Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Chromosome-scale genome of a twig-girdler longhorn beetle reveals clustered genomic organization of herbivory-linked functions

Diego de S. Souza
dsouza@fieldmuseum.org
Field Museum of Natural History
Tatiana A. Sepulveda
Field Museum of Natural History
Rowan L. K. French
University of Toronto
Bruno A. S. de Medeiros
Field Museum of Natural History
José O. Silva Júnior
Universidade Estadual de Montes Claros
Dayanna do N. Machado
Instituto Federal do Mato Grosso
Terrence Sylvester
University of Memphis
Luciane Marinoni
Federal University of Paraná
Eugenio H. Nearns
Field Museum of Natural History
Marcela L. Monné
National Museum, Federal University of Rio de Janeiro
Felix A. H. Sperling
University of Alberta

Research Article

Keywords: Cerambycidae, chemosensory genes, detoxification, tandem gene clusters, transcriptomics, host-plant interactions, evolution, gene expression, Acacia

Posted Date: August 13th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10013748/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Herbivorous insects have diversified through repeated interactions with host plants. However, the genomic organization and transcriptional deployment that enables host use and tolerance of chemically defended tissues remain poorly characterized in many phytophagous lineages. We present a chromosome-scale genome assembly and multi-tissue, multi-stage transcriptomes for the Neotropical twig-girdler *Oncideres impluviata* (Germar) (Cerambycidae: Lamiinae). This pest species girdles living branches prior to oviposition and exploits at least 21 host species across five plant families. The assembly comprises 474.7 Mb (N50 = 39.4 Mb; BUSCO = 99.4%), with telomeric and synteny evidence supporting organization into 10 chromosomes, consistent with the conserved Lamiinae karyotype. Comparative analyses across phytophagous beetles reveal expansions in chemosensory (including a 40-gene Obp56d tandem cluster), neural signaling, detoxification, and digestive gene families, with Gene Ontology enrichment confirming bias toward host-interface functions. Across stages and tissues, host-interface expression concentrates in adults, with chemosensory genes in antennae and detoxification genes partitioned between antennae (odorant clearance) and thorax (systemic processing); sex bias is limited despite behavioral dimorphism. At the chromosomal level, 178 tandem clusters show significant overrepresentation of differentially expressed genes, and seven multi-family superloci co-localize clusters from different gene families, suggesting coordinated deployment of complementary host-interface functions. These results identify clustered genome architecture as a plausible substrate for rapid host evolution: tandem arrays and multi-family superloci concentrate standing variation in genomic regions where dosage and regulatory change can mediate host breadth, host shifts, and complex plant-manipulation behavior. The genome and transcriptome resources reported here support future comparative genomics in phytophagous beetles and management of twig-girdling pests.

INTRODUCTION

Herbivorous insects have diversified extensively through evolutionary interactions with host plants, yet the molecular mechanisms enabling host plant selection and colonization of chemically defended hosts remain incompletely understood [1–5]. Transitions to novel hosts require coordinated adaptations across multiple functional systems, including sensory perception for host recognition, detoxification machinery for processing plant defenses, and digestive enzymes for nutrient extraction [6–10]. How these systems evolve and are deployed across life stages is a central question in evolutionary biology, with implications for understanding ecological specialization and for managing pest insects [11–13].

Longhorn beetles (Cerambycidae) illustrate both the scale and the mechanistic complexity of plant feeding evolution. The family comprises more than 39,000 described species [14], feeding on nearly all major woody plant lineages and exhibiting diverse larval strategies associated with xylem and cambium [15]. Despite this diversity, genomic resources remain sparse across Cerambycidae relative to their ecological and economic importance [10, 16–19]. This gap is particularly pronounced for Neotropical lineages, including the tribe Onciderini, in which host plant use is coupled to a distinctive reproductive behavior. Females of several species girdle living branches by severing the cambium, killing distal

tissues and creating oviposition sites where larvae subsequently develop [20]. This behavior, which has multiple origins in Onciderini [21], converts living, chemically defended tissue into a dead-wood substrate, reducing inducible plant defenses and facilitating microbial decomposition that can influence larval nutrition and performance [22].

Among onciderines, *Oncideres impluviata* [23] is a suitable system for investigating the genomic foundations of host breadth, host shifts, and girdling-associated life history. Native to the South American Atlantic Forest, this species is a damaging pest in southern Brazil and has successfully colonized black-wattle trees (*Acacia mearnsii* De Wild.) [24], a host native to Australia and introduced to Brazil in the early twentieth century [25, 26]. Black-wattle is characterized by high concentrations of tannins and related phenolics that deter many generalist herbivores [27], and its exploitation by *O. impluviata* highlights the capacity of this beetle to colonize new chemically defended hosts. More broadly, *O. impluviata* has been reported feeding on 21 host species across five plant families (Aquifoliaceae, Caesalpiniaceae, Euphorbiaceae, Fabaceae, and Myrsinaceae [28]), indicating substantial host breadth and providing a framework for examining genomic features associated with flexible host use.

Based on the well-documented ecology of *Oncideres impluviata*, we can make several testable predictions about the organization and expression of herbivory-linked genes in this species. First, we expect chemosensory and detoxification genes to be more strongly expressed in adult females than adult males, assuming the costs of poor host choices are especially high for females. This assumption has a clear biological basis: although both sexes feed on live plant tissue and likely use olfactory cues to locate hosts, females differ in that they actively choose the plant in which their larvae develop, and they also have sustained interactions with plant defenses during girdling. Thus, host-related chemosensory and detoxification functions may be more important for females than males, resulting in sexually dimorphic gene expression. In addition to this predicted sex difference, we expect an expression difference between life stages, because larvae and adults differ in how they interact with host plants [22, 29–31]. Specifically, larvae should show biased expression of digestion-related genes, relative to adults, because larval development occurs in dead, girdled tissues that are only weakly chemically defended, but are physically difficult to digest [32].

Finally, regarding genome organization, we expect herbivory-linked loci to be spatially clustered on chromosomes, possibly into tandem arrays and/or groupings of multiple gene families, like the superloci (or “supergenes”) underlying other complex traits [33] Because plant-feeding involves coordinated interactions among functionally divergent genes, physical gene linkage may be adaptive, enabling suites of plant-feeding traits to be co-inherited and expressed using shared regulatory mechanisms. Such coordination could be key in allowing species like *O. impluviata* to readily adapt to new hosts, despite the potential novelty of those plants’ toxins and scent profiles.

