




Genome report: Genome sequence of tuliptree scale, *Toumeyella liriodendri* (Gmelin), an ornamental pest insect

Andrew J. Mongue ^{1,*} Amanda Markee ^{1,2} Ethan Grebler,¹ Tracy Liesenfelt,¹ Erin C. Powell ³

¹Department of Entomology and Nematology, University of Florida, Gainesville, FL 32611, USA

²American Museum of Natural History, New York, NY 10024, USA

³Division of Plant Industry, Florida Department of Agriculture and Consumer Services, Florida State Collection of Arthropods, Gainesville, FL 32608, USA

*Corresponding author: University of Florida, Department of Entomology and Nematology, 1881 Natural Area Drive, Gainesville, FL 32611, USA. Email: andrew.mongue@ufl.edu

Scale insects are of interest both to basic researchers for their unique reproductive biology and to applied researchers for their pest status. In spite of this interest, there remain few genomic resources for this group of insects. To begin addressing this lack of data, we present the genome sequence of tuliptree scale, *Toumeyella liriodendri* (Gmelin) (Hemiptera: Coccoomorpha: Coccidae). The genome assembly spans 536 Mb, with over 96% of sequence assembled into one of 17 chromosomal scaffolds. We characterize roughly 66% of this sequence as repetitive and annotate 16,508 protein-coding genes. Then we use the reference genome to explore the phylogeny of soft scales (Coccidae) and evolution of karyotype within the family. We find that *T. liriodendri* is an early-diverging soft scale, less closely related to most sequenced soft scales than a species of the family Aclerdidae is. This molecular result corroborates a previous, morphology-based phylogenetic placement of Aclerdidae within Coccidae. In terms of genome structure, *T. liriodendri* has nearly twice as many chromosomes as the only other soft scale assembled to the chromosome level, *Ericerus pela* (Chavannes). In comparing the two, we find that chromosome number evolution can largely be explained by simple fissions rather than more complex rearrangements. These genomic natural history observations lay a foundation for further exploration of this unique group of insects.

Keywords: tuliptree scale; scale insect; paternal genome elimination; Coccoidea; Coccoomorpha; non-mendelian genetics; Hemiptera; chromosome-level assembly

Introduction

Scale insects are a uniquely interesting group of arthropods, as they hold great importance for both basic and applied research. On the basic side, this group displays remarkable reproductive diversity (Gavrilov 2007; Blackmon et al. 2017; Ross et al. 2022), from familiar sexual reproduction with chromosomal sex determination (Blackmon et al. 2017) to hermaphroditism (Hughes-Schrader 1925; Mongue et al. 2021) and a complex form of purely autosomal haplodiploidy known as paternal genome elimination (PGE) (Nur 1980). Very little is known about either the molecular mechanisms of these alternative sex determination systems or how the clade transitioned between them. On the applied side, many scale insects are globally invasive pests that require intervention to control (Caltagirone and Doutt 1989; Waterhouse 1991; García Morales et al. 2016), but others are both geographically limited and specialized on one or a few host plants (e.g. Spanish moss ensign scale; Morrison 1925). Understanding the factors that differentiate these 2 groups will help better target screening and management practices. Specifically, both applied and basic research goals are held back by a lack of modern genomic resources compared with other insect clades (e.g. Lepidoptera; Wright et al. 2024). In service of beginning to address this scarcity, we have sequenced the genome of the tuliptree scale.

Toumeyella liriodendri (Gmelin 1790), or tuliptree scale, is a soft scale insect (Hemiptera: Coccoomorpha: Coccidae). It reproduces

sexually, but with a fully autosomal genome and sex determined by PGE (Nur 1980). Under this system, females are fully diploid in karyotype and gene expression, but males silence and ultimately discard their paternally inherited chromosomes, making them functionally haploid (Gavrilov 2007). Unlike better-studied mealybugs (Pseudococcidae; Ross et al. 2010, 2024; de la Filia et al. 2021), which keep the entire silenced paternal genome in somatic cells, soft scales employ a variant of PGE in which some paternal chromosomes are lost during cell division, leading to a variable karyotype between cells within the same male (Gavrilov 2007). Coincident with this rare sex determination system, adults are extremely sexually dimorphic, with large, long-lived, sessile females and small, winged, non-feeding males (Fig. 1).

Ecologically, tuliptree scales are presumed native to the eastern United States and likely introduced to California (Hamon and Williams 1984) and Cuba (Novoa et al. 2011). This species can be a major pest of native trees including tulip tree, *Liriodendron tulipifera* L., which is commonly grown for timber, and trees in the genus *Magnolia* L., which are popular ornamental and shade trees (Burns and Donley 1970; Hamon and Williams 1984; Gill 1988). While this species prefers Magnoliaceae, it will feed on a variety of hosts across several families (Hamon and Williams 1984; Gill 1988; Miller and Williams 1995). To support the study of both the evolution of PGE and monitoring of a tree pest, we present a chromosome-level assembly of *T. liriodendri*.



Fig. 1. Adult *T. liriodendri* display extreme sexual dimorphism. Males (winged) are not only smaller but possess many features (wings, eyes) that are not present in adult females (larger patterned domed individuals). Top: Photograph of a colony of *T. liriodendri* with 2 males and 8 females. Bottom: Rendering of individual female (left) and male (right) from a dorsal view to highlight sexual dimorphism. Photograph by E. Powell, illustration by T. Liesenfelt.

