208–16220.

Boulain H, Legeai F, Guy E, Morlière S, Douglas NE, Oh J, Murugan M, Smith M, Jaquiéry J, Peccoud J, et al. 2018. Fast evolution and lineage-specific gene family expansions of aphid salivary effectors driven by interactions with host-plants. Genome Biol Evol. 10(6):1554–1572.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics. 10(1):421.

Carter C. 1971. Conifer woolly aphids (Adelgidae) in Britain. doi:10.5555/19716605345. https://www.cabidigitallibrary.org/doi/full/10.5555/19716605345.

Cartereau A, Taillebois E, Selvam B, Martin C, Graton J, Le Questel J-Y, Thany SH. 2020. Cloning and expression of cockroach α7 nicotinic acetylcholine receptor subunit. Front Physiol. 11:418.

Cembrowski MS, Phillips MG, DiLisio SF, Shields BC, Winnubst J, Chandrashekar J, Bas E, Spruston N. 2018. Dissociable structural and functional hippocampal outputs via distinct subiculum cell classes. Cell. 173(5):1280–1292.e18.

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

Chen C-Y, Liu Y-Q, Song W-M, Chen D-Y, Chen F-Y, Chen X-Y, Chen Z-W, Ge S-X, Wang C-Z, Zhan S, et al. 2019. An effector from cotton bollworm oral secretion impairs host plant defense signaling. Proc Natl Acad Sci U S A. 116(28):14331–14338.

Cornelissen T, Wilson Fernandes G, Vasconcellos-Neto J. 2008. Size does matter: variation in herbivory between and within plants and the plant vigor hypothesis. Oikos. 117(8):1121–1130.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res. 14(6):1188–1190.

Dancewicz K, Gabryś B, Morkunas I, Samardakiewicz S. 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 W, Rademakers R. 2023. NanoPack2: population-scale evaluation of long-read sequencing data. Bioinformatics. 39(5). doi:10.1093/bioinformatics/btad311. http://dx.doi.org/10.1093/bioinformatics/btad311.

Dial DT, Weglarz KM, Brunet BMT, Havill NP, von Dohlen CD, Burke GR. 2023. Whole-genome sequence of the Cooley spruce gall adelgid, Adelges cooleyi (Hemiptera: Sternorrhyncha: Adelgidae). G3 (Bethesda). 14(1). doi:10.1093/g3journal/jkad224. http://dx.doi.org/10.1093/g3journal/jkad224.

Dohlen CDVON, Moran NA. 2000. Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. Biol J Linn Soc Lond. 71(4):689–717.

von Dohlen CD, Spaulding U, Shields K, Havill NP, Rosa C, Hoover K. 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. Environ Microbiol. 15(7):2043–2062.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. 2017. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science. 356(6333):92–95.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. 2016. Juicebox provides a visualization system for hi-C contact maps with unlimited zoom. Cell Syst. 3(1):99–101.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer provides a one-click system for analyzing loop-resolution hi-C experiments. Cell Syst. 3(1):95–98.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.

Ellison AM, Orwig DA, Fitzpatrick MC, Preisser EL. 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). doi:10.3390/insects9040172. http://dx.doi.org/10.3390/insects9040172.

Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for

comparative genomics. Genome Biol. 20(1):238.

Favret C, Havill NP, Miller GL, Sano M, Victor B. 2015. Catalog of the adelgids of the world (Hemiptera, Adelgidae). Zookeys. 534(534):35–54.

Fei S, Morin RS, Oswalt CM, Liebhold AM. 2019. Biomass losses resulting from insect and disease invasions in US forests. Proc Natl Acad Sci U S A. 116(35):17371–17376.

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

Gharpure A, Noviello CM, Hibbs RE. 2020. Progress in nicotinic receptor structural biology. Neuropharmacology. 171(108086):108086.

Grice SJ, Liu J-L. 2011. Survival motor neuron protein regulates stem cell division, proliferation, and differentiation in Drosophila. PLoS Genet. 7(4):e1002030.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29(8):1072–1075.

Hart AJ, Ginzburg S, Xu MS, Fisher CR, Rahmatpour N, Mitton JB, Paul R, Wegrzyn JL. 2020. EnTAP: Bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. Mol Ecol Resour. 20(2):591–604.

Havill NP, Foottit RG. 2007. Biology and evolution of adelgidae. Annu Rev Entomol. 52:325–349.

Havill NP, Foottit RG, von Dohlen CD. 2007. Evolution of host specialization in the Adelgidae (Insecta: Hemiptera) inferred from molecular phylogenetics. Mol Phylogenet Evol. 44(1):357–370.