Here we combine a chromosome-scale genome assembly with transcriptomes from multiple tissues and developmental stages for the twig-girdler longhorn beetle *O. impluviata* to characterize, and test

hypotheses about, the diversity, spatial organization, and expression of herbivory-linked genes in this species. We show that chemosensory, detoxification, and digestive gene families are expanded and frequently organized in tandem chromosomal clusters, and that these repertoires exhibit strong tissue-specific deployment in adults, consistent with the ecological demands placed on this life stage during host selection and girdling. Developmental profiles further suggest stage-structured deployment of functions linked to interactions with host plants across the life cycle of *O. impluviata*. Together, these data describe the chromosomal architecture and adult functional genomics of a distinctive plant-feeding beetle. Our results provide a framework for understanding host breadth and host shifts in phytophagous beetles, alongside resources that may contribute to both comparative analyses and the development of targeted pest management approaches.

MATERIAL AND METHODS

1. Sample collection

Specimens of *O. impluviata* were collected from fallen, girdled branches of black-wattle (*Acacia mearnsii* De Wild.) at multiple localities in Rio Grande do Sul, Brazil, during 19–21 October 2021. Adults were collected directly in the field or obtained by rearing field-collected larvae under ambient laboratory conditions (25°C). All specimens were preserved in RNA*later* (Invitrogen) and stored at -80°C until processing. In total, seven individuals (larva and adults of both sexes) were used across genome and transcriptome work (Table S1).

2. DNA extraction and genome sequencing

High-molecular-weight genomic DNA was extracted from a single adult male *O. impluviata* using the MagAttract HMW DNA Kit (Qiagen). Tissue from the head and antennae was used for extraction, with the digestive tract removed to minimize microbial contamination. DNA quantity, purity, and fragment size distribution were assessed using Qubit fluorometry, Nanodrop spectrophotometry, and a Fragment Analyzer. A PCR-free library was prepared following the standard PacBio HiFi protocol using 5 µg of input DNA sheared to a 20 kb target, and whole-genome sequencing was performed on a single SMRT Cell using the PacBio Sequel II platform.

3. Genome assembly and chromosome-scale inference

The haploid genome size of *O. impluviata* was estimated using k-mer frequency analysis. K-mers (k = 21) were counted with Jellyfish v2.3.1 [34], and the resulting histogram was analyzed with GenomeScope v2.0 [35]. HiFi reads were assembled with Hifiasm v0.16.1 [36], generating a primary contig assembly that was converted to FASTA format using any2fasta (https://github.com/tseemann/any2fasta). Haplotypic duplications were identified and removed with Purge_dups v1.2.6 [37] to obtain a haploid genome assembly. Assembly statistics were computed with QUAST v5.2.0 [38], and genome

completeness was assessed using BUSCO v5.2.2 [39] with the insecta_odb10 lineage dataset. Contigs were sorted by size and renamed sequentially.

Chromosome-scale inference was based on the identification and localization of telomeric repeats using TIDK v0.2.41 [40]. Simple repetitive motifs were identified using tidk explore --minimum 5 --maximum 12, which detected the canonical insect telomeric repeat (TTAGG/AACCT). Telomeric repeat arrays were then localized across the genome using tidk search and mapped to scaffold termini. The presence of telomeric repeats at both ends of scaffolds was used to identify candidates corresponding to complete chromosome arms.

Putative chromosomes were further evaluated through synteny analysis using minimap2, assuming 5% sequence divergence, aligning the *O. impluviata* genome to the genome of the Japanese pine sawyer longhorn beetle *Monochamus alternatus* Hope, 1842 (GenBank accession GCA_037114965.1). This species was selected because it is a high-quality assembly produced with Hi-C with a haploid chromosome number of 10, typical of many Lamiinae species [41, 42], and because it belongs to Lamiini — the same tribe as *Anoplophora glabripennis* and the closest relative of Onciderini among the available genomes [43]. No annotated cerambycid genome is currently assembled to chromosome scale, so the unannotated *M. alternatus* assembly was used for the synteny comparison.

Alignments were performed against the unmasked *O. impluviata* genome, retaining only hits longer than 100 bp with mapping quality ≥ 30. For visualization, alignments were grouped by proximity, collapsing 20 links within 1 Mb window into single connections using the bundlelinks program in circos v0.23 [44]. Synteny plots were generated with the R package circlize v0.4.17 [45]. Scaffolds without telomeric evidence or clear syntenic assignment were retained as unplaced scaffolds and excluded from chromosome-scale visualizations and downstream chromosome-based summaries.

4. RNA extraction and transcriptome sequencing

RNA-seq was generated from adult antennae (two males and two females) and adult thorax (one male and two females), and from one larval individual sampled for head and thorax (Table S1). Total RNA was extracted using the Direct-zol RNA MiniPrep Kit (Zymo Research), with tissues homogenized directly in TRI Reagent, following the manufacturer's protocol. RNA quantity and purity were assessed by spectrophotometry and fluorometry prior to sequencing. Paired-end libraries (2 × 150 bp) were prepared and sequenced on an Illumina NovaSeq 6000 platform, targeting approximately 20 million read pairs per sample.

Raw reads were processed with fastp v1.0.1 [46] to remove adapters and low-quality bases, using default parameters. Quality-filtered reads were assembled *de novo* with Trinity [47] to generate sample-specific transcriptome assemblies used as evidence for genome annotation and gene model refinement (including validation of exon–intron boundaries and refinement of untranslated regions) and for downstream differential expression analyses.

5. Gene prediction and annotation

Repeat identification and masking

Genome-wide repeats were identified using a combined *de novo* and homology-based approach. A *de novo* consensus repeat sequence database was first generated with RepeatModeler (https://github.com/Dfam-consortium/RepeatModeler) using default parameters. This de novo library was then combined with a curated set of Coleoptera-specific repeats from the Dfam 3.8 Repbase 20181026 databases [48, 49] to create a custom repeat library. Repetitive elements were subsequently annotated and soft-masked with RepeatMasker v4.1.8 using rmblast as the search engine (http://www.repeatmasker.org).

Gene prediction

Gene prediction was performed with Funannotate v1.8.17 [50], integrating *ab initio* predictions, protein homology, and transcript evidence. *Ab initio* gene models were generated using Augustus [51], GeneMark-ES [52], SNAP [53], GlimmerHMM [54], and CodingQuarry [55]. Protein homology evidence was incorporated using Exonerate [56] with reference sequences from UniProtKB/Swiss-Prot [57] (Table S2).

Transcript evidence included both the *de novo* transcript assemblies generated with Trinity and genome-guided assemblies produced by aligning RNA-seq reads to the genome with HISAT2 [58] and StringTie [59]. Transcript alignments were performed with GMAP [60] and BLAT [61], and gene structures were refined using PASA [62]. Consensus gene models were generated with EvidenceModeler (EVM [63]), which integrates all evidence sources into a unified gene set.

