

Whole-genome sequence of the Cooley spruce gall adelgid, *Adelges cooleyi* (Hemiptera: Sternorrhyncha: Adelgidae)

Dustin T. Dial ^{1,*} Kathryn M. Weglarz,^{2,5} Bryan M.T. Brunet ³ Nathan P. Havill,⁴ Carol D. von Dohlen,² Gaelen R. Burke¹

¹Department of Entomology, University of Georgia, Athens, GA 30602, USA

²Department of Biology, Utah State University, Logan, UT 84322, USA

³Ottawa Research and Development Centre, Agriculture and Agri-Food Canada, Ottawa, ON, Canada K1A 0C6

⁴USDA Forest Service, Northern Research Station, Hamden, CT 06514, USA

⁵Present address: Biology Department, Westfield State University, Westfield, MA 01086, USA

*Corresponding author: Department of Entomology, University of Georgia, 419 Biological Sciences Building, 120 Cedar St, Athens, GA 30602, USA. Email: dtdial@uga.edu

The adelgids (Adelgidae) are a small family of sap-feeding insects, which, together with true aphids (Aphididae) and phylloxerans (Phylloxeridae), make up the infraorder Aphidomorpha. Some adelgid species are highly destructive to forest ecosystems such as *Adelges tsugae*, *Adelges piceae*, *Adelges laricis*, *Pineus pini*, and *Pineus boemeri*. Despite this, there are no high-quality genomic resources for adelgids, hindering advanced genomic analyses within Adelgidae and among Aphidomorpha. Here, we used PacBio continuous long-read and Illumina RNA-sequencing to construct a high-quality draft genome assembly for the Cooley spruce gall adelgid, *Adelges cooleyi* (Gillette), a gall-forming species endemic to North America. The assembled genome is 270.2 Mb in total size and has scaffold and contig N50 statistics of 14.87 and 7.18 Mb, respectively. There are 24,967 predicted coding sequences, and the assembly completeness is estimated at 98.1 and 99.6% with core BUSCO gene sets of Arthropoda and Hemiptera, respectively. Phylogenomic analysis using the *A. cooleyi* genome, 3 publicly available adelgid transcriptomes, 4 phylloxera transcriptomes, the *Daktulosphaira vitifoliae* (grape phylloxera) genome, 4 aphid genomes, and 2 outgroup coccoid genomes fully resolves adelgids and phylloxerans as sister taxa. The mitochondrial genome is 24 kb, among the largest in insects sampled to date, with 39.4% composed of noncoding regions. This genome assembly is currently the only genome-scale, annotated assembly for adelgids and will be a valuable resource for understanding the ecology and evolution of Aphidomorpha.

Keywords: *Adelges cooleyi*; adelgid; Adelgidae; phylloxera; aphid; Aphidomorpha; genome assembly; mitochondria

Introduction

The Adelgidae, Phylloxeridae, and Aphididae comprise Aphidomorpha, a clade of sap-feeding insects that are cyclically parthenogenetic and exhibit complex, multigenerational, and polyphenic life cycles (Havill and Footitt 2007). The highly host-specific Adelgidae and Phylloxeridae are comparatively species poor, containing about 60–70 species each, whereas Aphididae contains approximately 5,000 species (<http://aphid.speciesfile.org/>). This disparity is likely an outcome of an ancient switch from gymnosperm to angiosperm feeding within Aphididae, and the subsequent rapid radiation of angiosperms (von Dohlen and Moran 2000). In contrast, adelgids retained the comparatively species-poor gymnosperm conifers as hosts, and phylloxerids exploited only a few angiosperm families. Some adelgid species are invasive pests of major economic and ecological importance, such as the hemlock woolly adelgid (*Adelges tsugae*) (Reardon et al. 2011; Havill et al. 2016) and the balsam woolly adelgid (*Adelges piceae*) (Amman 1965; Havill et al. 2021). While several high-quality aphid genomes and one phylloxera genome have been published, there are currently no adelgid genomes, limiting the phylogenetic scope of genomic resources within Aphidomorpha.

The phylogenetic reconstruction of relationships within Adelgidae and among Aphidomorpha has been complicated by

complex polyphenism, a lack of informative morphological characteristics, and a lack of genomic data. The deep divergences among Adelgidae, Phylloxeridae, and Aphididae likely date to the Jurassic (Havill et al. 2007). Adelgids and phylloxerans are distinguished from Aphididae by a lack of cornicles, by reproduction by means of ovipary in every generation, and by a lack of *Buchnera*, the ancestral nutritional endosymbiont of aphids. While obligate nutritional symbioses are absent in the phylloxera, adelgids exhibit a diverse mosaic of symbionts wherein each major host lineage contains a unique pair (von Dohlen et al. 2017). While some studies have suggested that aphids and phylloxerans are sister groups on the basis of morphological characters such as reduced ovipositors (Grimaldi and Engel 2005), most studies support adelgids + phylloxerans as sister to the aphids with fossil data (Heie and Wegierek 2009), mitochondrial and/or ribosomal sequences (von Dohlen and Moran 1995; Havill et al. 2007; Yeh et al. 2020), or phylogenomic datasets (Rispe et al. 2020; Hardy et al. 2022; Owen and Miller 2022). However, these studies did not specifically address the relationship among adelgids, phylloxerans, and aphids, and often contained few taxa or little genomic data within Adelgidae.

Adelges cooleyi (Gillette), the Cooley spruce gall adelgid, is native to western North America and is distributed throughout the Rocky

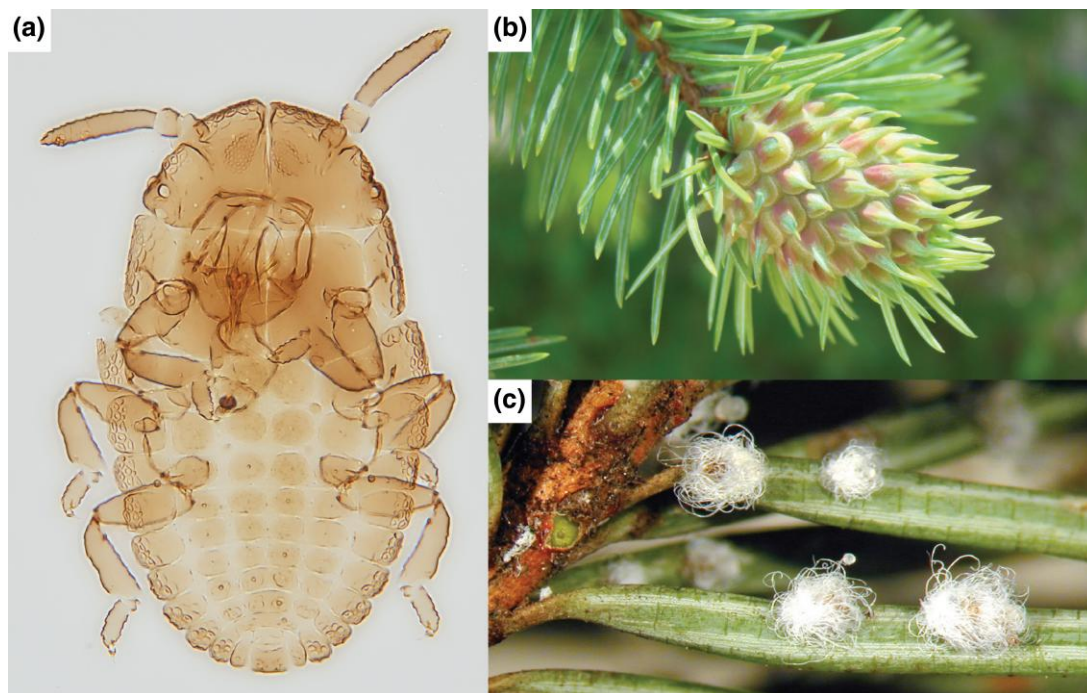


Fig. 1. Ecology of *A. cooleyi*. a) Slide-mounted 1st instar *A. cooleyi* (image credit: N. Havill). b) *A. cooleyi* gall (image credit: D. Dickinson). c) *A. cooleyi* on Douglas fir (image credit: M. Montgomery).

Mountains and Cascade Mountains (Fig. 1a; Gillette 1907). An analysis of 2 mitochondrial genes and amplified fragment length polymorphisms (AFLPs) indicated that there are at least 3 mitochondrial lineages within *A. cooleyi*: southeastern Arizona, Pacific Northwest, and Rocky Mountains (Ahern et al. 2009). AFLPs suggest that populations from Arizona are genetically isolated from the other 2 lineages. *Adelges cooleyi* feeds on spruce (*Picea*) species and Douglas fir (*Pseudotsuga menziesii*). On spruce, the insect induces the formation of large galls at the growing tips of branches and feeds on parenchyma tissue within the gall (Fig. 1b). On Douglas fir, it feeds on the phloem sap of needles, discoloring the needles and causing cosmetic damage (Fig. 1c; Annand 1928). Both trees are required to complete its 2-year life cycle, although cases of populations residing exclusively on spruce have been documented (Annand 1928; Cumming 1962). On spruce, the insects reproduce both sexually and asexually over the course of several generations, and winged asexuals emerge from galls to fly to Douglas fir in the summer (Annand 1928). On Douglas fir, they reproduce asexually, and winged generations return to spruce the following summer, or wingless asexuals can continue to reproduce on Douglas fir. As with the vast majority of other sap-feeding insects (Moran 2001), the plant sap diet of adelgids is supplemented with essential amino acids and vitamins by obligate intracellular bacteria residing within specialized tissues called bacteriomes. The alternation between different dietary sources (i.e. phloem and parenchyma) has been hypothesized to have promoted the multiple replacements of endosymbionts within Adelgidae (von Dohlen et al. 2017; Weglarz et al. 2018; Dial et al. 2022).