Materials and methods

Sample collection and identification

We collected adult female *T. liriodendri* from a single colony feeding on *Magnolia grandiflora* L. (Magnoliaceae) in Gainesville, Florida, USA (29.686122, -82.345128), on 2023 June 28. We confirmed species identity with slide-mounted specimens following Hamon and Williams (1984) and Kondo and Williams (2008). Specimens were deposited in the Florida State Collection of Arthropods (FSCA) in Gainesville, Florida (barcodes FSCA_00072042–FSCA_00072046).

DNA extraction and sequencing

We followed a sequencing strategy that has proven successful in other recent insect genome sequencing projects, including other scale insects (Mongue et al. 2024; Ross et al. 2024). First, we extracted DNA from a single adult female using a modified OmniPrep extraction protocol (G-Biosciences, St Louis, MO, USA); we followed the manufacturer's protocol for solid tissue, but extended digestion with proteinase K to overnight (~15 h) and extended the DNA precipitation step to 1 h at -20°C. After quality checking, we sequenced genomic DNA at the University of Florida's Interdisciplinary Center for Biotechnology Research on the PacBio Sequel IIe platform (Menlo Park, CA, USA) to generate HiFi long reads from a single SMRTCell. Because of the limited yield of DNA from a single individual, we chose a low-input library

preparation strategy that did not include DNA fragment size selection; this approach was a calculated tradeoff to avoid introducing additional haplotypes into the primary assembly. High molecular weight genomic DNA preparations were evaluated for read length distribution using the Agilent TapeStation, using a high-sensitivity genomic tape. DNA preparations (3–5 µg) were cleaned using the MoBio PowerClean DNA Cleanup Kit (# 12877-50). Samples were fragmented to the desired 12–15 kb range using G-tubes (Covaris Inc., # 520079). Following sample fragmentation, DNA was concentrated using AMPure beads, and ~3 µg of clean DNA was used for Sequel IIe library construction following the manufacturer's protocol, for a 30 h movie run. Separately, we pooled adult and juvenile females from the same colony, flash-froze them in liquid nitrogen, and sent tissue to Novogene, Inc. (Sacramento, CA, USA) to generate Arima Hi-C-linked reads (Carlsbad, CA, USA) for scaffolding. We completed the initial HiFi sequencing and genome size estimation (below) before Hi-C sequencing to ensure we chose adequate coverage depth for the latter.

Genome size estimate

Because little is known about *T. liriodendri* and tissue samples were limited and used for genomic DNA extraction, we sought to estimate the expected genome size directly from the raw PacBio reads by counting k-mer frequencies. To do so, we used Jellyfish v2.3.0

Table 1. Summary of currently available soft scale sequence data as well as 2 outgroups from other families: Acleridae and Pseudococcidae.

Species	Family	Data type	Accession
<i>Aclerda</i> sp.	Acleridae	Transcriptome	SRR1821892
<i>Planococcus citri</i>	Pseudococcidae	Genome	GCA_950023065.1
<i>Ericerus pela</i>	Coccidae	Genome	GCA_011428145.1
<i>Ceroplastes cirripediformis</i>	Coccidae	Transcriptome	SRR1821905
<i>Coccus</i> sp.	Coccidae	Transcriptome	SRR1821911
<i>Parthenolecanium corni</i>	Coccidae	Genome	GCA_038050395.1

We used this dataset to explore phylogenetic relationships within the family Coccidae.

(Marcais and Kingsford 2012) to generate frequencies and then used custom R scripts to estimate the genome size from these data.

Genome assembly

We assembled raw HiFi reads with hifiasm v0.16.1 (Cheng et al. 2021), first with all parameters set to defaults and then with the “-l 3” stringency parameter to more aggressively purge duplicated haplotigs. Based on higher contiguity of the latter approach (L50 = 97 vs 95), we proceeded with the more aggressively purged assembly. At this stage we performed additional curation steps, both characterizing potential co-bionts and contaminants using blob-tools v1.0 (Laetsch and Blaxter 2017) and searching for any remaining haplotigs using the purge_dups tool (https://github.com/dfguan/purge_dups), which searches for potential duplicate sequences based on coverage depth of input reads and self-alignment of the assembly; this approach did not identify any duplicate sequences. We also explored circular contigs to identify the mitochondrial sequence and potential co-bionts. We assessed baseline genome completeness with BUSCO v4.1.4 using the hemipteraodb10 dataset (Manni et al. 2021). With these steps complete, we aligned Hi-C-linked reads to the assembly using Arima’s pipeline (found at https://github.com/ArimaGenomics/mapping_pipeline/): briefly, we independently aligned left and right reads using bwa-mem2 v2.2.1 (Vasimuddin et al. 2019) and filtered out chimeric reads while keeping the 5’ end using scripts from the Arima git repository and following the User Guide pdf file. We then merged separate alignment files and removed PCR and optical duplicates using Picard tools v2.25.5’s MergeSamFiles and MarkDuplicates functions (“Picard toolkit” 2019). We input this curated alignment into YaHS v1.1 (Zhou et al. 2023). YaHS generated an initial scaffolded assembly, which we further explored by using the “juicer pre” command to generate JBAT files for visualizing HiC contacts in Juicebox 2.17 (Durand et al. 2016) for manual correction of the assembly. After visual inspection and correction of misplacements, we saved the updated linkage file and input it into YaHS to run “juicer post” to update the assembly to reflect changes made in Juicebox (Zhou et al. 2023). We again assessed BUSCO completeness. Based on the failure of the purge_dups approach to identify haplotigs in the primary assembly, we used BUSCO information to screen haplotigs as follows. We concatenated a list of all scaffolds with single-copy BUSCO sequences as well as a list of those with multicopy (duplicate) sequences. We compared the 2 lists and identified 68 non-chromosomal scaffolds that contained only duplicates and no single-copy BUSCOs. We filtered to remove these from our assembly and proceeded to repeat masking and annotation.