Functional annotation and feature inference

Functional annotation was performed with Funannotate v1.8.17 [50] using protein domain, homology, and functional evidence from Pfam [64], InterPro [65], eggNOG/COG [66, 67], Gene Ontology [68], MEROPS [69], CAZy [70], and UniProtKB [71]. Protease families were assigned by mapping InterPro peptidase signatures to MEROPS identifiers, and carbohydrate-active enzymes (CAZymes) were inferred from CAZy/dbCAN-related domains captured via InterPro and eggNOG. Subcellular localization was inferred from signal peptide predictions (SignalP or corresponding InterPro signatures) and transmembrane helix predictions (TMHMM or InterPro transmembrane signatures), enabling identification of putative secreted and membrane-associated proteins. To reduce false positives, gene models lacking support from transcript or protein evidence were removed to produce the final curated gene set. The curated gene set was used for downstream analyses. Annotation completeness before and after filtering was assessed with BUSCO using the insecta_odb10 dataset.

6. Comparative genomics analyses

Comparative genomic analyses were conducted to place the *O. impluviata* genome in a broader context within Phytophaga (Chrysomeloidea and Curculionoidea) and to identify lineage-specific gene family expansions and functional enrichments. We analyzed *O. impluviata* alongside 24 annotated beetle genomes representing Cerambycidae, Chrysomelidae, Curculionoidea, and the outgroup *Tribolium castaneum* (Herbst, 1797) (Tenebrionidae). Protein sets were obtained from the *O. impluviata* annotation and from NCBI reference genomes (Table S2). To ensure comparability, only the longest isoform per gene was retained.

Orthology inference was performed with OrthoFinder v2.5.5 [72] using DIAMOND [73] for all-versus-all similarity searches and MSA mode for orthogroup construction. Single-copy orthologs present in all taxa were extracted, aligned with MAFFT v7.520 [74], trimmed with trimAl v1.4 [75], and concatenated with AMAS v1.0 [76] to generate a supermatrix for phylogenetic inference in IQ-TREE v2.2.2 [77] under the LG + G4 model. Node support was assessed using 1,000 ultrafast bootstrap replicates and 1,000 SH-aLRT tests, and the tree was rooted with *T. castaneum*.

Orthogroups were classified into universal single-copy (1:1:1), multicopy universal (N:N:N), species-specific, lineage-specific (Cerambycidae, Chrysomelidae, Curculionoidea), Phytophaga-specific, other shared orthologs, and unassigned genes. Gene family expansions were inferred from orthogroup sizes by comparing *O. impluviata* against four reference sets: (1) *Anoplophora glabripennis* (Motschulsky, 1853), (2) other Cerambycidae species (excluding *O. impluviata*), (3) Chrysomelidae species, and (4) Curculionoidea species. Orthogroups were considered expanded when *O. impluviata* contained more genes than the maximum observed in the corresponding reference set, using configurable fold-change and minimum gene-difference thresholds (default: >1.0-fold and ≥ 1 gene).

Gene-to-GO term associations were obtained from eggNOG-mapper annotations, and the background set comprised all *O. impluviata* genes with GO annotations after identifier normalization. Gene Ontology (GO) enrichment analyses were performed with GOATOOLS v1.3.1 [78] using Fisher's exact test, propagation of gene counts through the GO hierarchy (propagate_counts = True), and Benjamini–Hochberg correction for multiple testing (FDR). Terms with FDR < 0.05 were considered significant. Analyses were conducted separately for each GO namespace (Biological Process, Molecular Function, Cellular Component) and comparative baseline, and GO ontology data (go-basic.obo) were obtained from the Gene Ontology Consortium (version 2024-01-17).

7. Transcriptomic profiling and gene family analysis

Differential expression analyses

Differential expression analyses were performed using DESeq2 v1.38 [79]. RNA-seq reads were quantified with Salmon v1.10 [80] using a decoy-aware transcriptome index constructed from predicted transcripts and the genome assembly. Quantification used automatic library type detection, sequence-specific bias correction, GC bias correction, and positional bias modeling. Transcript-level abundances

were summarized to gene-level counts using tximport [81] in R. Size factor normalization was applied to account for differences in library size, and differential expression was tested using the Wald test with Benjamini–Hochberg correction for multiple comparisons [82]. Genes were considered significantly differentially expressed at adjusted $p < 0.05$. Comparisons were performed among developmental stages (larva vs adult), sexes (male vs female), and tissues (antennae vs thorax), following the sampling design as shown in Table S1).

Target gene families and membership assignment

To characterize expression patterns associated with overrepresented functional categories identified by GO enrichment analyses, we curated and analyzed gene families involved in host-interface functions. These included chemosensory genes (odorant-binding proteins [OBPs], chemosensory proteins [CSPs], odorant receptors [ORs], gustatory receptors [GRs], ionotropic receptors [IRs], and sensory neuron membrane proteins [SNMPs]), detoxification enzymes (cytochrome P450s, glutathione S-transferases [GSTs], UDP-glucuronosyltransferases [UGTs], carboxyl/cholinesterases [CCEs], and ABC transporters), and digestive enzymes involved in plant tissue processing (serine [SerPro], cysteine [CysPro], aspartic [AspPro], and metalloproteases [MetPro], and carbohydrate-active enzymes [CAZymes]). Gene family membership was assigned using diagnostic Pfam domain annotations, MEROPS protease classifications, and CAZy assignments extracted from the genome annotation (Table S3). Target genes annotated as hypothetical proteins within target families were further evaluated by BLAST searches against Swiss-Prot, NCBI nr, and the curated beetle proteome database used for protein homology evidence during gene prediction to confirm family membership.

Family-level expression was calculated as the mean normalized expression across all genes within each family to represent typical per-gene deployment while reducing dependence on family size, enabling comparisons among families with different numbers of members. Differential expression within target gene families was assessed by filtering DESeq2 results to include only genes assigned to chemosensory, detoxification, or digestion families, and applying the same statistical thresholds used for genome-wide analyses.

8. Chromosomal organization of host-interface genes

Chromosome-scale positions for target gene family members were determined by extracting gene coordinates from the GFF3 annotation and converting scaffold-relative positions to chromosome-scale coordinates using scaffold-to-chromosome assignments derived from synteny analysis with *M. alternatus*. Scaffold orientation was determined from telomere positions (5' and 3' telomeric repeat arrays), and scaffolds were reversed where necessary to maintain consistent chromosome polarity. Cumulative coordinates were then calculated to map genes to chromosome-scale positions.

Tandem gene clusters were defined as two or more genes from the same family located within 100 kb on the same chromosome. Superloci were defined as regions where clusters from different gene

families co-localized within 50 kb, potentially indicating shared genomic coordinated regulation. Gene density was calculated using 500 kb sliding windows across each chromosome.