Here, we report a high-quality genome sequence of *A. cooleyi*, the first available genome representing Adelgidae. We include both nuclear and mitochondrial genomes. Our *A. cooleyi* genome data were generated alongside our efforts to recover the endosymbiont genomes (Dial et al. 2022). We report on the characteristics of both genomes and perform a phylogenomic analysis to test the

relationships within Aphidomorpha. These publicly available genome data will provide a resource to help address future questions relating to adelgid systematics and evolution and to facilitate comparative analyses with the insects of other groups.

Methods

Insect collection, DNA extraction, and sequencing

Adelgids used for sequencing were collected from Franklin County, ID, USA, on state highway ID-36 in Emigration Canyon in July 2019 and identified by C.D.v.D. Galls containing parthenogenetic females were collected from blue spruce (*Picea coloradensis*) and cut open to reveal insect galleries. Approximately 100 individual whole-body 4th instar individuals were collected from the gall, and DNA was extracted using a DNeasy Blood & Tissue Kit (Qiagen Inc., Germantown, MD, USA). Adelgids within a gall are likely all clonal offspring of a single fundatrix (Havill and Footitt 2007), but intergall migration is possible and could result in occupants belonging to different clones (Ozaki 1995). Size selection was performed by using the BluePippin system (Sage Science, Beverly, MA, USA), yielding DNA fragments of an average size of 29,055 bp. DNA was sequenced using the Pacific Biosciences (PacBio) (Menlo Park, CA, USA) Sequel II System in continuous long-read mode on a single SMRTCell at the Georgia Genomics and Bioinformatics Core.

Transcriptome sequencing

To aid in ab initio gene prediction, RNA sequencing was performed on a total of 4 whole-body *A. cooleyi* samples collected from both spruce (gallicolae) and Douglas fir (sistens). Total RNA was extracted from whole bodies with the Zymo Research DirectZol RNA miniprep kit. Quantification and quality assessment were performed by using an Agilent 2100 Bioanalyzer. Strand-specific libraries were prepared using the Illumina TruSeq Stranded mRNA Sample Prep kit with oligo dT selection and were

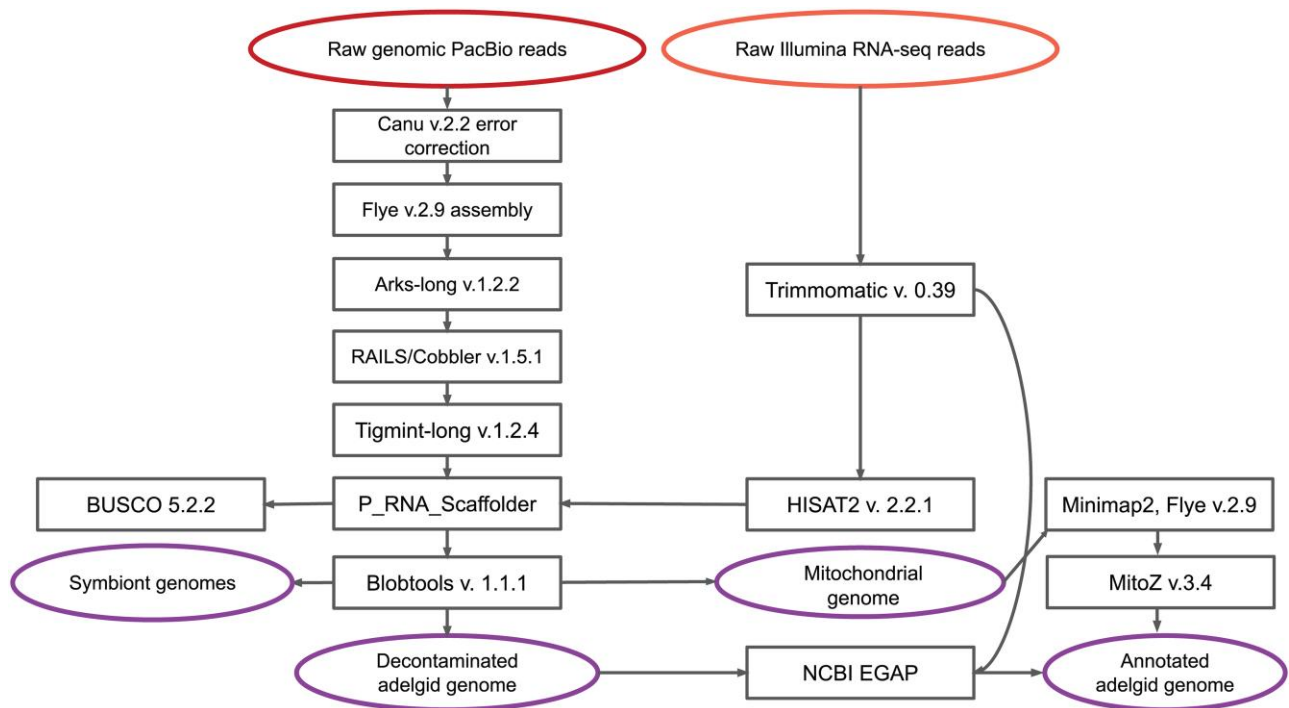


Fig. 2. Summary schematic of tools used for genome assembly and annotation.

sequenced with the HiSeq system with 125 bp paired-end chemistry at the University of Utah Huntsman Cancer Institute.

Genome assembly, contaminant filtering, and annotation

Genome assembly was performed in a multistep process (Fig. 2). First, consensus reads were generated from overlapping raw PacBio reads by correction with Canu v. 2.2 (Koren et al. 2017). The Canu-corrected reads were then assembled with Flye v. 2.9 (Kolmogorov et al. 2019) with the option `-pacbio-corr` (<3% error) and with 5 rounds of genome polishing. Scaffolding was initially performed with ARKS-long (Coombe et al. 2018), with options “`m = 8-10000c = 4l = 4 a = 0.3 k = 20 j = 0.05.`” Rails and Cobbler (Warren 2016) were then used for further scaffolding and gap closing with minimap2 (Li 2018) (options “`250 0.8 250 bp 2 pacbio.`”). The resulting assembly was corrected with Tigmint-long (Jackman et al. 2018) with options “`span = auto G = 272898873 dist = auto longmap = pb.`” Corrected reads were used in all scaffolding and gap-closing steps. Finally, P_RNA_Scaffolder (Zhu et al. 2018) was used to scaffold the genome with quality-trimmed RNA-seq reads aligned to the genome with HISAT2 (Kim et al. 2019). Settings were adjusted such that the minimal number of RNA-seq pairs to join contigs was 5 instead of the default value of 2 for additional stringency.

Contaminants were screened and filtered with BlobTools (Laetsch and Blaxter 2017); corrected PacBio reads were aligned to the genome with minimap2, and open reading frames (ORFs) ≥ 100 bp were called using EMBOSS (Rice et al. 2000) and searched with DIAMOND v. 2.0.2 (Buchfink et al. 2021) against the RefSeq (Pruitt et al. 2007) database modified to include only arthropods, bacteria, archaea, viruses, nematodes, and fungi with TaxonKit (Shen and Ren 2021) for taxonomic assignment. The plots illustrating coverage, GC content, and taxonomy were generated with BlobTools and used to remove all contigs that were not hemipteran (Fig. 3a). DIAMOND search results of contigs assigned to noninsect

groups (i.e. Proteobacteria, Nematoda, Ascomycota, and viruses) were manually checked and searched against NR before exclusion. In total, only 3 symbiont sequences (2 chromosomes and a plasmid), 16 kb contig belonging to Burkholderia, the mitochondrial genome, and contigs <200 bp were removed. BUSCO v. 5.3.2 (Manni et al. 2021) was used to assess genome quality and completeness, with MetaEuk (Levy Karin et al. 2020) used for gene prediction by default (Hemiptera_odb10 database, $n = 2,510$). Contigs with no hit (300 sequences) were retained as they may contain lineage-specific genes. After decontamination, assembly statistics were determined using the stats.sh script from BBmap version 38.98 (Bushnell 2014).

The resulting insect genome was annotated with the NCBI Eukaryotic Genome Annotation Pipeline (EGAP). This pipeline utilized our *A. cooleyi* RNA-seq data for gene prediction, in addition to protein alignments of NCBI RefSeq proteins from *Myzus persicae*, *Diuraphis noxia*, *Halyomorpha halys*, *Acyrtosiphon pisum*, *Bemisia tabaci*, *Drosophila melanogaster*, *Cimex lectularius*, 118,571 Insecta GenBank proteins, and 9,034 Insecta RefSeq proteins. RNA-seq and protein alignments were performed with STAR (Dobin et al. 2013) and ProSplign (<https://www.ncbi.nlm.nih.gov/sutils/static/prosplign/prosplign.html>), respectively, and were used for gene model prediction with GNOMON, the NCBI eukaryotic gene prediction tool (Souvorov et al. 2010). Repeats were masked with WindowMasker (Morgulis et al. 2006), and transcript and protein alignments were performed on the repeat-masked genome. The transcriptome and proteome generated by this pipeline were checked for completeness with BUSCO v4.1.4. The annotation release for *A. cooleyi* can be viewed online at https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Adelges_cooleyi/100/ (GenBank assembly accession GCA_023614345.1).