Repeat masking

We characterized repeats in this final curated genome as follows. First, we modeled repeats de novo using RepeatModeler v2.0

(Flynn et al. 2020) including a search for long terminal repeats using the “-LTRstruct” parameter. This generated a set of species-specific repeats, which we concatenated to the end of a custom library which consisted of the 2020 Repbase arthropod and hemipteran repeat databases (Bao et al. 2015), combined with repeats identified with RepeatModeler in other high-quality genomes: *Icerya purchasi* Maskell (Mongue et al. 2024), *Planococcus citri* (Risso) (Ross et al. 2024), and another soft scale *Ericerus pela* (Chavannes) (Yang et al. 2019). We imported this curated database of hemipteran repeats augmented with scale insect specific and *T. liriodendri* specific repeats into RepeatMasker v4.0.9 (Smit et al. 2019) to generate a final soft-masked assembly and summary of repetitive elements.

Gene annotation

Lacking enough samples to generate an RNAseq dataset, we chose a de novo approach to gene annotation, using the machine learning tool helixer (Holst et al. 2023). Helixer requires only a genome sequence and a general lineage (in our case “invertebrate”) to annotate. We passed the softmasked chromosomal assembly to helixer for annotation but note that the helixer tool claims to not be impacted by the presence or absence of masking.

Phylogeny of Coccidae

We sought to use our newly generated genome to explore phylogenetic relationships between soft scale species. To do so, we downloaded existing datasets for other soft scales and outgroups, as shown in Table 1. To analyze the transcriptomic data, we first assembled the transcriptomes using Trinity v2.9.0 (Grabherr et al. 2011), then ran BUSCO v4.1.4 in transcriptome mode (Manni et al. 2021) to extract single-copy orthologs for phylogenetic inference; for genome assemblies, we ran BUSCO (Manni et al. 2021) directly on the genomes. Next, we used a BUSCO_phylogenomics pipeline (https://github.com/jamiemcg/BUSCO_phylogenomics) to gather single-copy BUSCOs present in at least 75% of our sample species and used FastTree v2.1.11 (Price et al. 2009) to create individual gene trees for each protein sequence using the JTT model of evolution (Jones et al. 1992) with the CAT approximation for different rates at each site (Stamatakis 2006). Then we used ASTER’s (<https://github.com/chaoszhang/ASTER>) ASTRAL tool (Zhang et al. 2018) to generate a consensus species tree with the “-R” more subsampling and placements options, and choosing *P. citri* to root as the outgroup. We assessed confidence in the tree with the local posterior probability using a quartet-based algorithm (Sayari and Mirarab 2016).

Karyotype evolution between scale insect species

To date, scale insects of the better-studied mealybug family Pseudococcidae with chromosome-level genome assemblies all have shown a karyotype of $n=5$ (Gavrilov 2007; Li et al. 2020; Ross et al. 2024), but with additional chromosome-level resources

for scale insects comes the opportunity to explore how genome architecture has evolved. We sought to compare our newly generated *T. liriodendri* assembly ($n = 17$) with that of *E. pela* ($n = 9$; [Chen et al. 2021](#)). For this analysis, we removed shorter scaffolds, leaving only the chromosomal pseudomolecule scaffolds and ran Satsuma v2 ([Grabherr et al. 2010](#)) SatsumaSynteny2 to find orthologous matches across the genome. We then processed these matches with “BlockDisplaySatsuma” and visualized them with “ChromosomePaint” to understand the relationship between karyotypes.

Results and discussion

Sequencing and assembly

We generated 31 Gb of raw PacBio HiFi reads (raw data accessions found in [Table 2](#)), with a mean read length of 5,221 bp and a median quality score of 43. We first used these data to set expectations of genome size using k-mer counting analyses. Our analyses suggested we generated an average of $31 \times$ coverage on a roughly 515 Mb genome. The karyotype of *T. liriodendri* has not been previously reported, but the contact map of the downstream assembly showed 17 clear linkage groups, even before curation, suggesting the same karyotype as *Neolecanium cornuparvum* (Thro) ([Gavrilov 2007](#)). After manual curation of the contact map, 96.5% of the total assembled length (517,894,882 bp) is contained in one of 17 chromosomal scaffolds ([Table 3](#)). The size of the genome assigned to chromosomes matches well with the k-mer-based genome size estimates, but we note that at present, we lack an estimate of genome size from flow cytometry to independently validate the genome size of *T. liriodendri*.