RESULTS

1. Genome sequencing

PacBio sequencing yielded 25.36 Gb of HiFi circular consensus reads, corresponding to approximately 53.4× genome coverage (Table 1). K-mer frequency analysis of the HiFi reads with GenomeScope v2.0 (k = 21; diploid model) estimated a haploid genome size of ~ 408 Mb, including ~ 107 Mb of repetitive sequence, and a heterozygosity rate of ~ 2.07%. The inferred read error rate was 0.18%, and model fit ranged from 76.0% to 98.5% (Table S4; Supplementary file 1).

2. Genome assembly and chromosome-scale inference

The final genome assembly comprised 29 scaffolds with a total length of 474.7 Mb and GC content of 32.5%. The assembly was highly contiguous (scaffold N50 = 39.4 Mb; L50 count = 6; BUSCO completeness 99.4%), suggesting overall high integrity of the genome (Table 1). We detected the canonical insect telomeric motif at both ends of scaffolds 1, 4, 7, 8, and 11, consistent with telomere-to-telomere assembly for these five scaffolds (Fig. 1a). Eight additional large scaffolds carried telomeric arrays at a single terminus (scaffolds 2, 3, 5, 6, 9, 10, 12, and 13), consistent with near-complete chromosomes lacking one telomere (scaffolds 3 and 6) or chromosome arms represented by scaffold pairs (scaffolds 2 + 12, 5 + 9, 10 + 13) (Fig. 1a).

Table 1
Summary of sequencing, assembly, and annotation statistics for the Oncideres impluviata genome.

PacBio sequencing data statistics	
Number of reads	**1,724,749**
Total bases (bp)	25,358,460,417
Total bases (Gb)	25.36
Mean read length (bp)	14,703
Read length N50 (bp)	14,561
Minimum read length (bp)	40
Maximum read length (bp)	51,365
% bases with Q ≥ 20	98.8%
% bases with Q ≥ 30	97.4%
GC content (%)	32.7%
Mean mapped depth (×)	53.4
Assembly statistics	
Number of scaffolds	29
Assembly size (bp)	474,744,405
Scaffold N50 (bp)	39,412,979
Longest scaffold (bp)	56,020,172
Shortest scaffold (bp)	16,256
GC content (%)	32.49
Number of ambiguous bases (N)	184
Genome completeness (BUSCO v. 5.3.2, insecta_odb10, total BUSCOs: 1,367)	
Complete BUSCOs	99.4% (1,359)
Single-copy BUSCOs	98.9% (1,352)
Duplicated BUSCOs	0.5% (7)
Fragmented BUSCOs	0.2% (3)
Missing BUSCOs	0.4% (5)
Chromosomes statistics	

PacBio sequencing data statistics	
Number of reads	**1,724,749**
Number of complete chromosomes	5
Chromosome 3 (bp)	56,020,172
Chromosome 5 (bp)	42,955,755
Chromosome 8 (bp)	37,712,326
Chromosome 9 (bp)	31,820,999
Chromosome 10 (bp)	26,561,689
Number of fragmented chromosomes	8
Chromosome 1 (bp)	71,364,639
Chromosome 2 (bp)	69,491,285
Chromosome 4 (bp)	43,747,616
Chromosome 6 (bp)	41,847,146
Chromosome 7 (bp)	39,412,979
Unplaced scaffolds	16

Whole-genome synteny analysis against *M. alternatus* revealed extensive macrosynteny. Seven *O. impluviata* scaffolds (1, 3, 4, 6, 7, 8, and 11) showed one-to-one correspondence with individual *M. alternatus* chromosomes, and several scaffold pairs (3 and 10, 5 and 9, and 2 and 12) mapped to the same *M. alternatus* chromosome (Fig. 1b). Together, telomeric signals and conserved macrosynteny indicate that most of the largest scaffolds represent complete or nearly complete chromosomes.

3. Gene prediction and functional annotation

Repeat annotation combining *de novo* and homology-based approaches identified 49.7% of the *O. impluviata* genome as repetitive (Table S5). Transposable elements (TEs) accounted for most of this fraction (50.77%; ~241 Mb), dominated by unclassified interspersed repeats (28.62%), LINEs (7.6%), DNA transposons (5.3%), and LTR retrotransposons (8.7%) (Fig. S1a). The repeat landscape of *O. impluviata* differed substantially depending on whether unclassified repeats were included. Classified repeats alone showed a pronounced low-divergence peak near Kimura distance 0, consistent with recent or ongoing TE activity (Fig. S1b). Including unclassified repeats, increased coverage across all divergence categories, with elevated signal at intermediate Kimura distances (approximately 5 and 17) and additional accumulation at higher divergence (approximately 30 and 42) (Fig. S1c).

The initial annotation of *O. impluviata* predicted 38,368 protein-coding gene models based on *ab initio* predictions, transcript evidence, and homology alignments. Functional annotation assigned putative functions to a majority of genes: 12,236 (52.8%) with Pfam domains, 12,588 (54.3%) with GO terms, 15,639 (67.5%) with InterPro annotations, and 19,364 (83.6%) with eggNOG orthologous groups, along with 916 predicted proteases (MEROPS) and 413 carbohydrate-active enzymes (CAZymes) (Tables S6, S7). After filtering to retain only models supported by RNA-seq evidence or protein homology, the curated annotation comprised 23,161 high-confidence protein-coding genes. BUSCO completeness remained at 95.2% for both gene sets, indicating that filtering reduced gene count without compromising annotation coverage (Table S8). The curated gene set was submitted to NCBI (BioProject PRJNA1256903).

4. Gene orthology and comparative genomic analyses

A total of 47 single-copy conserved orthologs shared by the Phytophaga species were used for the phylogenetic analysis, which recovered *O. impluviata* within Cerambycidae, with *A. glabripennis* as the closest relative (Fig. 2). OrthoFinder assigned 20,516 *O. impluviata* genes to 11,695 orthogroups (of 23,451 total across all species; Table S9). Orthogroup size comparisons identified 9,967 *O. impluviata* genes in expanded orthogroups relative to *A. glabripennis*, 7,069 relative to other Cerambycidae, 4,927 relative to Chrysomelidae, and 7,264 relative to Curculionoidea (Table S10).

GO enrichment analyses were performed on the functionally annotated subsets of expanded genes (4,131, 2,657, 1,529, and 2,468 genes, respectively; Fig. 3), yielding 1,216–1,850 significantly enriched GO terms per comparison (FDR < 0.05; Table S11; Supplementary file 2). Within Cerambycidae, expanded orthogroups were consistently enriched for neural and sensory functions. Relative to *A. glabripennis* (4,131 genes; 1,746 enriched terms), the strongest signals included detection of chemical stimulus (1.95-fold, FDR = 7.98×10^{-19}), nervous system process (1.33-fold, FDR = 1.90×10^{-18}), G protein-coupled receptor activity (1.79-fold, FDR = 3.54×10^{-13}), and neurotransmitter receptor activity (1.95-fold, FDR = 1.05×10^{-12}). Relative to other Cerambycidae (2,657 genes; 1,850 enriched terms), sensory perception of smell was prominently enriched (2.33-fold, FDR = 2.18×10^{-19}), alongside fatty acid biosynthesis (2.64-fold) and terpenoid metabolism (2.78-fold).