Mitogenome assembly and annotation

Our assembly screen revealed a noncircular contig with exceptionally high coverage and low GC content (50,776 bp and 14.6% GC). This contig was searched against the NCBI nt database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

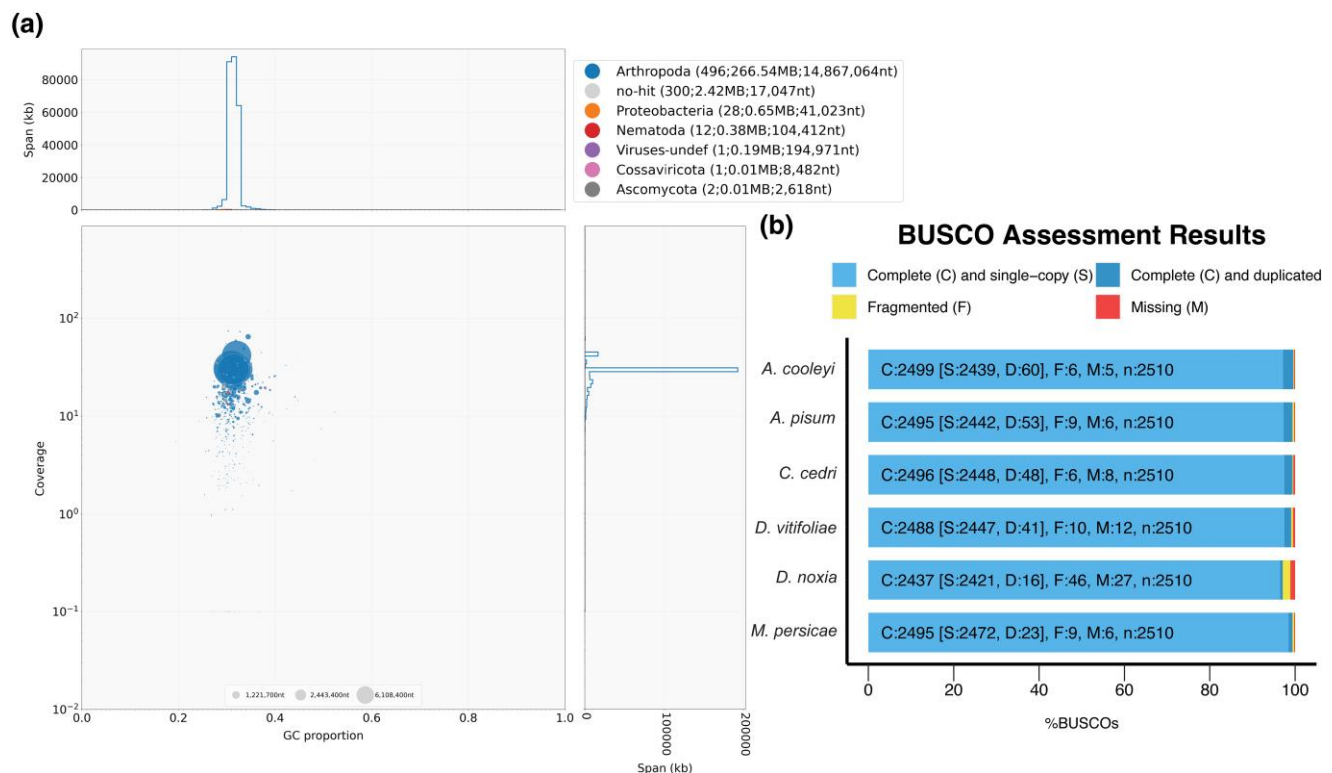


Fig. 3. Quality assessment of the *A. cooleyi* genome assembly. a) A taxon-annotated GC-proportion and coverage plot of the *A. cooleyi* genome. Each blob represents a scaffold and is scaled by length and colored by taxonomy assigned by BlobTools with a custom RefSeq database. The X axis is the average GC content of each scaffold, and the Y axis corresponds to the read coverage based on the alignment of corrected PacBio reads. Most scaffolds assigned as noninsect taxa were revealed to belong to the insect genome upon searching their ORFs against NCBI's NR database. Contigs assigned to the "no-hit" category were not discarded. b) A comparison of the BUSCO results of *A. cooleyi* and several published high-quality aphid genomes using the Hemiptera gene set of 2,510 conserved single-copy genes.

nih.gov/Blast.cgi) with BLASTn (Altschul et al. 1990) and matched to the *A. tsugae* mitochondrial genome (Yeh et al. 2020). Further inspection revealed that this contig contained 2 complete copies of the mitochondrial genome, approximately 24.2 and 26.6 kb in length, differing in the length of their control and repeat regions (4.2–5.6 and 5.2–6.2 kb, respectively). Corrected reads were mapped to this contig with minimap2 (515 reads) and reassembled with Flye v 2.9 in Geneious Prime 2021.1.1 (<https://www.geneious.com>) with default settings, yielding a circular 24 kb contig with 206-fold coverage. The genic content of this contig is nearly identical to that of the mitogenome copies present in the original 50.8 kb scaffold. To ensure that we excluded nuclear mitochondrial reads from the mitochondrial genome assembly, we mapped the mitochondrial reads back to the clean nuclear assembly with minimap2, and no hits were obtained. Furthermore, we employed a strategy that involved mapping all reads to the nuclear genome, extracting and mapping unmapped reads to the mitochondrial genome, and reassembling the reads that were mapped. This produced a 23.9 kb contig highly similar to the 24 kb contig derived earlier. Mitogenome annotations were produced using MitoZ v.3.4 (Meng et al. 2019) and compared with the *A. tsugae* mitogenome. The *trnE* gene was not annotated by MitoZ, but a comparison of this region with *A. tsugae* and related aphid *trnE* genes showed high-sequence similarity and was manually annotated. Tandem Repeat Finder was used to identify tandem repeats in the noncoding regions of the mitochondrial genome (Benson 1999).

Incorporation of publicly available RNA-seq data

To increase taxon representation within Adelgidae and Phylloxeridae in our species tree, we included 3 adelgid and 4 phylloxeran species

that each have an available transcriptome. Adelgid species included *A. tsugae* (SRR1198669), *Pineus* sp. pispAD (SRR5134714), and *Adelges* sp. AdspAS (SRR5134707). While the latter 2 taxa were not identified to the species level prior to sequencing, the extraction of full-length COI genes from assembled transcripts showed 99.85 and 100% sequence identity to *Pineus strobi* and *Adelges abietis*, respectively. Phylloxeran species included *Phylloxera nyssae* (SRR23289299), *Phylloxera* sp. PhspAP35 (SRR5134737), *Phylloxera* sp. PE11a (SRR23290230), and *Phylloxera* sp. H7a (SRR23290233). Raw RNA-seq reads for these species were retrieved from the Sequence Read Archive (SRA) by using the SRA-toolkit v. 3.0.1, assembled with Trinity v. 2.8.5 (Grabherr et al. 2011), and set to trim the sequences automatically with Trimmomatic (Bolger et al. 2014). The longest isoform of each transcript was selected, and the prediction of proteins was performed by TransDecoder (<https://github.com/TransDecoder/TransDecoder/>) and set to output ORFs > 100 amino acids. The resulting ORFs were searched with BLASTp (e-value cutoff $1e^{-5}$) against a database of 18 species within the Aphidomorpha and 2 coccoid proteomes (Supplementary Table 1), and ORFs with no hits were removed with the TransDecoder Predict function. BUSCO v. 5.3.2 was used to assess the completeness of each quality-filtered transcriptome (Supplementary Table 2).

Orthology inference and species phylogeny

We performed orthology clustering of predicted protein sequences from the *A. cooleyi* genome, the 3 adelgid transcriptomes, and 4 phylloxeran transcriptomes described above; from the genome sequences of *Daktulosphaira vitifoliae* (grape phylloxera) (Rispe et al. 2020), aphid species *Cinara cedri* (Julca et al. 2020), *Aphis*

Table 1. Summary statistics of *A. cooleyi* RNA-seq reads used for gene prediction.

Sequence Read Archive accession	Material	Read pairs	Number of bases (Gb)	Percent aligned reads
All	Aggregate of all aligned samples	237,817,810	59.4	92
SRR21113675	Whole-body 4th instar collected from Douglas fir	54,998,082	13.7	92
SRR21113676	Whole-body 4th instar collected from spruce	54,421,594	13.6	93
SRR21113677	Whole-body 4th instar collected from spruce	60,718,604	15.2	90
SRR21113678	Whole-body 4th instar collected from Douglas fir	67,679,530	16.9	92

Table 2. *Adelges cooleyi* genome assembly statistics.

Total assembly size (Mb)	270.2
Number of contigs	962
Contig N50 (Mb)	7.18
Contig L50	11
Contig L90	106
Max contig length (Mb)	21.44
Number of scaffolds	840
Scaffold N50 (Mb)	14.87
Scaffold L50	8
Scaffold L90	59
Max scaffold length (Mb)	24.43
Number of scaffolds >50 kb	194
Main genome in scaffolds >50 kb	96.62%

glycines (Mathers 2020), *A. pisum* (Li et al. 2019), and *M. persicae* (Mathers et al. 2017); and from 2 coccoid species, *Ericerus pela* (Yang et al. 2019) and *Phenacoccus solenopsis* (Li et al. 2020b) with OrthoFinder v2.5.2 (Emms and Kelly 2019). The longest isoform of each annotated gene was extracted and used as an input into OrthoFinder. We ran OrthoFinder in multiple sequence alignment mode (-M msa), and alignment and tree inference were performed within OrthoFinder with MAFFT (Kato and Standley 2013) and FastTree (Price et al. 2010). This method estimates a species tree based on a concatenated alignment of all single-copy orthogroups present in all species. Alignments were lightly trimmed by OrthoFinder with default settings. To confirm the tree topology generated by OrthoFinder's utilization of FastTree, the concatenated alignment was used as input into IQ-Tree with automatic model selection with ModelFinder (JTT + F + R4) (Kalyaanamoorthy et al. 2017) and 1,000 ultrafast bootstraps. Additionally, we ran OrthoFinder in its default species tree inference mode, which uses STAG (Emms and Kelly 2018) for species tree inference. This method creates a consensus species tree from the most closely related genes within both single-copy and multicopy orthogroups. Trees were rooted by STRIDE (Emms and Kelly 2017) using gene duplication information.