In the context of other scale insects, there are no closely related species, i.e. congeners, for comparison to set expectations for genome size. That said, the closest well-characterized genome is the soft scale *E. pela*, which has a genome of 650 Mb ([Yang et al. 2019](#)). Other better-studied scale insects in the mealybug family Pseudococcidae have genomes of roughly 400 Mb ([Vea et al. 2021](#); [Ross et al. 2024](#)). From this perspective, the ~540 Mb genome

Table 2. Location of all primary and assembly data generated for *T. liriodendri*.

Data	Accession (Bioproject: PRJNA1100599)
PacBio HiFi reads	SRX24290617
Illumina Hi-C reads	SRX24290618
Assembly	GCA_041937245.1
Annotation	DOI:10.5281/zenodo.12611068

Table 3. Assembly statistics.

Assembly statistics	Primary HiFi assembly (Hifiasm)	Hi-C scaffolded assembly (YaHS)	Final curated and haplotig purged
Total length	543,812,424 bp	543,870,724 bp	536,198,248 bp
Sequence count	869	435	367
N50	1,574,055 bp	30,207,406 bp	30,207,406 bp
L50	95	8	8
GC content	35.44%	35.44%	35.44%
N content	0 (0.00%)	58,300 (0.01%)	58,300 (0.01%)
BUSCO completeness (hemipteraodb10, $n = 2510$)	C: 90.2% [S: 77.8%, D: 12.4%], F: 2.6%, M: 7.2%	C: 90.3% [S: 79.4%, D: 10.9%], F: 2.6%, M: 7.1%	Genome mode: C: 90.0% [S: 82.6%, D: 7.4%], F: 2.7%, M: 7.3% Annotation: C: 90.4% [S: 81.8%, D: 8.6%], F: 2.1%, M: 7.5%

Assembly size, contiguity, and BUSCO completeness for the primary HiFi assembly, the HiC scaffolded assembly, and the final curated assembly.

of *T. liriodendri* fits well within the established trend and serves to highlight that the scale insect *I. purchasi* is an outlier with a genome of over 1 Gb in length ([Mongue et al. 2024](#)), likely due to its unique reproductive ecology as a self-fertile hermaphrodite ([Mongue et al. 2021](#)).

Working with the final curated assembly (the far right column of [Table 2](#)), we used a combination of de novo repeat finding and matching against a database of known hemipteran repeats. In total, we masked 66% of the genome, with results from RepeatMasker summarized in [Table 4](#). This repeat percentage is higher than the ~55% reported from the related *E. pela* ([Yang et al. 2019](#)), despite the smaller genome size of *T. liriodendri*. This difference is partially attributable to our use of a larger hemipteran repeat database, as applying it to the *E. pela* genome masked ~60% of bases, but it does not fully explain the increase in repeats.

Finally, gene content in scale insects is also poorly characterized. On the low end, annotations of the genome of *E. pela* and *Phenacoccus solenopsis* Tinsley report 12,022 genes ([Yang et al. 2019](#)) and 11,880 genes ([Li et al. 2020](#)), respectively. On the high end, species cataloged on the web resource Mealybug Base (<https://ensembl.mealybug.org/index.html>) range from ~22,000 [*Pseudococcus longispinus* (Targioni Tozzetti)] to ~40,000 (*P. citri*) predicted genes. Our annotation of *T. liriodendri* is closer to the low end, with 16,508 protein-coding genes. At present, we lack RNA

Table 4. Summary of masked repeats in the *T. liriodendri* genome.

Repetitive element	Count	Sequence length (bp)	Percentage of sequence
SINEs	3,162	529,083	0.10
LINEs	36,639	10,916,041	2.04
LINE1	3,326	1,083,625	0.20
LINE2	1,587	515,529	0.10
L3/CR1	1,298	423,116	0.08
LTR elements	109,745	47,806,576	8.92
ERV	4	253	0.00
ERV Class I	15,834	4,356,406	0.81
ERV Class II	53	93,459	0.02
DNA elements	79,859	16,639,012	3.10
hAT-Charlie	75	5,663	0.00
TcMar-Tigger	13,514	1,954,000	0.36
Unclassified	1,270,206	270,951,242	50.53
Total interspersed repeats		34,684,195	64.69
Small RNA	1,961	398,467	0.07
Satellites	56	10,072	0.00
Simple repeats	86,079	4,594,864	0.86
Low complexity	11,037	586,349	0.11%

Output is based on the RepeatMasker summary table.