Havill NP, Montgomery ME, Yu G, Shiyake S, Caccone A. 2006. Mitochondrial DNA from Hemlock Woolly Adelgid (Hemiptera: Adelgidae) Suggests Cryptic Speciation and Pinpoints the Source of the Introduction to Eastern North America. esaa. 99(2):195–203.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47(D1):D309–D314.

Jiménez-Pompa A, Arribas RL, McIntosh JM, Albillos A. 2023. Differential tyrosine and serine/threonine phosphorylation/dephosphorylation pathways regulate the expression of α7 versus α3β4 nicotinic receptor subtypes in mouse hippocampal neurons. Biochem Biophys Res Commun. 684(149115):149115.

Jones AK, Brown LA, Sattelle DB. 2007. Insect nicotinic acetylcholine receptor gene families: from genetic model organism to vector, pest and beneficial species. Invert Neurosci. 7(1):67–73.

Joseph SV, Kristine Braman S, Quick J, Hanula JL. 2011. The Range and Response of Neonicotinoids on Hemlock Woolly Adelgid, Adelges tsugae (Hemiptera: Adelgidae)1. J Environ Hort. 29(4):197–204.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 14(6):587–589.

Katoh K, Toh H. 2008. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform. 9(4):286–298.

Kim D, Song L, Breitwieser FP, Salzberg SL. 2016. Centrifuge: rapid and sensitive classification of metagenomic sequences. Genome Res. 26(12):1721–1729.

Kinahan IG, Grandstaff G, Russell A, Rigsby CM, Casagrande RA, Preisser EL. 2020. A Four-Year, Seven-State Reforestation Trial with Eastern Hemlocks (Tsuga canadensis) Resistant to Hemlock Woolly Adelgid (Adelges tsugae). For Trees Livelihoods. 11(3):312.

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

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27(5):722–736.

Kumar S, Jones M, Koutsovoulos G, Clarke M, Blaxter M. 2013. Blobology: exploring raw genome data for contaminants, symbionts and parasites using taxon-annotated GC-coverage plots. Front Genet. 4:237.

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 35(6):1547–1549.

Kuznetsov D, Tegenfeldt F, Manni M, Seppey M, Berkeley M, Kriventseva EV, Zdobnov EM. 2023. OrthoDB v11: annotation of orthologs in the widest sampling of organismal diversity. Nucleic Acids Res. 51(D1):D445–D451.

Laetsch DR, Blaxter ML. 2017. BlobTools: Interrogation of genome assemblies. F1000Res. 6(1287):1287.

Letunic I, Bork P. 2024. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. Nucleic Acids Res. 52(W1):W78–W82.

Limbu S, Keena MA, Whitmore MC. 2018. Hemlock woolly adelgid (Hemiptera: Adelgidae): A non-native pest of hemlocks in eastern north America. J Integr Pest Manag. 9(1):27.

Little E, Luther LA. 1971. Atlas of United States Trees:(no. 1146). Conifers and important hardwoods. Little EL, editor. Jr. 1(1146).

Lovell JT, Sreedasyam A, Schranz ME, Wilson M, Carlson JW, Harkess A, Emms D, Goodstein DM, Schmutz J. 2022. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. Elife. 11. doi:10.7554/eLife.78526. [accessed 2024 Nov 4]. https://elifesciences.org/articles/78526.

Maccallini P, Bavasso F, Scatolini L, Bucciarelli E, Noviello G, Lisi V, Palumbo V, D'Angeli S, Cacchione S, Cenci G, et al. 2020. Intimate functional interactions between TGS1 and the Smn complex revealed by an analysis of the Drosophila eye development. PLoS Genet. 16(5):e1008815.

Makhnovskii DA, Murzina GB, Tretyakova MS, Pivovarov AS. 2013. The role of Serine/threonine and tyrosine protein kinases in depression of the cholinosensitivity of neurons in the common snail in a cellular analog of habituation. Neurosci Behav Physiol. 43(1):22–33.

Manni M, Berkeley MR, Seppey M, Zdobnov EM. 2021. BUSCO: Assessing Genomic Data Quality and Beyond. Curr Protoc. 1(12):e323.

Mao M, Simmonds TJ, Stouthamer CM, Kehoe TM, Geib SM, Burke GR. 2023. A chromosome scale assembly of the parasitoid wasp Venturia canescens provides insight into the process of virus domestication. G3 (Bethesda). 13(10):jkad137.

Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. 2017. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. Bioinformatics. 33(4):574–576.

Mathers TC, Chen Y, Kaithakottil G, Legeai F, Mugford ST, Baa-Puyoulet P, Bretaudeau A, Clavijo B, Colella S, Collin O, et al. 2017. Rapid transcriptional plasticity of duplicated gene clusters enables a clonally reproducing aphid to colonise diverse plant species. Genome Biol. 18(1):27.