Results and discussion

Adelges cooleyi genome characteristics

We generated 84.26 Gb (2,729,794 reads) of long-read PacBio data from a pool of clonal *A. cooleyi* individuals collected from a single gall and a total of 59.4 Gb of RNA-seq data from *A. cooleyi* sampled from spruce and Douglas fir hosts (Table 1). The error correction of PacBio reads with Canu generated 8.80 Gb of data (229,338 reads, N50/N90 54,907/42,252), and these reads were used to de novo assemble the *A. cooleyi* genome. The final assembled, scaffolded, corrected, and contaminant-filtered genome had an average coverage of 30.0x (calculated from corrected consensus reads), was 270.2 Mb in total length, and contained 962 contigs and 840 scaffolds (Table 2). The genomic GC content was low (31.4%) but comparable with other aphid genomes. The genome is highly

contiguous, with contig and scaffold N50 statistics of 7.18 and 14.87 Mb, respectively. The *A. cooleyi* genome is contained within $2n = 22$ chromosomes (Steffan 1968), and 90% or more of the current assembly is contained within 59 scaffolds (L90 = 59) (Table 2). The prefiltered assembly possessed a low degree of contamination; most of the contigs identified by BlobTools as potential contaminants were either symbiont sequences or contained insect genes as verified by BLAST searches against NR (Fig. 3a). These contigs containing insect genes were retained in the final assembly. Although not assembled to the level of chromosomes, this genome is highly complete, with an excellent representation of conserved hemipteran genes, which meets or exceeds that of published genomes of related species (Fig. 3b, Table 3). Our assembly pipeline utilized error-corrected consensus reads from high-coverage PacBio data and likely produced an assembly with a low error rate (Dohm et al. 2020). However, we note that the lack of high-accuracy short-read data inhibits us from obtaining higher-level quality metrics such as kmer spectra or Merqury QV statistics (Rhie et al. 2020). The 2 endosymbiont species *Ca. Vallotia cooleyia* and *Ca. Gillettella cooleyia* were described previously with the same long-read dataset (Dial et al. 2022). WindowMasker repeat-masked 43.67% of the genome, which falls within the range of masked content in related insect genomes. Gene prediction with NCBI EGAP produced 24,967 coding sequences (CDSs) and 27,577 transcripts (Table 4).

Adelges cooleyi has large mitochondrial control and repeat regions and a gene order identical to the ancestral Insecta

Our initial assembly yielded a 50.8 kb linear concatemer containing 2 copies of the mitochondrial genome, each with complete gene sets but differing in the length of the control and repeat regions. These noncoding regions share 98% nucleotide identity between copies but differ in length. Concatemers are common when read length exceeds the length of the sequence being assembled (Hunt et al. 2015), and the length variation of noncoding regions in the mitogenome copies may be attributed to the presence of different clones within the same gall or heteroplasmy within a single clone. Upon mapping corrected reads to this contig and reassembling these isolated mitochondrial reads, we obtained a final circular, 24,090 bp contiguous sequence with even 206.0x coverage and 85.2% AT content (Fig. 4). This genome size is consistent with the length of mitogenome sequences as seen in individual reads. The gene content and order are identical to the ancestral insect (Cameron 2014), with 13 protein-coding genes, 22 tRNA genes, and 2 rRNA genes, with 83.7% genic AT content. Notably, the *A. cooleyi* mitochondrial genome possesses a long 5.14 kb repeat region and a 4.31 kb control region with AT contents of 86.1 and 88.4%, respectively. Consequently, 39.4% of the mitochondrial genome is composed of noncoding DNA. These regions in the final 24 kb mitogenome are most similar to the first copy in the 50.8 kb contig. These noncoding regions in *A. cooleyi* are much longer than in the *A. tsugae* mitogenome, but the genic

Table 3. Comparison of genome statistics between *A. cooleyi* and related species.

Species	Reference	Contig count (N50 Mb)	Scaffold count (N50 Mb)	Total length (Mb)	CDS count
<i>Adelges cooleyi</i>	This study	962 (7.18)	840 (14.87)	270.20	24,967
<i>Daktulosphaira vitifoliae</i>	GCF_025091365.1	17,226 (0.72)	8,544 (45.89)	282.59	29,510
<i>Cinara cedri</i>	Julca et al. (2020)	6,160 (0.28)	2,842 (1.23)	396.03	22,503
<i>Aphis glycines</i>	Mathers (2020)	1,024 (2.0)	941 (2.51)	303.15	19,750
<i>Rhopalosiphum maidis</i> ^a	Chen et al. (2019)	960 (9.05)	220 (93.3)	326.02	17,629
<i>Sitobion miscanthi</i> ^a	Jiang et al. (2019)	1,148 (2.05)	656 (36.26)	397.9	16,006
<i>Diuraphis noxia</i>	Nicholson et al. (2015)	49,379 (0.01)	5,641 (0.40)	393.0	19,097
<i>Myzus persicae</i> ^a	Mathers et al. (2021)	915 (4.17)	360 (69.5)	395.14	27,663
<i>Acyrthosiphon pisum</i> ^a	Mathers et al. (2021)	2,298 (0.53)	558 (126.6)	525.80	30,784

^a Scaffolds have been assigned to chromosomes.

sequences of these assemblies share high sequence identity and the same gene order. While the overall size of the *A. cooleyi* mitogenome is significantly greater than the size of a typical insect mitogenome, it is not exceptionally large. For example, unusually large mitogenomes up to 36 kb have been described in weevils, up to 28 kb in Hercules beetles, and up to 26 kb in seed beetles due to large, often heteroplasmic control regions (Boyce et al. 1989; Sayadi et al. 2017; Morgan et al. 2022).

The control region is thought to function as a binding site for proteins involved in replication and transcription (Saito et al. 2005). Reconstructing the control region is challenging because of the presence of tandem repeat units that are often larger than the read length generated by short-read technologies. While often only partial or missing in public databases (Cameron 2014), it can vary in size from a few hundred to several thousand bases within the Hemiptera (Li and Liang 2018) and can be as long as 13.7 kb in the Coleoptera (Morgan et al. 2022). Within the Cicadidae, polymerase chain reaction-amplified control regions can range between 1.7 and 5 kb (Lukasik et al. 2019). The application of long-read sequencing to the study of mitogenomes has recently revealed large noncoding regions that were previously unassembled with short-read data in insects (Sayadi et al. 2017; Xu et al. 2018; Filipovic et al. 2021; Morgan et al. 2022) and other animals (Gan et al. 2019; Kinkar et al. 2020; Novosolov et al. 2022; Sharbrough et al. 2023). Our long-read data have allowed for the reconstruction of the complete control region in *A. cooleyi*, and while large, the size falls within the known range for related insects. Tandem Repeat Finder found 13 repeats in the control region, but the largest and highest scoring repeats occur in 3 stretches of 299 bp repeats with copy numbers of 3.5, 4.1, and 4.1 spanning a total of 3.5 kb.

Many aphid mitochondrial genomes exhibit a tandem repeat region typically located between trnE and trnF (Wang et al. 2013; Zhang et al. 2014, 2021, 2022; Wei et al. 2019). The size and tandem repeat copy number vary widely across different aphid lineages, ranging from a few hundred to a few thousand base pairs. The repeat region of *A. cooleyi* contains 61 tandem repeats, 59 of which are <30 bp and the 2 largest and highest scoring tandem repeat units are each composed of 9.0 and 11.1 copies of an identical 231 bp repeat spanning a total of 4.6 kb. This repeat matches the partially assembled repeat regions of *A. tsugae* (77.56% identity, 95% length) and the aphid species *Eutrichosiphum pasaniae* (71.23% identity, 91% length) (Li et al. 2020a) and *Cavariella salicicola* (77.27% identity, 34% length) (Wang et al. 2013) in BLASTn searches against NR. Recent studies have proposed that the repeat region either evolved in the common ancestor of aphids and was lost in several aphid lineages or evolved independently multiple times (Chen et al. 2019; Song et al. 2019). The presence of this region in *A. cooleyi* and *A. tsugae*

supports the hypothesis that the repeat region is an ancestral feature of Aphidomorpha. The ancient origin of the repeat region and its persistence throughout their ~245 My evolution (Johnson et al. 2018; Szewedo and Nel 2011) has raised questions about its origin and function. It has been proposed that the repeat region in aphids evolved from an ancient duplication of the control region and functions as a second origin of replication to enhance mitochondrial replication and transcription (Wang et al. 2015). Song et al. (2019) reported that the pairwise identity between the control and repeat regions for several aphid species is on average 38%. In *A. cooleyi*, these regions share 48.6% identity (2,199 identical sites). A second origin of replication may be related to the origin of cyclical parthenogenesis; a single female can produce hundreds of offspring over the course of her lifetime, a highly energy-demanding process (Song et al. 2019). Additionally, many host plants produce oxidizing molecules in defense against insects; increased rates of mitochondrial replication and recycling may be a mechanism by which aphids and adelgids deal with these host defenses.