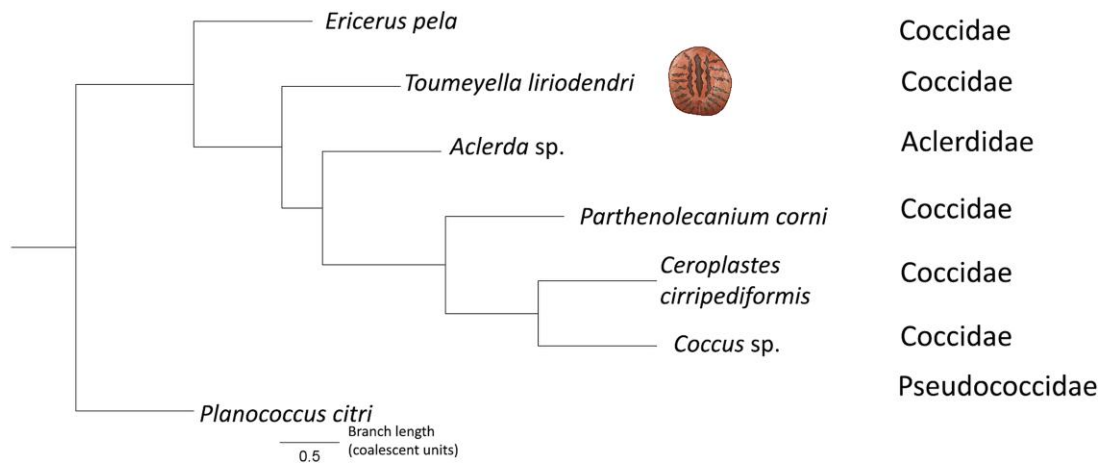


Fig. 2. Molecular phylogeny of available Coccidae species. *T. liriodendri* is an early-diverging soft scale (indicated with a graphic), with *E. pela* as the outgroup to the rest of the family. The only sequenced *Aclerda* species is nested within the family Coccidae. All branches have a local posterior probability of 1.

evidence to directly validate our gene predictions, but this lower number is in line with expectations from other insects. Moreover, the 8.6% BUSCO duplication we observe is very similar to the 8.0% observed in the new chromosome-level assembly of *P. citri* (Ross et al. 2024), which has yet to be formally annotated. Future studies of *T. liriodendri* could improve genetic resources by focusing on generating expression data and revisiting the annotation with this added line of evidence.

Coccidae phylogeny

We used available genomic resources to explore the phylogenetic relationship of sequenced coccid species. Specifically, we used conserved hemipteran BUSCO ortholog sequences to build 2,136 gene trees from which we inferred the overall species tree. We found that *E. pela* is the outgroup to other sequenced soft scales, with *T. liriodendri* being the next branching species (Fig. 2). Interestingly, we recover the unspecified *Aclerda* species from Johnson et al. (2018) within the Coccidae, despite its placement in the family Acleridae; a larger, mostly morphological phylogeny of scale insects also recovered Acleridae within Coccidae (Vea and Grimaldi 2016). This congruence of molecular and morphological results suggests that the classification of Acleridae and/or Coccidae may require revision.

Soft scale karyotype evolution

Scale insect genome architecture evolution is poorly studied, but our newly generated genomic resources can begin to make sense of the differences in chromosome number between lineages. Members of Coccidae range in karyotype from $n = 5$ to $n = 18$ chromosomes, based on currently characterized species (Gavrilov 2007), with $n = 17$ for *T. liriodendri* and $n = 9$ for *E. pela* based on inference from the genome assemblies. We explored synteny, the conservation of features along chromosomes, between the 2 species and found high levels of conservation despite the different karyotypes (Fig. 3, colored blocks). It is not worth speculating on the exact series of events that created these differences, because with only 2 chromosome-level genomes for Coccidae, either the *T. liriodendri* karyotype or the *E. pela* karyotype (or indeed both) may be more recently derived compared with the genome of their most recent common ancestor. More genomic resources will be necessary to confidently resolve this history.

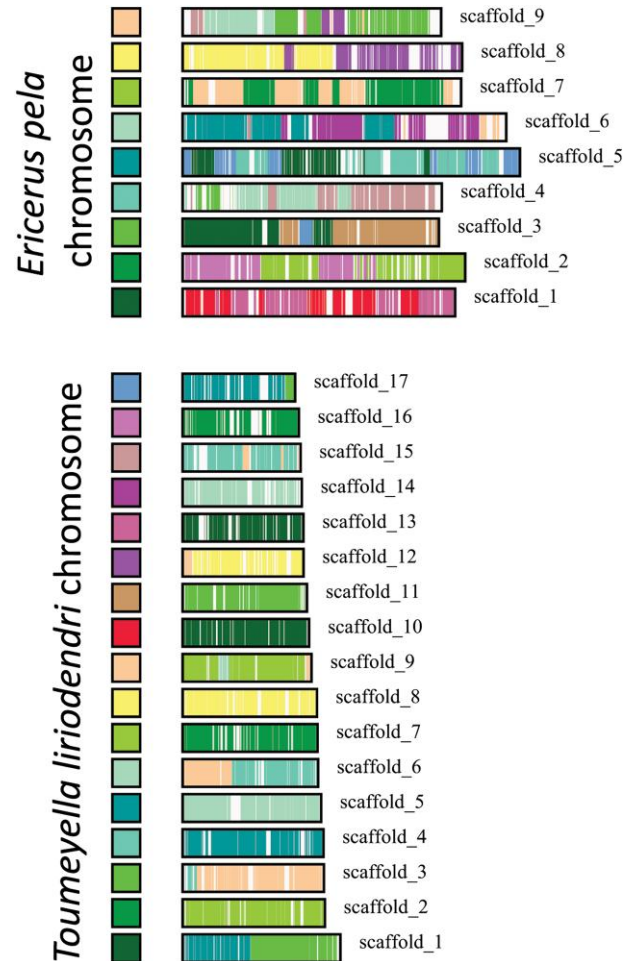


Fig. 3. Macrosynteny between the 2 chromosome-level coccid assemblies. Investigation of synteny between the *E. pela* ($n = 9$, top) and *T. liriodendri* ($n = 17$, bottom) genomes. Colored blocks on the left of each chromosome in a given panel encode the color of a syntenic match to a chromosome in the other species [e.g. *T. liriodendri* scaffold 10 (red) matches exclusively to *E. pela* scaffold 1 (dark green)]. Overall, karyotype evolution follows a pattern of chromosomal fissions: a given *T. liriodendri* chromosome matches to only one *E. pela* chromosome, but that same *E. pela* chromosome matches to multiple *T. liriodendri* chromosomes.