Mathers TC, Wouters RHM, Mugford ST, Swarbreck D, van Oosterhout C, Hogenhout SA. 2021. Chromosome-scale genome assemblies of aphids reveal extensively rearranged autosomes and long-term conservation of the X chromosome. Mol Biol Evol. 38(3):856–875.

Matsuda K, Buckingham SD, Kleier D, Rauh JJ, Grauso M, Sattelle DB. 2001. Neonicotinoids: insecticides acting on insect nicotinic acetylcholine receptors. Trends Pharmacol Sci. 22(11):573–580.

Mayfield AE III, Bittner TD, Dietschler NJ, Elkinton JS, Havill NP, Keena MA, Mausel DL, Rhea JR, Salom SM, Whitmore MC. 2023. Biological control of hemlock woolly adelgid in North America: History, status, and outlook. Biol Control. 185(105308):105308.

McClure MS, Cheah CAS-J. 1999. Reshaping the Ecology of Invading Populations of Hemlock Woolly Adelgid, Adelges tsugae (Homoptera: Adelgidae), in Eastern North America. Biol Invasions. 1(2/3):247–254.

Mendes FK, Vanderpool D, Fulton B, Hahn MW. 2021. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics. 36(22-23):5516–5518.

Meng X, Zhang Y, Bao H, Liu Z. 2015. Sequence analysis of insecticide action and detoxification-related genes in the insect pest natural enemy Pardosa pseudoannulata. PLoS One. 10(4):e0125242.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 37(5):1530–1534.

Montgomery ME, Bentz SE, Olsen RT. 2009. Evaluation of hemlock (Tsuga) species and hybrids for resistance to Adelges tsugae (Hemiptera: Adelgidae) using artificial infestation. J Econ Entomol. 102(3):1247–1254.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 32(1):268–274.

Orwig DA, Foster DR. 1998. Forest response to the introduced hemlock woolly adelgid in southern New England, USA. J Torrey Bot Soc. 125(1):60.

Petersen M, Armisén D, Gibbs RA, Hering L, Khila A, Mayer G, Richards S, Niehuis O, Misof B. 2019. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. BMC Evol Biol. 19(1):11.

Pilichowski S, Giertych MJ. 2018. Does Hartigiola annulipes (Diptera: Cecidomyiidae) distribute its galls randomly? Eur J Entomol. 115(1):504–511.

P McCarty E, Addesso KM. 2019. Hemlock woolly adelgid (Hemiptera: Adelgidae) management in forest, landscape, and nursery production. J Insect Sci. 19(2):iez031.

Rajendra TK, Gonsalvez GB, Walker MP, Shpargel KB, Salz HK, Matera AG. 2007. A Drosophila melanogaster model of spinal muscular atrophy reveals a function for SMN in striated muscle. J Cell Biol. 176(6):831–841.

Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 21(1):245.

Rider SD Jr, Srinivasan DG, Hilgarth RS. 2010. Chromatin-remodelling proteins of the pea aphid, Acyrthosiphon pisum (Harris): Aphid chromatin genes. Insect Mol Biol. 19 Suppl 2(0 2):201–214.

Rispe C, Legeai F, Nabity PD, Fernández R, Arora AK, Baa-Puyoulet P, Banfill CR, Bao L, Barberà M, Bouallègue M, et al. 2020. The genome sequence of the grape phylloxera provides insights into the evolution, adaptation, and invasion routes of an iconic pest. BMC Biol. 18(1):90.

Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics. 19(1):460.

Ruzzante L, Feron R, Reijnders MJMF, Thiébaut A, Waterhouse RM. 2022. Functional constraints on insect immune system components govern their evolutionary trajectories. Mol Biol Evol. 39(1):msab352.

Sanders M, Tardani R, Locher A, Geller K, Partridge CG. 2023. Development of novel early detection technology for hemlock woolly adelgid, Adelges tsugae (Hemiptera: Adelgidae). J Econ Entomol. 116(1):168–180.

Sano M, Ozaki K. 2012. Variation and evolution of the complex life cycle in Adelgidae (Hemiptera): Complex life cycle of adelgids. Entomol Sci. 15(1):13–22.

Shen W, Sipos B, Zhao L. 2024. SeqKit2: A Swiss army knife for sequence and alignment processing. Imeta. 3(3):e191.

Shimada S, Kamiya M, Shigetou S, Tomiyama K, Komori Y, Magara L, Ihara M, Matsuda K. 2020. The mechanism of loop C-neonicotinoid interactions at insect nicotinic acetylcholine receptor α1 subunit predicts resistance emergence in pests. Sci Rep. 10(1):7529.

Shinoda T, Itoyama K. 2003. Juvenile hormone acid methyltransferase: a key regulatory enzyme for insect metamorphosis. Proc Natl Acad Sci U S A. 100(21):11986–11991.