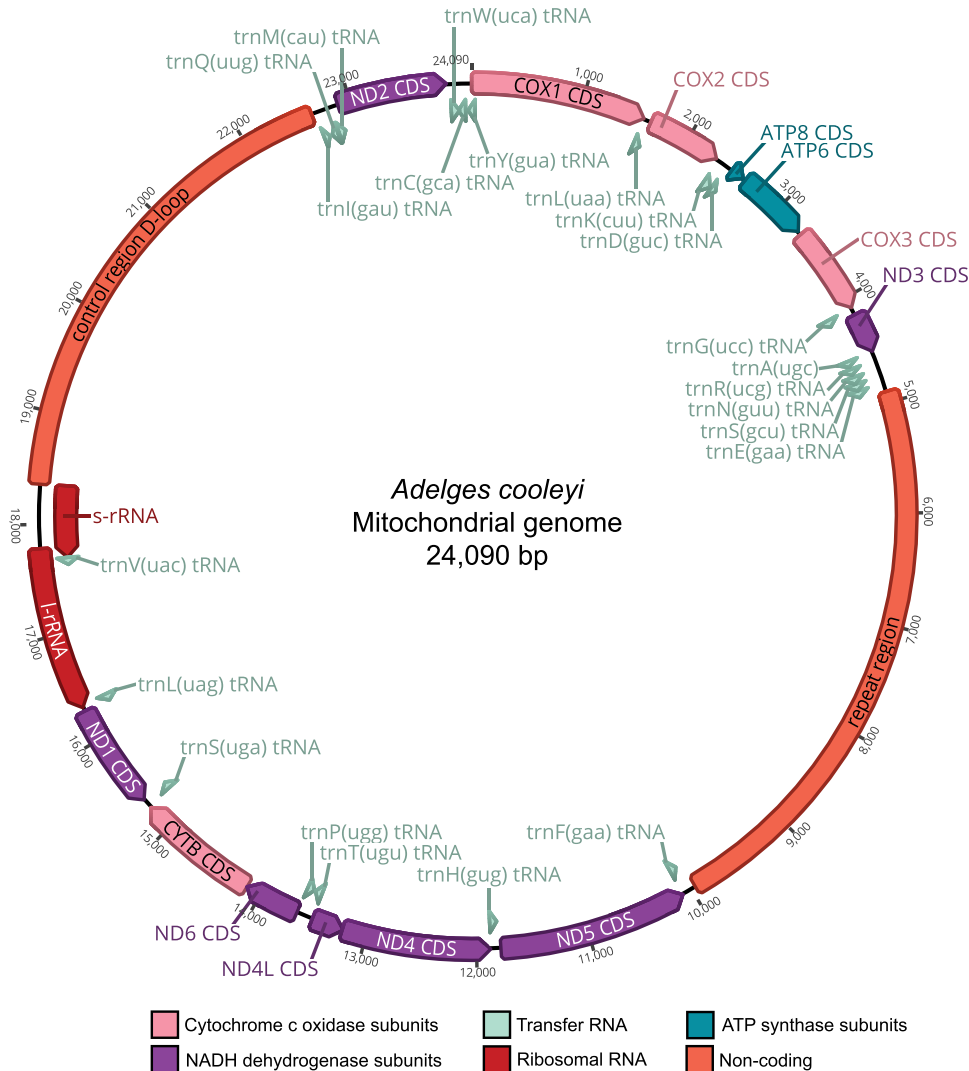
Phylogenomics supports Adelgidae and Phylloxeridae as sister groups

To investigate the phylogenetic position of *A. cooleyi* in relation to other adelgid species and to assess the position of Adelgidae relative to Phylloxeridae and Aphididae, we performed orthology clustering of protein CDS from 4 adelgid species, 5 phylloxera species, 4 aphid species, and 2 coccoid species. This sampling represents the 2 adelgid genera (*Adelges* and *Pineus*) and 4 of the 5 major adelgid lineages (those that feed on pine, hemlock, Douglas fir, and larch secondary hosts), 3 phylloxera genera (*Daktulosphaira*, *Phylloxera*, and *Phylloxera*), 3 aphid tribes (Macrosphini, Aphidini, and Eulachnini), and the Coccidae and Pseudococcidae. OrthoFinder assigned 179,577 genes (91.6% of the total number of genes) to 21,768 orthogroups, with 3,583 orthogroups and 961 single-copy orthologs shared by all species. It should be noted that this does not necessarily represent a complete shared gene set, as transcriptome data were used for some species.

Maximum likelihood analysis with IQ-tree of the 961 concatenated single-copy orthologs generated a fully resolved species tree with 100% support at all nodes (Fig. 5). This topology is identical to the species trees produced by FastTree and STAG within OrthoFinder. Our species tree identifies phylloxerans as sister to adelgids and Adelgidae and Phylloxeridae as sister to Aphididae. The topology within Aphididae is consistent with previous analyses (von Dohlen et al. 2006; Novakova et al. 2013; Julca et al. 2020; Hardy et al. 2022; Owen and Miller 2022; Smith et al. 2022), and the topology within Adelgidae mirrors the relationships established using 3 mitochondrial genes and 1 nuclear gene (Havill et al. 2007).

Table 4. Gene annotation summary statistics for *A. cooleyi*.

Feature	Count	Mean length (bp)	Median length (bp)	Min length (bp)	Max length (bp)
Genes	15,100	10,785	4,267	68	414,770
All transcripts	27,577	2,330	1,781	68	62,780
mRNA	24,967	2,447	1,871	96	62,780
misc_RNA	627	2,258	1,881	193	9,857
tRNA	212	74	73	71	84
lncRNA	1,613	1,050	758	83	34,541
rRNA	65	316	119	119	4,328
CDSs	24,967	1,778	1,275	96	61,812
Exons	128,259	271	171	1	34,139
Introns	110,635	1,695	189	30	404,309

**Fig. 4.** Circular mitochondrial genome map for *A. cooleyi*.

Our phylogenomic results are consistent with the morphological and reproductive features that support the grouping of adelgids and phylloxerans to the exclusion of aphids. The possession of intimate symbioses with obligate nutritional bacteria is ubiquitous throughout the aphids and adelgids but absent in the phylloxerans and is thus most likely a plesiomorphic trait (Ponsen 1997). It has been hypothesized that the ancestral phylloxera lost their obligate symbionts and symbiotic cells known as bacteriocytes as a result of specialized feeding on modified parenchyma cells

containing essential amino acids, relaxing selection on the maintenance of genes underlying symbiont nutritional pathways (Buchner 1965; Warwick and Hildebrandt 1966; von Dohlen et al. 2017). One intriguing question is the developmental and evolutionary origin of the cells that contain symbionts within Aphidomorpha. Parsimony would suggest that bacteriocytes were present in the common ancestor of the Aphidomorpha but were lost in the phylloxera and maintained within the adelgids and aphids. However, similar symbiotic cells have evolved