That said, it does not appear to be a particularly complex history based on our preliminary observations. By our count, 15 of 17 *T. liriodendri* chromosomes appear to be mainly fragments of single *E. pela* chromosomes. For instance, all of *T. liriodendri* chromosomes 8 and 12 map to 2 halves of chromosome 8 in *E. pela*, which could be explained by a simple fission event (or fusion, depending on the ancestral karyotype). The 2 exceptions, *T. liriodendri* chromosomes 1 and 6 map to 2 different *E. pela* chromosomes, indicating fusion/fission events in the opposite direction of the prevailing pattern. On top of this are more localized regions of incongruence that may represent gene trafficking or small-scale rearrangements. What drives these genomic changes between scale insect species remains an unanswered question.

This variability may be partly explained by the fact that scale insect chromosomes are holocentric, i.e. homologous pairs align along the length of the chromosome rather than at a single centromere point (Parida and Ghosh 1986; Gavrillov 2007). It has been argued that this chromosomal system is more permissive of fusions and fissions, as distinct karyotypes can still successfully align during cell division (Márquez-Corro et al. 2019); however, overall rates of chromosome number evolution do not appear to be higher in clades with holocentric taxa (Ruckman et al. 2020), and in another group of holocentric insects, the Lepidoptera, chromosome number appears to be remarkably conserved, with the exception of fusions involving the sex chromosome (Mongue et al. 2017; Wright et al. 2024). Thus, other factors must impact the propensity of chromosome number to vary between scale insect species. Indeed, in chromosomally sex-determined taxa, genomic rearrangements involving the sex chromosomes (e.g. X or Z chromosomes) may have some adaptive benefit (Mongue et al. 2022) but also come with the cost of potential dosage problems between males and females that requires the evolution of novel gene regulation (Gu et al. 2019). For a purely autosomal system such as PGE, in contrast, there is no clear adaptive advantage to chromosomal rearrangements *per se*, but the genome-wide consistency of ploidy may also make it more permissive of fusions and fissions. In either case, a broader sampling across scale insect families will be required to test these predictions.

Conclusion

Here, we report a chromosome-level assembly for the ornamental pest tuliptree scale, *T. liriodendri*. This resource will open avenues in both basic research of reproductive diversity of scale insects as well as tracking and management of this ornamental pest. To demonstrate its usefulness, we perform preliminary phylogenetic and synteny analyses with available soft scale data. A deeper exploration of the molecular mechanisms and consequences of PGE will require more data, but this genome assembly is a key component.

Data availability

All data used in this project are summarized with appropriate accessions and DOIs in Table 1 (for existing data used in comparisons) and Table 2 (newly generated sequence data, genome, annotation data). Ancillary scripts can be found at <https://github.com/amongue/ToliGenome/>.

Acknowledgments

Thanks to the University of Florida's Interdisciplinary Center for Biotechnology Research for support in the PacBio sequencing. Thanks as well to University of Florida Research Computing for

providing computational resources and support that enabled the results reported in this publication (<http://www.rc.ufl.edu>) and to the Florida Department of Agriculture and Consumer Services, Division of Plant Industry (FDACS-DPI), for supporting the taxonomy in this work.

Funding

This project was funded by University of Florida start-up funding to AJM.

Conflicts of interest

The author(s) declare no conflict of interest.