Comparisons with other Phytophaga clades instead highlighted digestive and detoxification functions. Relative to Chrysomelidae (1,529 genes; 1,460 enriched terms), serine-type endopeptidase activity (4.04-fold, FDR = 1.18×10^{-34}), odorant binding (5.33-fold, FDR = 7.34×10^{-29}), and peptidase activity (2.37-fold, FDR = 1.83×10^{-22}) dominated the enrichment profile. Relative to Curculionoidea (2,468 genes; 1,216 enriched terms), the strongest enrichments were in oxidative metabolism: alkane 1-monooxygenase activity (4.63-fold, FDR = 2.02×10^{-31}), monooxygenase activity (3.48-fold, FDR = 3.54×10^{-31}), and oxidoreductase activity acting on paired donors (4.48-fold, FDR = 5.12×10^{-30}).

5. Transcriptomic profiling and gene family analyses

To examine how expanded gene repertoires are deployed across development and adult tissues, we analyzed RNA-seq data spanning two developmental stages (larva, adult), both sexes, and two adult tissues (antennae, thorax). PCA of variance-stabilized expression data identified developmental stage as the dominant source of transcriptomic variation (PC1 = 52.6%), separating larvae from adults (Fig. 4a). Within adults, samples clustered by tissue rather than by sex (Fig. 4b), indicating that tissue identity explains a larger fraction of adult transcriptomic divergence than sexual dimorphism.

Genome-wide differential expression

Larva-adult comparisons detected 8,164 differentially expressed genes (DEGs; padj < 0.05, |log2FC| > 1), of which 6,702 were adult-biased and 1,462 larva-biased (4.6:1 ratio; Table S12; Fig. S2a). Adult-biased counts were consistent when larvae were compared separately to adult males (7,042 DEGs) and adult females (7,492 DEGs), with 4,538 genes differentially expressed across all contrasts (Fig. 5; Fig. S2b–c; Supplementary file 3).

Sex-biased expression was comparatively negligible. Across tissues, 283 genes differed between adult males and females (67 male-biased; 216 female-biased; Table S12; Fig. S3a). Within antennae and thorax, only 110 and 18 genes were sex-biased, respectively (70 male-biased; 40 female-biased in antennae; Fig. S3b; 10 male-biased; 8 female-biased in thorax; Fig. S3c), whereas 2,868 genes differed between antennae and thorax (1,374 antenna-biased; 1,494 thorax-biased; Fig. S3d; Supplementary file 3).

Deployment of chemosensory, detoxification, and digestion gene families

We next analyzed differential expression within the curated chemosensory, detoxification, and digestion gene families identified as enriched in GO analyses. Across these families, 629 genes were differentially expressed between larvae and adults (Table S13; Fig. 6; Supplementary file 4).

Developmental deployment. Chemosensory genes showed strong adult-associated expression: 80 of 81 DEGs were adult-biased, including chemosensory proteins (CSP3_1, CSP3_3), odorant-binding proteins (Obp56d_14, Obp56d_11), and sensory neuron membrane proteins (Snmp1_2, scav-3_1) The sole larva-biased chemosensory gene was GRIN1_1 (glutamate receptor ionotropic NMDA 1; 20.6-fold) (Fig. 6a). Detoxification genes showed a more complex developmental pattern. Although most DEGs were higher in adults (217 of 237), several of the most significantly expressed genes were larva-biased, including proteasome components (PSMC3, PSMC6, PSMC4) and the ATPase TER94. These genes are associated with protein turnover rather than the enzymatic detoxification of plant secondary metabolites. In contrast, genes encoding Phase I–III enzymes were consistently adult-biased, including glutathione S-transferases (GSTD1_1, GSTE3_1), cytochrome P450s (CYP4V2_1, Cyp6a9_36, Cyp6a9_38), carboxylesterases (Est-6_8), and UDP-glucuronosyltransferases (UGT2A1_18) (Fig. 6a). Digestion-related genes comprised the largest set of DEGs (311), with 281 adult-biased (Fig. 6a). The serine protease Tpsg1_1 showed higher expression in larvae, potentially reflecting digestive specialization for dead wood-feeding, whereas the most significant adult-associated genes included cysteine proteases

(Cathepsin B and Cathepsin F), metallopeptidases (CNDP2_1, METAP2), and the carbohydrate-active enzyme MAN2B1_9. Among CAZymes broadly, 150 of 171 DEGs were adult-biased, including additional mannosidases (MAN2B1_12) and glycoside hydrolases, and 98 of 105 differentially expressed serine proteases were adult-biased.

Tissue-specific deployment in adults. Antennae-thorax comparisons revealed tissue-appropriate partitioning across gene families (Table S14; Fig. 6b). Chemosensory genes were overwhelmingly antenna-biased (48 of 52 DEGs), with the strongest enrichment in CSP3_3, odorant-binding proteins (Obp69a_2, Obp56d_38), odorant receptors (Or13a_17, GPROR40_3), and ionotropic receptors (IR25A) (Fig. 6b), consistent with concentration of chemosensory functions in antennal tissues. Detoxification genes, by contrast, were more evenly distributed (43 antenna-biased; 78 thorax-biased). Antenna-biased detoxification genes included ABC transporters (ABCG4, ABCC4_8) and carboxylesterases (CES5A_15), whereas thorax-biased genes included cytochrome P450s (Cyp6a9_41, Cyp6a9_7), carboxylesterases (Est-6_2), and glutathione S-transferases (GstD1_1, GSTO1_4) (Fig. 6b). Digestion genes were similarly balanced between tissues (104 antenna-biased; 132 thorax-biased), with the alpha-amylase Mal-A4_2 antenna-biased and serine proteases (OVCH1_4, prss16_1) and cysteine proteases (#4395 and #4396, BLAST-based identified as Cathepsin_B and Cathepsin_F, respectively; Supplementary file 5) thorax-biased (Fig. 6b).

Sex-biased expression. Sex differences within host-interface gene families were minimal (Table S15). In pooled adults, 9 genes were sex-biased (2 chemosensory; 7 detoxification; Table S15a). Within antennae, 12 genes differed (3 chemosensory; 2 detoxification; 7 digestion; Table S15b), and in thorax, only 4 digestion genes were sex-biased (Table S15c). Only three chemosensory genes showed consistent sex bias: one uncharacterized odorant receptor (#37451, male-biased in pooled adults and antennae) and two odorant-binding proteins with moderate sex differences in antennae (Obp56d_36 male-biased; Obp56d_10 female-biased; Supplementary file 4).