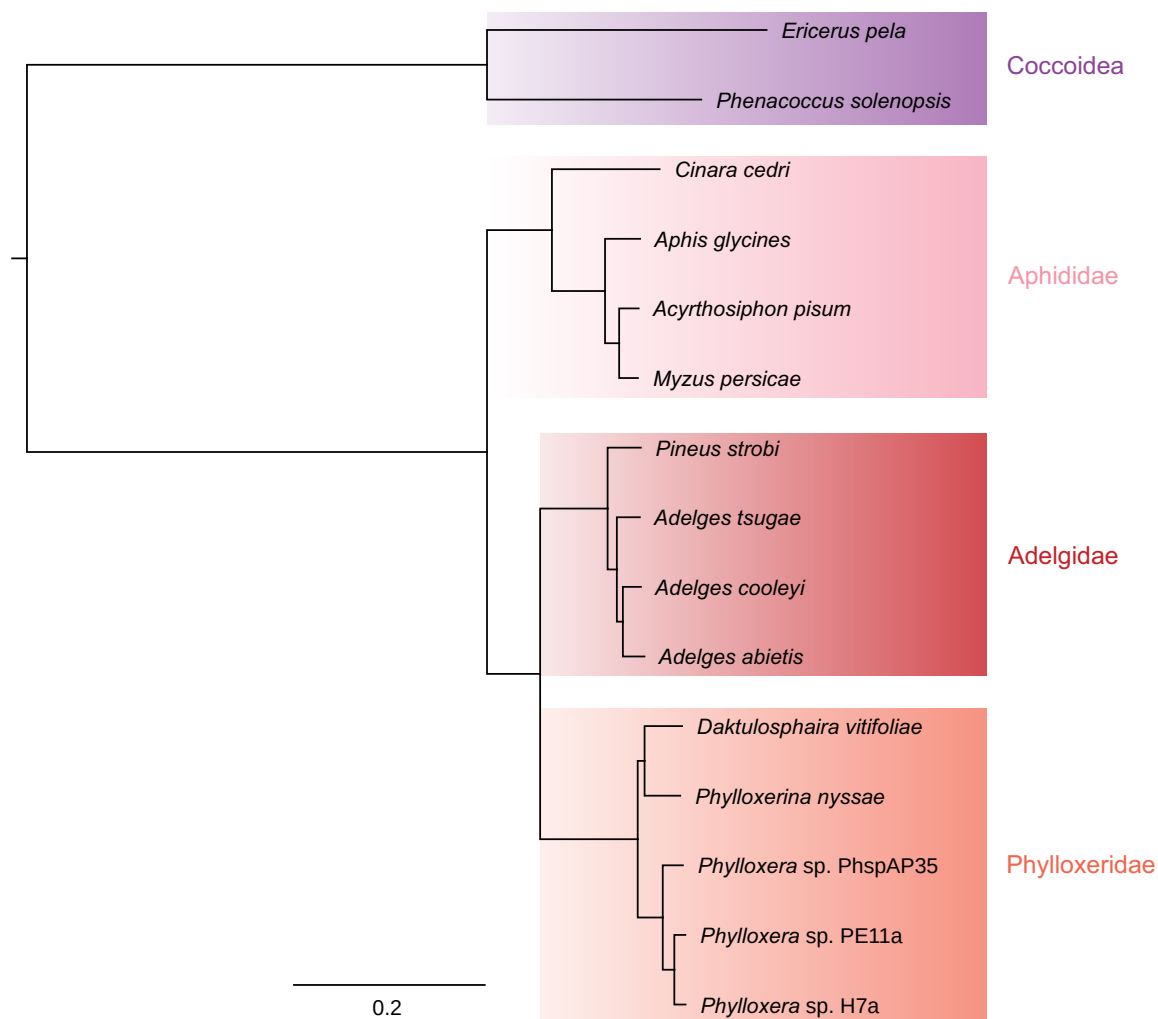


Fig. 5. Maximum-likelihood phylogeny of selected aphid, adelgid, phylloxeran, and coccoid species with sequenced genomes and transcriptomes based on a concatenated alignment of 961 single-copy orthologs and 1,000 ultrafast bootstraps in IQ-tree. *Adelges tsugae*, *A. abietis* (*Adelges* sp. adspAD), *P. strobi* (*Pineus* sp. pispAD), *P. nyssae*, *Phylloxera* sp. PhspAP35, *Phylloxera* sp. PE11a, and *Phylloxera* sp. H7a gene sets originate from transcriptomic data, and the remaining taxa are from genomic datasets. Clades are colored by family, and branch lengths represent amino acid substitutions per site. The bootstrap support at all branches was 100%.

independently in multiple insect orders (Buchner 1965) and even within families (Kuechler et al. 2019), suggesting that the true evolutionary scenario could be more complex. Alternatively, adelgids may have evolved their bacteriocytes independently from aphids, which, if true, may bear some influence on the comparative characteristics of symbiosis within the adelgids and aphids. Future work exploring these questions could reveal a more comprehensive picture of the evolution of symbiosis within this group.

Conclusions

We have generated a high-quality draft genome of *A. cooleyi* using PacBio long-read data and high-coverage Illumina RNA-seq data. This genome is the first representative of the adelgid lineage and adds to a small but growing list of high-quality genomes within Aphidomorpha. As an outgroup of Aphididae, it will aid in the detailed comparative analysis of aphid biology and evolution. Furthermore, we anticipate that it will enhance the studies of insect-microbe interactions relating to adelgid bacterial endosymbionts and the history of symbioses within Aphidomorpha.

Data availability

The nuclear and mitochondrial genome assemblies and annotations for this project are available under RefSeq accession GCF_023614345.1 and GenBank accession OQ991941.1, respectively. PacBio long-read sequences and RNA-seq reads are available under the BioProject accession numbers PRJNA827457 and PRJNA870591, respectively. The code used for the assembly pipeline is available on GitHub (https://github.com/dustin-dial/Adelges_cooleyi_genome).

Supplemental material available at G3 online.

Acknowledgments

The support and resources from the Georgia Genomics and Bioinformatics Core and the Georgia Advanced Computing Research Center at the University of Georgia are gratefully acknowledged.

Funding

The research was supported by National Science Foundation grants to C.D.v.D. (DEB-1655182) and G.R.B. (DEB-1655177) and the Utah Agricultural Experiment Station (approved as journal paper #9695).

Conflicts of interest

The author(s) declare no conflict of interest.

Literature cited

- Ahern RG, Hawthorne DJ, Raupp MJ. 2009. Phylogeography of a specialist insect, *Adelges cooleyi*: historical and contemporary processes shape the distribution of population genetic variation. *Mol Ecol*. 18(2):343–356. doi:10.1111/j.1365-294X.2008.04034.x.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*. 215(3):403–410. doi:10.1016/S0022-2836(05)80360-2.
- Amman GD. 1965. Balsam woolly aphid in the southern Appalachians. *J For*. 63(1):18–20. doi:10.1093/jof/63.1.18.
- Annand PN. 1928. A Contribution Toward a Monograph of the Adelginae (Phylloxeridae) of North America. Redwood City (CA): Stanford University Press.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 27(2):573–580. doi:10.1093/nar/27.2.573.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30(15):2114–2120. doi:10.1093/bioinformatics/btu170.
- Boyce TM, Zwick ME, Aquadro CF. 1989. Mitochondrial-DNA in the bark weevils—size, structure and heteroplasmy. *Genetics*. 123(4):825–836. doi:10.1093/genetics/123.4.825.
- Buchfink B, Reuter K, Drost HG. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 18(4):366–368. doi:10.1038/s41592-021-01101-x.
- Buchner P. 1965. Endosymbiosis of Animals with Plant Microorganisms. New York (NY): John Wiley and Sons. p. 901.
- Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. In: 9th Annual Genomics of Energy & Environment Meeting, Walnut Creek (CA), March 17–20. <https://sourceforge.net/projects/bbmap/>.
- Cameron SL. 2014. How to sequence and annotate insect mitochondrial genomes for systematic and comparative genomics research. *Syst Entomol*. 39(3):400–411. doi:10.1111/syen.12071.
- Chen WB, Shakir S, Bigham M, Richter A, Fei ZJ, Jander G. 2019b. Genome sequence of the corn leaf aphid (*Rhopalosiphum maidis* Fitch). *Gigascience*. 8(4):4. doi:10.1093/gigascience/giz033.
- Chen J, Wang Y, Qin M, Jiang LY, Qiao GX. 2019a. The mitochondrial genome of *Greenidea psidii* van der Goot (Hemiptera: Aphididae: Greenideinae) and comparisons with other Aphididae aphids. *Int J Biol Macromol*. 122:824–832. doi:10.1016/j.ijbiomac.2018.10.209.
- Coombe L, Zhang J, Vandervalk BP, Chu J, Jackman SD, Birol I, Warren RL. 2018. Arks: chromosome-scale scaffolding of human genome drafts with linked read kmers. *BMC Bioinform*. 19(1):234. doi:10.1186/s12859-018-2243-x.
- Cumming ME. 1962. A monomorphic cycle of *Adelges cooleyi* (Gill.) (Homoptera: Phylloxeridae) living only on spruce. *Can Entomol*. 94(11):1190–1195. doi:10.4039/Ent941190-11.
- Dial DT, Weglarz KM, Aremu AO, Havill NP, Pearson TA, Burke GR, von Dohlen CD. 2022. Transitional genomes and nutritional role reversals identified for dual symbionts of adelgids (Aphidoidea: Adelgidae). *ISME J*. 16(3):642–654. doi:10.1038/s41396-021-01102-w.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 29(1):15–21. doi:10.1093/bioinformatics/bts635.
- Dohm JC, Peters P, Stralis-Pavese N, Himmelbauer H. 2020. Benchmarking of long-read correction methods. *NAR Genom Bioinform*. 2(2):lqaa037. doi:10.1093/nargab/lqaa037.
- Emms DM, Kelly S. 2017. Stride: species tree root inference from gene duplication events. *Mol Biol Evol*. 34(12):3267–3278. doi:10.1093/molbev/msx259.
- Emms DM, Kelly S. 2018. STAG: species tree inference from all genes. *bioRxiv* 267914. <https://doi.org/10.1101/267914>, preprint: not peer reviewed.
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 20(1):238. doi:10.1186/s13059-019-1832-y.
- Filipović I, Hereward JP, Rašić G, Devine GJ, Furlong MJ, Etebari K. 2021. The complete mitochondrial genome sequence of *Oryctes rhinoceros* (Coleoptera: Scarabaeidae) based on long-read nanopore sequencing. *PeerJ*. 9:e10552. doi:10.7717/peerj.10552.
- Gan HM, Grandjean F, Jenkins TL, Austin CM. 2019. Absence of evidence is not evidence of absence: nanopore sequencing and complete assembly of the European lobster (*Homarus gammarus*) mitogenome uncovers the missing *nad2* and a new major gene cluster duplication. *BMC Genomics*. 20(1):335. doi:10.1186/s12864-019-5704-3.
- Gillette CP. 1907. *Chermes* of Colorado conifers. *Proc Acad Nat Sci Philadelphia*. 59(1): 3–22.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat Biotechnol*. 29(7):644–U130. doi:10.1038/nbt.1883.
- Grimaldi D, Engel MS. 2005. *Evolution of the Insects*. Cambridge University Press.
- Hardy NB, Kaczvinsky C, Bird G, Richter R, Dettman JR, Maw E, Brunet BMT, Footitt RG. 2022. Geographic isolation drives speciation in Nearctic aphids. *Commun Biol*. 5(1):796. doi:10.1038/s42003-022-03771-5.
- Havill NP, Footitt RG. 2007. Biology and evolution of Adelgidae. *Annu Rev Entomol*. 52(1):325–349. doi:10.1146/annurev.ento.52.110405.091303.
- Havill NP, Footitt 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. doi:10.1016/j.ympev.2006.11.008.
- Havill NP, Griffin BP, Andersen JC, Footitt RG, Justesen MJ, Caccone A, D'Amico V, Elkinton JS. 2021. Species delimitation and invasion history of the balsam woolly adelgid, *Adelges (Dreyfusia) piceae* (Hemiptera: Aphidoidea: Adelgidae), species complex. *Syst Entomol*. 46(1):186–204. doi:10.1111/syen.12456.
- Havill NP, Shiyake S, Lamb Galloway A, Footitt RG, Yu G, Paradis A, Elkinton J, Montgomery ME, Sano M, Caccone A. 2016. Ancient and modern colonization of North America by hemlock woolly adelgid, *Adelges tsugae* (Hemiptera: Adelgidae), an invasive insect from East Asia. *Mol Ecol*. 25(9):2065–2080. doi:10.1111/mec.13589.
- Heie OE, Wegierek P. 2009. A classification of the Aphidomorpha (Hemiptera: Sternorrhyncha) under consideration of the fossil taxa. *Redia*. 92:69–77.
- Hunt M, De Silva N, Otto TD, Parkhill J, Keane JA, Harris SR. 2015. Circlator: automated circularization of genome assemblies using long sequencing reads. *Genome Biol*. 16(1):294. doi:10.1186/s13059-015-0849-0.
- Jackman SD, Coombe L, Chu J, Warren RL, Vandervalk BP, Yeo S, Xue Z, Mohamadi H, Bohlmann J, Jones SJM, et al. 2018. Tigrint: correcting assembly errors using linked reads from large molecules. *BMC Bioinformatics*. 19(1):393. doi:10.1186/s12859-018-2425-6.