Literature Cited

- Bao W, Kojima KK, Kohany O. 2015. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 6(1):11. doi:10.1186/s13100-015-0041-9.
- Blackmon H, Ross L, Bachtrog D. 2017. Sex determination, sex chromosomes, and karyotype evolution in insects. *J Hered*. 108(1): 78–93. doi:10.1093/jhered/esw047.
- Burns DP, Donley DE. 1970. Biology of the tuliptree scale, *Toumeyella liriodendri* (Homoptera: Coccidae). *Ann Entomol Soc Am*. 63(1): 228–235. doi:10.1093/aesa/63.1.228.
- Caltagirone LE, Douthett RL. 1989. The history of the vedalia beetle importation to California and its impact on the development of biological control. *Annu Rev Entomol*. 34(1):1–16. doi:10.1146/annurev.en.34.010189.000245.
- Chen H, Lu Q, Chen X, Ling X, Liu P, Liu N, Wang W, Zhang J, Qi Q, Ding W, et al. 2021. Genome assembly and methylome analysis of the white wax scale insect provides insight into sexual differentiation of metamorphosis in hexapods. *Mol Ecol Resour*. 21(6):1983–1995. doi:10.1111/1755-0998.13376.
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 18(2):170–175. doi:10.1038/s41592-020-01056-5.
- de la Filia AG, Mongue AJ, Dorrens J, Lemon H, Laetsch DR, Ross L. 2021. Males that silence their father's genes: genomic imprinting of a complete haploid genome. *Mol Biol Evol*. 38(6):2566–2581. doi:10.1093/molbev/msab052.
- 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. doi:10.1016/j.cels.2015.07.012.
- 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. doi:10.1073/pnas.1921046117.
- García Morales M, Denno BD, Miller DR, Miller GL, Ben-Dov Y, Hardy NB. 2016. ScaleNet: a literature-based model of scale insect biology and systematics. *Database*. 2016:bav118. doi:10.1093/database/bav118.
- Gavrillov IA. 2007. A catalog of chromosome numbers and genetic systems of scale insects (Homoptera: Coccinea) of the world. *Isr J Entomol*. 37:1–45.
- Gill R. 1988. The Scale Insects of California: Part 1. The Soft Scales (Homoptera: Coccoidea: Coccidae). Sacramento (CA): California Department of Food & Agriculture. p. 132.
- Gmelin JF. 1790. *Systema Naturae*. Editio decima tertia. Acuta, reformata, cara. Lipsiae 1517–2224.

- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 29(7):644–652. doi:10.1038/nbt.1883.
- Grabherr MG, Russell P, Meyer M, Mauceli E, Alföldi J, Di Palma F, Lindblad-Toh K. 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics.* 26(9): 1145–1151. doi:10.1093/bioinformatics/btq102.
- Gu L, Reilly PF, Lewis JJ, Reed RD, Andolfatto P, Walters JR. 2019. Dichotomy of dosage compensation along the neo Z chromosome of the monarch butterfly. *Curr Biol.* 29(23):4071–4077.e3. doi:10.1016/j.cub.2019.09.056.
- Hamon AB, Williams ML. 1984. *The Soft Scale Insects of Florida* (Homoptera: Coccoidea: Coccidae). Sacramento (CA): Florida Department of Agriculture and Consumer Services, Division of Plant Industry.
- Holst F, Bolger A, Günther C, Maß J, Triesch S, Kindel F, Kiel N, Saadat N, Ebenhöf O, Usadel B, et al. 2023. Helixer—de novo prediction of primary eukaryotic gene models combining deep learning and a hidden Markov model. *bioRxiv* 2023-02. doi:10.1101/2023.02.06.527280.
- Hughes-Schrader S. 1925. Cytology of hermaphroditism in *Icerya purchasi* (Coccidae). *Z Zellforsch Mikrosk Anat.* 2(2):264–290. doi:10.1007/BF00376393.
- 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.
- Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics.* 8(3):275–282. doi:10.1093/bioinformatics/8.3.275.
- Kondo T, Williams DJ. 2008. Neotype designation and redescription of *Toumeyella lirioidendri* (Gmelin) (Hemiptera: Coccoidea: Coccidae). *J Insect Sci.* 8(56):1–6. doi:10.1673/031.008.5601.
- Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Res.* 6:1287. doi:10.12688/f1000research.12232.1.
- Li M, Tong H, Wang S, Ye W, Li Z, Omar MAA, Ao Y, Ding S, Li Z, 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.
- Manni M, Berkeley MR, Seppely M, Simão 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.
- Marcais G, Kingsford C. 2012. Jellyfish: a fast k-mer counter. *Tutorialis e Manuais.* 1:1–8. doi: 10.1093/bioinformatics/btr011.
- Márquez-Corro JI, Martín-Bravo S, Pedrosa-Harand A, Hipp AL, Luceño M, Escudero M. 2019. Karyotype evolution in holocentric organisms. *eLS.* 1–7. doi: 10.1002/9780470015902.a0028758.
- Miller GL, Williams ML. 1995. Systematic analysis of the adult males of *Toumeyella* group, including *Mesolecanium nigrofasciatum*, *Neolecanium cornuparvum*, *Pseudophilippia quaintancii*, and *Toumeyella* spp. (Homoptera: Coccidae). *Contrib Am Entomol Inst.* 28: 68.
- Mongue AJ, Hansen ME, Walters JR. 2022. Support for faster and more adaptive Z chromosome evolution in two divergent lepidopteran lineages. *Evolution.* 76(2):332–345. doi:10.1111/evo.14341.
- Mongue AJ, Michaelides S, Coombe O, Tena A, Kim D-S, Normark BB, Gardner A, Hoddle MS, Ross L. 2021. Sex, males, and hermaphrodites in the scale insect *Icerya purchasi*. *Evolution.* 75(11): 2972–2983. doi:10.1111/evo.14233.
- Mongue AJ, Nguyen P, Volenikova A, Walters JR. 2017. Neo-sex chromosomes in the monarch butterfly, *Danaus plexippus*. *G3* (Bethesda). 7:3281–3294. doi:10.1534/g3.117.300187.
- Mongue AJ, Ross L, Watson G; Darwin Tree of Life Consortium. 2024. The genome sequence of the cottony cushion scale, *Icerya purchasi* (Maskell, 1879). *Wellcome Open Res.* 9:21. doi:10.12688/wellcomeopenres.20653.1.
- Morrison H. 1925. Classification of scale insects of the subfamily ortheziinae. *J Agric Res.* 30:97.
- Novoa NM, Hamon A, Evans G, Kondo T, Oliver PH, Marrero AH, Alonso AA. 2011. The cocoideos (Hemiptera: Sternorrhyncha: Coccoidea) present in the Cordillera de Guaniguanico, Pinar del Rio, Cuba, and the relationship with their hosts. (Los cocoideos (Hemiptera: Sternorrhyncha: Coccoidea) presentes en la Cordillera de Guaniguanico, Pinar del Rio, Cuba, y la relación con sus hospedantes.) *Insecta Mundi*, No.0183:1-25. <http://fulltext10.fcla.edu/cgi/t/text/text-idx?c=mundi;cc=mundi;idno=SN07496737;view=text;rgn=main;a=33>.
- Nur U. 1980. Evolution of unusual chromosome systems in scale insects (Coccoidea: Homoptera). In: Blackman RL, Hewitt GM, Ashburner M, editors. *Insect Cytogenetics*. Oxford, UK: Blackwell Scientific Publications. p. 97–117.
- Parida B, Ghosh D. 1986. Somatic chromosome complements of nine armored scale insects (Homoptera: Coccoidea: Diaspididae) of India and their karyological relationship. *Caryologia.* 39: 335–345. doi:10.1080/00087114.1986.10797795.
- Picard toolkit. 2019. Broad Institute. GitHub Repository. <https://broadinstitute.github.io/picard/>.
- Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol.* 26:1641–1650. doi:10.1093/molbev/msp077.
- Ross L, Langenhof MBW, Pen I, Beukeboom LW, West SA, Shuker DM. 2010. Sex allocation in a species with paternal genome elimination: the roles of crowding and female age in the mealybug *Planococcus citri*. *Evol Ecol Res.* 12:89–104.
- Ross L, Mongue AJ, De La Filia A; Darwin Tree of Life Consortium. 2024. The genome sequence of the citrus mealybug, *Planococcus citri* (Risso, 1913). *Wellcome Open Res.* 9:22. doi:10.12688/wellcomeopenres.20686.1.
- Ross L, Mongue AJ, Hodson CN, Schwander T. 2022. Asymmetric inheritance: the diversity and evolution of non-Mendelian reproductive strategies. *Annu Rev Ecol Evol Syst.* 53:1–23. doi:10.1146/annurev-ecolsys-021822-010659.
- Ruckman SN, Jonika MM, Casola C, Blackmon H. 2020. Chromosome number evolves at equal rates in holocentric and monocentric clades. *PLoS Genet.* 16:e1009076. doi:10.1371/journal.pgen.1009076.
- Sayyari E, Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol Biol Evol.* 33: 1654–1668. doi:10.1093/molbev/msw079.
- Smit A, Hubley R, Green P. 2019. 2013–2015. RepeatMasker Open-4.0. <http://www.repeatmasker.org>.
- Stamatakis A. 2006. Phylogenetic models of rate heterogeneity: a high performance computing perspective. In: *Proceedings of the 20th IEEE International Parallel & Distributed Processing Symposium (IPDPS2006)*. Washington (DC): IEEE Computer Society Press. p. 278–286.
- Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: *IEEE International Parallel and Distributed Processing Symposium.* 314–324. doi: 10.1109/IPDPS.2019.00041.
- Veja IM, de la Filia AG, Jaron KS, Mongue AJ, Ruiz-Ruano FJ, Barlow SEJ, Nelson R, Ross L. 2021. The B chromosome of *Pseudococcus*