6. Chromosomal organization of differentially expressed genes

Across the 10 chromosomes, we identified 178 tandem clusters encompassing 873 genes (4.0% of all annotated genes; Table S16; Table S17; Fig. 7). Of the 1,293 members in the 16 examined gene families, 873 (67.5%) were arranged in tandem arrays, indicating that tandem duplication has been a pervasive mechanism of copy-number enrichment. Differentially expressed genes were significantly overrepresented within tandem clusters compared with the genome-wide baseline (Fisher's exact test, $OR = 1.86$, $p = 3.43 \times 10^{-19}$). Clustering was especially pronounced in CSPs (94.4% of family members in arrays), UGTs (93.0%), OBPs (81.7%), and P450s (81.5%). The largest arrays included an OBP cluster on chromosome 1 (40 genes, 275.8 kb), serine protease clusters on chromosomes 3 and 6 (30 and 27 genes), a UGT array on chromosome 3 (25 genes, 198.7 kb), and a P450 array on chromosome 6 (25 genes, 224.8 kb). CAZymes formed the most numerous clusters overall (50 arrays, 216 genes), consistent with their expanded representation in the genome.

Chromosome-scale mapping identified seven superloci where tandem clusters from different gene families co-localize within 50 kb (Table S18; Fig. 7). The largest, on chromosome 3 (53.5–53.9 Mb), combined ABC transporter and serine protease clusters totaling 32 genes across 325.7 kb. A second prominent superlocus on chromosome 2 (3.3–3.5 Mb) paired UGT and cysteine protease clusters (14 genes, 152.2 kb), and the most complex superlocus, on chromosome 8 (~ 0.9–1.1 Mb), combined GR, IR, and CAZyme clusters totaling 20 genes across 140.6 kb. Additional superloci paired OR with SerPro (chromosome 1), GR with CysPro (chromosome 5), and SerPro with CAZyme (chromosome 8), indicating that multi-family co-localization spans the chemosensory, detoxification, and digestion axes).

DISCUSSION

Using a newly generated chromosome-scale genome assembly and multi-tissue, multi-stage transcriptomes for the twig-girdler *Oncideres impluviata*—the first genomic resource for the Neotropical tribe Onciderini—we show that host-interface gene families are expanded, physically clustered in tandem arrays and multi-family superloci, and deployed through stage- and tissue-structured expression programs in which host-interface gene expression is concentrated in adults. These patterns are consistent with a life history in which adults locate, assess, and girdle brances that are living and chemically defended, while larvae develop in girdled substrates transitioning toward dead wood. These data establish a comparative framework for Cerambycidae and inspire testable hypotheses about how clustered genome architecture and regulatory partitioning contribute to host breadth, host shifts, and plant manipulation in phytophagous beetles. For example, future work might test whether tandem array size in chemosensory and detoxification families scales with host-range breadth across cerambycids, whether genes within multi-family superloci show coordinated transcriptional responses to host-plant chemistry, and whether the magnitude of adult-versus-larval expression bias varies with reliance on plant-manipulation behaviors such as girdling.

1. A chromosome-scale genome with deeply conserved macrosynteny

The *O. impluviata* assembly provides a high-quality genomic resource for Cerambycidae. Since the first cerambycid genome (*Anoplophora glabripennis* [16]), long-read sequencing has transformed assembly contiguity across the family, enabling reliable annotation of tandem arrays, inference of multi-gene neighborhoods, and chromosome-scale synteny comparisons [10, 18, 19, 83–85]. The near-absence of unresolved bases and long scaffold lengths in *O. impluviata* reduce fragmentation-driven inflation of apparent copy number and improve inference of clustered gene-family architecture, central to the mechanistic questions addressed here.

Genome size in Cerambycidae varies by roughly an order of magnitude, from ~ 262 Mb in *Exocentrus adspersus* [10] to ~ 2,022 Mb in *Rutpela maculata* [86]. At ~ 475 Mb, *O. impluviata* falls toward the lower end of this distribution. Much of this variation is driven by transposable element content [10], and the

moderate repeat fraction in *O. impluviata* is consistent with lineage-specific TE histories shaping cerambycid genome architecture. Because TE proliferation can drive local duplication and regulatory remodeling, this repeat-rich background is relevant when interpreting gene-family expansions, although conclusions about causation would require additional evidence [87].

Assembly contiguity enabled robust gene prediction, yielding a high-confidence gene set comparable to other cerambycid annotations [16]. The repeat landscape is dominated by transposable elements, with a large fraction remaining unclassified, as is common in insect and arthropod genomes [88]. The landscape includes a pronounced low-divergence peak consistent with recent or ongoing TE activity, and elevated coverage at intermediate divergence levels suggesting older expansion episodes [87]. Landscapes restricted to classified families can substantially underestimate total repeat content; improved characterization of unclassified elements will clarify whether observed patterns reflect vertical persistence, horizontal transfer, or lineage-specific innovations [88, 89].

Multiple lines of evidence support chromosome-scale organization for most of the assembly. Terminal telomeric motifs and extensive macrosynteny with *Monochamus alternatus* corroborate the inferred chromosome structure, and scaffold pairs that map to single *M. alternatus* chromosomes are most parsimoniously interpreted as chromosomes split across two scaffolds rather than true fission events. The resulting organization of approximately 10 chromosomes matches the predominant number in cerambycid assemblies and the karyotypic conservatism reported for the family [90–93]. We treat this as an inference from sequence features and comparative mapping rather than a cytogenetic assignment, but it provides a framework for interpreting genome organization in Onciderini. Karyotype data within Onciderini are sparse. *Oncideres amputator* has an XO system (2n = 19 in males; 2n = 20 in females [94]), an unusual condition within Cerambycidae. Our data derive from a single male genome and do not permit sex-chromosome assignment; whether *O. impluviata* shares this condition remains a hypothesis requiring targeted validation.

The strong macrosynteny between *O. impluviata* and *M. alternatus* is notable given the phylogenetic distance between Onciderini and Lamiini, separated since the Mesozoic/early Cenozoic [21]. However, it is consistent with coleopteran evidence that chromosomal gene content is conserved over long timescales while local gene order is extensively reshuffled [95]. Within Cerambycidae, emerging chromosome-level comparisons similarly suggest broad conservation of syntenic blocks across substantial phylogenetic distances, with lineage-specific exceptions associated with rearrangement and fission [10, 19].