- Jiang X, Zhang Q, Qin Y, Yin H, Zhang S, Li Q, Zhang Y, Fan J, Chen J. 2019. A chromosome-level draft genome of the grain aphid *Sitobion miscanthi*. *Gigascience*. 8(8):giz101. doi:[10.1093/gigascience/giz101](https://doi.org/10.1093/gigascience/giz101).
- Johnson KP, Dietrich CH, Friedrich F, Beutel RG, Wipfler B, Peters RS, Allen JM, Petersen M, Donath A, Walden KKO, et al. 2018. Phylogenomics and the evolution of hemipteroid insects. *Proc Natl Acad Sci U S A*. 115(50):12775–12780. doi:[10.1073/pnas.1815820115](https://doi.org/10.1073/pnas.1815820115).
- Julca I, Marcet-Houben M, Cruz F, Vargas-Chavez C, Johnston JS, Gómez-Garrido J, Frias L, Corvelo A, Loska D, Cámara F, et al. 2020. Phylogenomics identifies an ancestral burst of gene duplications predating the diversification of Aphidomorpha. *Mol Biol Evol*. 37(3):730–756. doi:[10.1093/molbev/msz261](https://doi.org/10.1093/molbev/msz261).
- 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. doi:[10.1038/nmeth.4285](https://doi.org/10.1038/nmeth.4285).
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780. doi:[10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010).
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 37(8):907–915. doi:[10.1038/s41587-019-0201-4](https://doi.org/10.1038/s41587-019-0201-4).
- Kinkar L, Young ND, Sohn WM, Stroehlein AJ, Korhonen PK, Gasser RB. 2020. First record of a tandem-repeat region within the mitochondrial genome of *Clonorchis sinensis* using a long-read sequencing approach. *PLoS Negl Trop Dis*. 14(8):e0008552. doi:[10.1371/journal.pntd.0008552](https://doi.org/10.1371/journal.pntd.0008552).
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 37(5):540–546. doi:[10.1038/s41587-019-0072-8](https://doi.org/10.1038/s41587-019-0072-8).
- 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. doi:[10.1101/gr.215087.116](https://doi.org/10.1101/gr.215087.116).
- Kuechler SM, Fukatsu T, Matsuura Y. 2019. Repeated evolution of bacteriocytes in lygaeoid stinkbugs. *Environ Microbiol*. 21(11):4378–4394. doi:[10.1111/1462-2920.14804](https://doi.org/10.1111/1462-2920.14804).
- Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Research*. 6:1287. doi:[10.12688/f1000research.12232.1](https://doi.org/10.12688/f1000research.12232.1).
- Levy Karin E, Mirdita M, Soding J. 2020. Metaeuk-sensitive, high-throughput gene discovery, and annotation for large-scale eukaryotic metagenomics. *Microbiome*. 8(1):48. doi:[10.1186/s40168-020-00808-x](https://doi.org/10.1186/s40168-020-00808-x).
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 34(18):3094–3100. doi:[10.1093/bioinformatics/bty191](https://doi.org/10.1093/bioinformatics/bty191).
- Li C, Jiang L, Zhang X, Chen J, Qiao G. 2020a. The complete mitochondrial genome of *Eutrichosiphum pasaniae* (Okajima, 1908) (Hemiptera: Aphididae: Greenideinae). *Mitochondrial DNA B Resour*. 5(3):3668–3669. doi:[10.1080/23802359.2020.1831997](https://doi.org/10.1080/23802359.2020.1831997).
- Li K, Liang AP. 2018. Hemiptera mitochondrial control region: new sights into the structural organization, phylogenetic utility, and roles of tandem repetitions of the noncoding segment. *Int J Mol Sci*. 19(5):1292. doi:[10.3390/ijms19051292](https://doi.org/10.3390/ijms19051292).
- Li Y, Park H, Smith TE, Moran NA. 2019. Gene family evolution in the pea aphid based on chromosome-level genome assembly. *Mol Biol Evol*. 36(10):2143–2156. doi:[10.1093/molbev/msz138](https://doi.org/10.1093/molbev/msz138).
- Li MZ, Tong HJ, Wang SP, Ye WY, Li ZC, Omar MAA, Ao Y, Ding SM, Li ZH, Wang Y, et al. 2020. A chromosome-level genome assembly provides new insights into paternal genome elimination in the cotton mealybug *Phenacoccus solenopsis*. *Mol Ecol Resour*. 20(6):1733–1747. doi:[10.1111/1755-0998.13232](https://doi.org/10.1111/1755-0998.13232).
- Łukasik P, Chong RA, Nazario K, Matsuura Y, Bublitz AC, Campbell MA, Meyer MC, Van Leuven JT, Pessacq P, Veloso C, et al. 2019. One hundred mitochondrial genomes of cicadas. *J Hered*. 110(2):247–256. doi:[10.1093/jhered/esy068](https://doi.org/10.1093/jhered/esy068).
- Manni M, Berkeley MR, Seppey M, Simao FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 38(10):4647–4654. doi:[10.1093/molbev/msab199](https://doi.org/10.1093/molbev/msab199).
- Mathers TC. 2020. Improved genome assembly and annotation of the soybean aphid (*Aphis glycines* Matsumura). *G3 (Bethesda)*. 10(3):899–906. doi:[10.1534/g3.119.400954](https://doi.org/10.1534/g3.119.400954).
- Mathers TC, Chen YZ, 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. doi:[10.1186/s13059-016-1145-3](https://doi.org/10.1186/s13059-016-1145-3).
- 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. doi:[10.1093/molbev/msaa246](https://doi.org/10.1093/molbev/msaa246).
- Meng G, Li Y, Yang C, Liu S. 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res*. 47(11):e63. doi:[10.1093/nar/gkz173](https://doi.org/10.1093/nar/gkz173).
- Moran NA. 2001. The coevolution of bacterial endosymbionts and phloem-feeding insects. *Ann Mo Bot Gard*. 88(1):35–44. doi:[10.2307/2666130](https://doi.org/10.2307/2666130).
- Morgan B, Wang TY, Chen YZ, Moctezuma V, Burgos O, Le MH, Huang JP. 2022. Long-read sequencing data reveals dynamic evolution of mitochondrial genome size and the phylogenetic utility of mitochondrial DNA in Hercules beetles (Dynastes; Scarabaeidae). *Genome Biol Evol*. 14(10):evac147. doi:[10.1093/gbe/evac147](https://doi.org/10.1093/gbe/evac147).
- Morgulis A, Gertz EM, Schäffer AA, Agarwala R. 2006. WindowMasker: window-based masker for sequenced genomes. *Bioinformatics*. 22(2):134–141. doi:[10.1093/bioinformatics/bti774](https://doi.org/10.1093/bioinformatics/bti774).
- Nicholson SJ, Nickerson ML, Dean M, Song Y, Hoyt PR, Rhee H, Kim C, Puterka GJ. 2015. The genome of *Diuraphis noxia*, a global aphid pest of small grains. *BMC Genomics*. 16(1):429. doi:[10.1186/s12864-015-1525-1](https://doi.org/10.1186/s12864-015-1525-1).
- Nováková E, Hypša V, Klein J, Footitt RG, von Dohlen CD, Moran NA. 2013. Reconstructing the phylogeny of aphids (Hemiptera: Aphididae) using DNA of the obligate symbiont *Buchnera aphidicola*. *Mol Phylogenet Evol*. 68(1):42–54. doi:[10.1016/j.ympev.2013.03.016](https://doi.org/10.1016/j.ympev.2013.03.016).
- Novosolov M, Yahalomi D, Chang ES, Fiala I, Cartwright P, Huchon D. 2022. The phylogenetic position of the enigmatic, *Polypodium hydri-forme* (Cnidaria, Polypodiozoa): insights from mitochondrial genomes. *Genome Biol Evol*. 14(8):evac112. doi:[10.1093/gbe/evac112](https://doi.org/10.1093/gbe/evac112).
- Owen CL, Miller GL. 2022. Phylogenomics of the Aphididae: deep relationships between subfamilies clouded by gene tree discordance, introgression and the gene tree anomaly zone. *Syst Entomol*. 47(3):470–486. doi:[10.1111/syen.12542](https://doi.org/10.1111/syen.12542).
- Ozaki K. 1995. Intergall migration in aphids—a model and a test of ESS dispersal rate. *Evol Ecol*. 9(5):542–549. doi:[10.1007/BF01237835](https://doi.org/10.1007/BF01237835).
- Ponsen MB. 1997. A Histological Description of the Alimentary Tract and Related Organs of Phylloxeridae (Homoptera, Aphidoidea). No. 97-1. Wageningen Agricultural University.
- Price MN, Dehal PS, Arkin AP. 2010. FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS One*. 5(3):e9490. doi:[10.1371/journal.pone.0009490](https://doi.org/10.1371/journal.pone.0009490).
- Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res*. 35(Database issue):D61–D65. doi:[10.1093/nar/gkl842](https://doi.org/10.1093/nar/gkl842).