Smit, AFA, Hubley, R & Green, P. 2013-2015. RepeatMasker Open-4.0. RepearMasker. http://www.repeatmasker.org.

Steffan AW. 1968. Evolution und Systematik der Adelgidae (Homoptera: Aphidina) : Eine Verwandtschaftsanalyse auf vorwiegend ethologischer, zytologischer und karyologischer Grundlage.

Stothard P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques. 28(6):1102, 1104.

Suchail S, Guez D, Belzunces LP. 2001. Discrepancy between acute and chronic toxicity induced by imidacloprid and its metabolites in Apis mellifera. Environ Toxicol Chem. 20(11):2482.

Szabó G, Schulz F, Manzano-Marín A, Toenshoff ER, Horn M. 2022. Evolutionarily recent dual obligatory symbiosis among adelgids indicates a transition between fungus- and insect-associated lifestyles. ISME J. 16(1):247–256.

Toshima K, Kanaoka S, Yamada A, Tarumoto K, Akamatsu M, Sattelle DB, Matsuda K. 2009. Combined roles of loops C and D in the interactions of a neonicotinoid insecticide imidacloprid with the alpha4beta2 nicotinic acetylcholine receptor. Neuropharmacology. 56(1):264–272.

Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. arXiv [csDC]. http://arxiv.org/abs/1907.12931.

Wang H, Liu B, Zhang Y, Jiang F, Ren Y, Yin L, Liu H, Wang S, Fan W. 2020.

Estimation of genome size using k-mer frequencies from corrected long reads. arXiv [q-bioGN]. [accessed 2024 Oct 31]. http://arxiv.org/abs/2003.11817.

Wang J, Zhang Z, Yu N, Wu X, Guo Z, Yan Y, Liu Z. 2024. Cys-loop ligand-gated ion channel superfamily of Pardosa pseudoannulata: Implication for natural enemy safety. Comp Biochem Physiol Part D Genomics Proteomics. 49(101190):101190.

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee T-H, Jin H, Marler B, Guo H, et al. 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. 40(7):e49.

Weglarz KM, Havill NP, Burke GR, von Dohlen CD. 2018. Partnering with a pest: Genomes of hemlock woolly adelgid symbionts reveal atypical nutritional provisioning patterns in dual-obligate bacteria. Genome Biol Evol. 10(6):1607–1621.

Wiesner A, Fuhrer C. 2006. Regulation of nicotinic acetylcholine receptors by tyrosine kinases in the peripheral and central nervous system: same players, different roles. Cell Mol Life Sci. 63(23):2818–2828.

Wu C, Chakrabarty S, Jin M, Liu K, Xiao Y. 2019. Insect ATP-binding cassette (ABC) transporters: Roles in xenobiotic detoxification and Bt insecticidal activity. Int J Mol Sci. 20(11):2829.

Wu Y-T, Ma R, Wei J-W, Song L-W, Dewer Y, Wang S-S, Liu L, Zhou J-J. 2024. ApCarE4 and ApPOD3 participate in the adaptation of pea aphids to different alfalfa varieties. Sci Rep. 14(1):25444.

Xu J, Zhang Y. 2010. How significant is a protein structure similarity with TM-score = 0.5? Bioinformatics. 26(7):889–895.

Yeh H-T, Ko C-C, Wu L-W. 2020. The first complete mitochondrial genome of Adelges tsugae Annand (Hemiptera: Adelgidae). Mitochondrial DNA B Resour. 5(3):2288–2290.

Zhang L, Li Y, Xu X, Feng M, Turak R, Liu X, Pan H. 2024. Functional analysis of AgJHAMT gene related to developmental period in Aphis gossypii Glover. Bull Entomol Res. 114(5):1–10.

Zhang S, Gao X, Wang L, Jiang W, Su H, Jing T, Cui J, Zhang L, Yang Y. 2022. Chromosome-level genome assemblies of two cotton-melon aphid Aphis gossypii biotypes unveil mechanisms of host adaption. Mol Ecol Resour. 22(3):1120–1134.

Zhang Y, Liu B, Zhang Z, Wang L, Guo H, Li Z, He P, Liu Z, Fang J. 2018. Differential expression of P450 genes and nAChR subunits associated with imidacloprid resistance in Laodelphax striatellus (Hemiptera: Delphacidae). J Econ Entomol. 111(3):1382–1387.

Zhang Y, Skolnick J. 2004. Scoring function for automated assessment of protein structure template quality. Proteins. 57(4):702–710.

Zhang Z, Wood WI. 2003. A profile hidden Markov model for signal peptides

generated by HMMER. Bioinformatics. 19(2):307–308.