2. Gene-family expansion targets herbivory-linked functions

Enrichment in *O. impluviata* is concentrated in functions mediating interactions with host plants: chemical perception, neural signaling, metabolism of xenobiotics (i.e., plant secondary metabolites and other foreign chemicals), and digestion (Fig. 3). The same functional categories recur across multiple phytophagous beetle lineages, including *A. glabripennis*, other Cerambycidae, Chrysomelidae,

Curculionoidea. Within Cerambycidae, chemosensory and neural functions dominate (Fig. 3a, b), a pattern reported across other wood-boring insect groups [16, 96, 97]. Comparisons against other Phytophaga instead highlight digestive and detoxification functions, including proteolysis against Chrysomelidae (Fig. 3c)—potentially reflecting the divergence between cerambycid alkaline-gut and chrysomelid acidic-gut digestive strategies [98]—and oxidative metabolism against Curculionoidea (Fig. 3d).

Chemosensory enrichment provides a plausible genomic basis for the detection of a broad range of host plants [96, 99, 100]. The 60 OBPs in *O. impluviata*, including a 40-gene Obp56d tandem cluster, are comparable to the 52–61 OBPs reported for *A. glabripennis* [16, 101]. Across wood-boring Coleoptera, chemosensory complements of this scale have been associated with broader host use [96]. Cerambycid OBPs bind host volatiles; for example, MaltOBP1 in *M. alternatus* shows high affinity for pine monoterpenes [102].

Our comparative results also indicate enrichment of neural signaling functions, a pattern not previously reported in Cerambycidae but potentially relevant to the behavioral complexity of twig girdling [103, 104]. AMPA glutamate receptor regulation and ion channel activities feature prominently, reflecting the coordinated motor and sensory demands of girdling [20]. Because insect odorant receptors function as ligand-gated ion channels rather than G protein-coupled receptors (GPCRs [105]), the observed GPCR enrichment (Fig. 3a) more parsimoniously reflects neuromodulatory receptors involved in behavioral state regulation than odorant receptor biology per se. With only one girdler genome currently available, additional genomes for girdling and non-girdling Onciderini will be needed to test these neuromodulatory hypotheses.

Detoxification signals are most pronounced (Fig. 3d) in comparisons against Curculionoidea, where enrichment maps oxidative metabolism functions associated with P450-mediated detoxification of plant defense compounds. The 130 P450 genes in *O. impluviata* are comparable to elevated repertoires in other beetles, including the 106 P450s reported for *A. glabripennis* [16, 106, 107]. The prominence of insecticide-related GO terms is consistent with the well-documented involvement of CYP6 and CYP9 lineages in xenobiotic metabolism. These patterns align with *O. impluviata*'s exposure to chemically distinct host defenses, including tannin-rich Fabaceae and chemically diverse Euphorbiaceae and Aquifoliaceae [27, 28, 108]. Broad detoxification enrichment may have facilitated host breadth and recent colonization of introduced *Acacia mearnsii*, although direct testing will require population genomic data and selection analyses on candidate paralogs.

Digestive enrichment, particularly proteolysis, is most evident in comparisons with Chrysomelidae, consistent with differences in feeding substrate and midgut physiology. For instance, serine-type endopeptidase activity is strongly enriched incerambycids, which use serine proteases to digest plant tissues under alkaline conditions, whereas chrysomelids use cysteine proteases under acidic conditions [98, 109].Intriguingly, serine proteases were also the most abundant gene family in our analysis of *O. impluviata*, and these genes were arranged into tandem clusters, much like genes from that same family

in *A. glabripennis* (another cerambycid) [16]. Beyond protein limitation in woody substrates, protease proliferation in *O. impluviata* and other cerambycids may provide a route to tolerance of plant defense inhibitors, such as the Kunitz-type and Bowman-Birk peptides present in Fabaceae hosts [110–112].

3. Transcriptomic deployment of host-interface repertoires

Genomic proliferation in host-interface gene families is common in phytophagous insects, but its relevance depends on whether expanded genes are deployed where they could mediate ecological performance. In *O. impluviata*, chemosensory, detoxification, and digestion families are not constitutively upregulated but organized into expression programs structured by developmental stage and tissue identity, with limited sex effects. This hierarchy, expected in holometabolous insects where larvae and adults occupy distinct niches [113], is amplified in twig-girdlers where adults interact with living, chemically defended tissues while larvae develop in girdled substrates transitioning toward dead wood.

Stage-biased deployment: adults as the primary host-interface stage

Target gene families are overwhelmingly adult-associated. Among differentially expressed genes, the vast majority are adult-biased across olfactory transduction, Phase I–III detoxification, and digestive enzymes (Fig. 6a). Although larval sampling was limited to a single individual, the coherence of this signal across functional categories supports a biological interpretation: adults deploy the principal host-interface repertoire, consistent with the demands of host location, branch assessment, and processing of defended tissues during girdling [114]. Larvae, by contrast, face reduced selective pressure for a broad detoxification response, unlike *A. glabripennis* larvae, which feed within living hosts and upregulate detoxification and digestive machinery [115, 116]. Deployment thus depends on which stage bears the primary burden of chemical confrontation.

The few larva-biased signals align with larval physiology rather than host-defense processing. Proteasome components (PSMC3, PSMC6, PSMC4) and TER94 (VCP/p97 ortholog) likely reflect elevated protein turnover in nutrient-limited wood [117, 118], while the larva-biased serine protease Tpsg1_1 is consistent with the alkaline digestive strategy of cerambycids and constraints on nutrient extraction from wood [98, 109]. Adults instead show coordinated upregulation of CYP6 P450s with conjugation and transport components, consistent with a complete detoxification pipeline deployed against tannins, phenolics, and other defenses in living tissues [116, 119].

Tissue, not sex, partitions expression

Sex contributes little to host-interface deployment despite female-specific girdling. Only a small number of target-family genes are sex-biased, consistent with coleopteran cases where sexes share peripheral chemosensory repertoires despite behavioral dimorphism [120–123]. A parallel logic likely applies to detoxification: although girdling itself is performed exclusively by females, both sexes of most Lamiinae undergo maturation feeding on living plant tissue—chewing bark, phloem, leaves, and twigs for one to three weeks after emergence—and therefore share exposure to the same plant defense compounds [15,

115]. Dimorphism likely arises through sex-specific central processing rather than divergent peripheral expression, consistent with circuit-level models where shared receptors feed into dimorphic neural pathways [124–126]. The extreme male-biased expression of one uncharacterized odorant receptor (#37451) is a conspicuous exception and a strong candidate for sex-specific communication, given that beetle pheromone receptors arise independently across OR phylogeny [127, 128] and cerambycid systems include both aggregation and female-produced pheromones [129, 130]. Ligand identity and behavioral role will require heterologous expression and physiological validation.