- Reardon RC, Onken B, United States. Forest Health Technology Enterprise Team. 2011. Implementation and Status of Biological Control of the Hemlock Woolly Adelgid. Morgantown (WV): U.S. Department of Agriculture, Forest Service, Forest Health Technology Enterprise Team.
- Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 21(1):1–27. doi:[10.1186/s13059-020-02134-9](https://doi.org/10.1186/s13059-020-02134-9).
- Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16(6):276–277. doi:[10.1016/S0168-9525\(00\)00204-2](https://doi.org/10.1016/S0168-9525(00)00204-2).
- Rispe C, Legeai F, Nabity PD, Fernández R, Arora AK, Baa-Puyoulet P, Banfill CR, Bao L, Barberà M, Bouallegue 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. doi:[10.1186/s12915-020-00820-5](https://doi.org/10.1186/s12915-020-00820-5).
- Saito S, Tamura K, Aotsuka T. 2005. Replication origin of mitochondrial DNA in insects. *Genetics.* 171(4):1695–1705. doi:[10.1534/genetics.105.046243](https://doi.org/10.1534/genetics.105.046243).
- Sayadi A, Immonen E, Tellgren-Roth C, Arnqvist G. 2017. The evolution of dark matter in the mitogenome of seed beetles. *Genome Biol Evol.* 9(10):2697–2706. doi:[10.1093/gbe/evx205](https://doi.org/10.1093/gbe/evx205).
- Sharbrough J, Bankers L, Cook E, Fields PD, Jalinsky J, McElroy KE, Neiman M, Logsdon JM, Boore JL. 2023. Single-molecule sequencing of an animal mitochondrial genome reveals chloroplast-like architecture and repeat-mediated recombination. *Mol Biol Evol.* 40(1):1. doi:[10.1093/molbev/msad007](https://doi.org/10.1093/molbev/msad007).
- Shen W, Ren H. 2021. TaxonKit: a practical and efficient NCBI taxonomy toolkit. *J Genet Genomics.* 48(9):844–850. doi:[10.1016/j.jgg.2021.03.006](https://doi.org/10.1016/j.jgg.2021.03.006).
- Smith TE, Li Y, Perreau J, Moran NA. 2022. Elucidation of host and symbiont contributions to peptidoglycan metabolism based on comparative genomics of eight aphid subfamilies and their *Buchnera*. *PLoS Genet.* 18(5):e1010195. doi:[10.1371/journal.pgen.1010195](https://doi.org/10.1371/journal.pgen.1010195).
- Song HJ, Donthu RK, Hall R, Hon L, Weber E, Badger JH, Giordano R. 2019. Description of soybean aphid (*Aphis glycines* Matsumura) mitochondrial genome and comparative mitogenomics of Aphididae (Hemiptera: Sternorrhyncha). *Insect Biochem Mol Biol.* 113:103208. doi:[10.1016/j.ibmb.2019.103208](https://doi.org/10.1016/j.ibmb.2019.103208).
- Souvorov A, Kapustin Y, Kiryutin B, Chetvermin V, Tatusova T, Lipman D. 2010. Gnomon—NCBI Eukaryotic Gene Prediction Tool. Bethesda (MD): National Center for Biotechnology Information. p. 1–24.
- Steffan AW. 1968. Evolution und Systematik der Adelgidae (Homoptera: Aphidina): Eine Verwandtschaftsanalyse auf vorwiegend ethologischer, zytologischer und karyologischer Grundlage. *Zoologica.* 115:1–139.
- Szwedo J, Nel A. 2011. The oldest aphid insect from the Middle Triassic of the Vosges, France. *Acta Palaeontol Pol.* 56(4):757–766. doi:[10.4202/app.2010.0034](https://doi.org/10.4202/app.2010.0034).
- von Dohlen CD, Moran NA. 1995. Molecular phylogeny of the Homoptera—a paraphyletic taxon. *J Mol Evol.* 41(2):211–223. doi:[10.1007/BF00170675](https://doi.org/10.1007/BF00170675).
- von Dohlen CD, Moran NA. 2000. Molecular data support a rapid radiation of aphids in the Cretaceous and multiple origins of host alternation. *Biol J Linn Soc.* 71(4):689–717. doi:[10.1006/bjll.2000.0470](https://doi.org/10.1006/bjll.2000.0470).
- von Dohlen CD, Rowe CA, Heie OE. 2006. A test of morphological hypotheses for tribal and subtribal relationships of Aphidinae (Insecta: Hemiptera: Aphididae) using DNA sequences. *Mol Phylogenet Evol.* 38(2):316–329. doi:[10.1016/j.ympev.2005.04.035](https://doi.org/10.1016/j.ympev.2005.04.035).
- von Dohlen CD, Spaulding U, Patch KB, Weglarz KM, Foottit RG, Havill NP, Burke GR. 2017. Dynamic acquisition and loss of dual-obligate symbionts in the plant-sap-feeding Adelgidae (Hemiptera: Sternorrhyncha: Aphidoidea). *Front Microbiol.* 8:1037. doi:[10.3389/fmicb.2017.01037](https://doi.org/10.3389/fmicb.2017.01037).
- Wang Y, Chen J, Jiang LY, Qiao GX. 2015. The complete mitochondrial genome of *Mindarus keteleerifoliae* (Insecta: Hemiptera: Aphididae) and comparison with other Aphididae insects. *Int J Mol Sci.* 16(12):30091–30102. doi:[10.3390/ijms161226219](https://doi.org/10.3390/ijms161226219).
- Wang Y, Huang XL, Qiao GX. 2013. Comparative analysis of mitochondrial genomes of five aphid species (Hemiptera: Aphididae) and phylogenetic implications. *PLoS One.* 8(10):e77511. doi:[10.1371/journal.pone.0077511](https://doi.org/10.1371/journal.pone.0077511).
- Warick RP, Hildebrandt AC. 1966. Free amino acid contents of stem and phylloxera gall tissue cultures of grape. *Plant Physiol.* 41(4):573–578. doi:[10.1104/pp.41.4.573](https://doi.org/10.1104/pp.41.4.573).
- Warren RL. 2016. Rails and cobbler: scaffolding and automated finishing of draft genomes using long DNA sequences. *J Open Source Softw.* 1(7):116. doi:[10.21105/joss.00116](https://doi.org/10.21105/joss.00116).
- 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. doi:[10.1093/gbe/evy114](https://doi.org/10.1093/gbe/evy114).
- Wei DD, Lang N, Tao Y, He W, Tu YQ, Miao ZQ, Yang L, Wang JJ. 2019. The mitochondrial genome of the brown citrus aphid *Aphis (Toxoptera) citricidus*: insights into the repeat regions in aphids and phylogenetic implications. *Int J Biol Macromol.* 136:531–539. doi:[10.1016/j.ijbiomac.2019.06.101](https://doi.org/10.1016/j.ijbiomac.2019.06.101).
- Xu JS, Liao CQ, Guo QY, Long CP, Dai XH. 2018. Mitochondrial genome of a leaf-mining beetle *Prionispa champaka* Maulik (Coleoptera: Chrysomelidae: Cassidinae). *Mitochondrial DNA B Resour.* 3(1):147–148. doi:[10.1080/23802359.2017.1413318](https://doi.org/10.1080/23802359.2017.1413318).
- Yang P, Yu SH, Hao JJ, Liu W, Zhao ZL, Zhu ZR, Sun T, Wang XQ, Song QS. 2019. Genome sequence of the Chinese white wax scale insect *Ericerus pela*: the first draft genome for the Coccidae family of scale insects. *Gigascience.* 8(9):giz113. doi:[10.1093/gigascience/giz113](https://doi.org/10.1093/gigascience/giz113).
- Yeh HT, Ko CC, Wu LW. 2020. The first complete mitochondrial genome of *Adelges tsugae* Annand (Hemiptera: Adelgidae). *Mitochondrial DNA B Resour.* 5(3):2288–2290. doi:[10.1080/23802359.2020.1772682](https://doi.org/10.1080/23802359.2020.1772682).
- Zhang H, Liu Q, Lu C, Deng J, Huang X. 2021. The first complete mitochondrial genome of Lachninae species and comparative genomics provide new insights into the evolution of gene rearrangement and the repeat region. *Insects.* 12(1):55. doi:[10.3390/insects12010055](https://doi.org/10.3390/insects12010055).
- Zhang H, Lu CC, Liu Q, Zou TM, Qiao GX, Huang XL. 2022. Insights into the evolution of aphid mitogenome features from new data and comparative analysis. *Animals (Basel).* 12(15):1970. doi:[10.3390/ani12151970](https://doi.org/10.3390/ani12151970).
- Zhang B, Ma C, Edwards O, Fuller S, Kang L. 2014. The mitochondrial genome of the Russian wheat aphid *Diuraphis noxia*: large repetitive sequences between trnE and trnF in aphids. *Gene.* 533(1):253–260. doi:[10.1016/j.gene.2013.09.064](https://doi.org/10.1016/j.gene.2013.09.064).
- Zhu BH, Xiao J, Xue W, Xu GC, Sun MY, Li JT. 2018. P_RNA_scaffolder: a fast and accurate genome scaffolder using paired-end RNA-sequencing reads. *BMC Genomics.* 19(1):175. doi:[10.1186/s12864-018-4567-3](https://doi.org/10.1186/s12864-018-4567-3).