- viburni: a selfish chromosome that exploits whole-genome meiotic drive. bioRxiv 2021-08. doi:[10.1101/2021.08.30.458195](https://doi.org/10.1101/2021.08.30.458195).
- Vea IM, Grimaldi DA. 2016. Putting scales into evolutionary time: the divergence of major scale insect lineages (Hemiptera) predates the radiation of modern angiosperm hosts. *Sci Rep.* 6:23487. doi:[10.1038/srep23487](https://doi.org/10.1038/srep23487).
- Waterhouse D. 1991. Possibilities for the biological control of the breadfruit mealybug, *Icerya aegyptiaca*, on Pacific Atolls. *Micronesica.* 117–122.
- Wright CJ, Stevens L, Mackintosh A, Lawniczak M, Blaxter M. 2024. Comparative genomics reveals the dynamics of chromosome evolution in Lepidoptera. *Nat Ecol Evol.* 8:777–790. doi:[10.1038/s41559-023-02311-6](https://doi.org/10.1038/s41559-023-02311-6).
- Yang P, Yu S, Hao J, Liu W, Zhao Z, Zhu Z, Sun T, Wang X, Song Q. 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:giz113. doi:[10.1093/gigascience/giz113](https://doi.org/10.1093/gigascience/giz113).
- Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics.* 19:153. doi:[10.1186/s12859-018-2129-y](https://doi.org/10.1186/s12859-018-2129-y).
- Zhou C, McCarthy SA, Durbin R. 2023. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics.* 39:btac808. doi:[10.1093/bioinformatics/btac808](https://doi.org/10.1093/bioinformatics/btac808).

Editor: K. Vogel