Within adults, tissue identity structures deployment and suggests regulatory partitioning among paralogs (Fig. 6b). Chemosensory genes concentrate in antennae, consistent with localization of volatile detection [131, 132]. Detoxification genes partition between tissues: antenna-enriched esterases and transporters are consistent with odorant-degrading enzymes for signal termination [133–135], while thorax-enriched P450s and ABC transporters suggest systemic detoxification [136–139]. Digestive families show more heterogeneous distribution, consistent with both alimentary and non-alimentary roles [140–143]. Together, these patterns indicate that host-interface families are deployed through structured regulatory programs, motivating chromosome-scale analyses of tandem arrays and multi-family neighborhoods as genomic substrates for coordinated expression and paralog partitioning.

4. Host-interface genes show clustered genomic organization

Genome architecture links the enrichment and expression patterns described above to physical organization. In *O. impluviata*, most host-interface family members reside in tandem arrays, and DEGs are significantly overrepresented within those arrays. Tandem clustering may reflect functional coordination—co-regulation through shared chromatin environments or enhancers—and/or the mechanistic legacy of tandem duplication, the primary mode of gene-family growth [144]. The combination of significant DEG overrepresentation within arrays and context-appropriate expression shifts among linked paralogs is compatible with regulatory coordination.

Within the largest arrays, substantial fractions of paralogs are differentially expressed in functionally relevant tissue contrasts. The chromosome 1 OBP super-array exemplifies this, linking three levels of evidence: a very large physical array (40 OBPs), a clear host-interface role (odorant binding and perireceptor processing), and a concentrated signal of tissue-structured deployment (multiple paralogs differentially expressed between antennae and thorax, including Obp56d-like genes). P450 and CSP arrays show parallel patterns, with Cyp6a9 paralogs displaying adult-associated, tissue-structured shifts consistent with detoxification and odorant clearance in tissues that directly confront host chemistry [119, 139].

Superloci extend this logic to multi-family neighborhoods. The co-localization of ABC transporters with serine proteases (chromosome 3), gustatory and ionotropic receptors with CAZymes (chromosome 8), and detoxification with digestive families across multiple chromosomes suggests that host-interface functions are organized as multi-family neighborhoods potentially sharing regulatory domains,

analogous in principle to eukaryotic metabolic gene clusters where pathway-relevant genes co-localize under shared chromatin-level control [144, 145].

The tandem arrays and multi-family neighborhoods point to a unifying mechanistic interpretation. This architecture provides a plausible substrate for host breadth and host shifts by concentrating standing variation where dosage and regulatory change can act. A rapid shift such as colonization of *Acacia* over roughly the last century, for instance, would not require *de novo* pathway invention but could draw on copy-number variation, paralog divergence, and expression changes within clustered repertoires [146]. Testing whether recent host use has shaped specific clusters will require population genomic sampling and selection analyses in candidate regions.

CONCLUSIONS

We link a new chromosome-scale genome of *O. impluviata* to the ecological demands of twig girdling, integrating comparative genomics, expression profiling, and chromosomal mapping of host-interface gene families. Enriched repertoires span chemosensory, neural signaling, detoxification, and digestive functions, and are deployed through structured programs in adults, the stage that confronts living, chemically defended tissues, with strong tissue partitioning and limited sex bias despite behavioral dimorphism. Chromosome-scale organization adds a mechanistic layer: host-interface families concentrate in tandem arrays where differentially expressed genes are overrepresented, and multi-family superloci place complementary functions in shared genomic neighborhoods. Together, these results demonstrate that host-interface repertoires are expanded, transcriptionally partitioned by stage and tissue, and physically clustered in architectures that concentrate standing variation where dosage and regulatory change can act. This architecture gives rise to testable hypotheses about how tandem duplication, regulatory partitioning, and neighborhood effects contribute to host breadth, host shifts, and plant manipulation. Moreover, the resources generated here provide a substrate for comparative work on Cerambycidae, as well as a framework for prioritizing candidate loci for functional assays and population genomic tests relevant to twig-girdling pests.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary files. Raw PacBio HiFi reads, the genome assembly, GFF3 annotation and RNA-seq reads

have been deposited in NCBI under BioProject PRJNA1256903. Accession numbers: PacBio HiFi genome reads (SRR38808688); genome assembly and GFF3 annotation (JBRAVC000000000); and RNA-seq reads (SRR38815781, SRR38815780, SRR38815779, SRR38815778, SRR38815777, SRR38815776, SRR38815775, SRR38815774, SRR38815773). Custom scripts and analysis pipelines are available at https://github.com/diego-ds-souza/oimp_genome.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (process 201.917/2020, DSS), a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant (RGPIN-2018-04920), and a Grainger Bioinformatics Center Research Grant at the Field Museum of Natural History.

Authors' contributions

DSS, FAHS, and RLKF conceived the study and designed the research framework. DSS led the study, performed the genome assembly, annotation, and comparative genomic and transcriptomic analyses, produced the data visualizations, interpreted the results, and wrote the original draft. RLKF contributed to investigation and interpretation of results BASM and TAS contributed to methodological development, bioinformatic analyses, investigation and interpretation of results. TS contributed to synteny and repeat analyses and interpretation of results. DNM and JOSJ contributed to specimen acquisition, field sampling, taxonomic context, and interpretation of results. LM provided resources and contributed to interpretation of results. EHN provided taxonomic context and contributed to interpretation of results. MLM provided taxonomic context and resources, supervised the work, acquired funding, and contributed to interpretation of results. FAHS supervised the work, acquired funding, and contributed to interpretation of results. All authors contributed to manuscript revision and read and approved the final manuscript.

Acknowledgements

We thank Angela Echeverry for assistance during fieldwork and laboratory experiments; Henrique Pinton Greff and Jeferson de Oliveira for facilitating access to field sites; Shawn Abraham, Oksana Vernygora, Erin Campbell and Janet Sperling for discussions and assistance with bioinformatics during early stages of this project; Taylor Heins and Yukun Sun for analytical support and insightful discussions; and Osvaldo Marinotti for valuable suggestions on the final draft of the manuscript. Whole-genome sequencing was performed at the Genomics and Cell Characterization Core Facility, University of Oregon, Eugene, USA. We especially thank Maggie Weitzman, Tina Arredondo, and the facility team for their assistance with sample processing and whole-genome sequencing. RNA sequencing was performed by Novogene Corporation Inc., Sacramento, USA. Analyses were performed using computational resources

provided by the Digital Research Alliance of Canada and the Grainger Bioinformatics Center at the Field Museum of Natural History.

References

1. Ehrlich PR, Raven PH. Butterflies and plants: a study in coevolution. Evolution. 1964;18(4):586-608. https://doi.org/10.2307/2406212
2. Simon J-C, d'Alencon E, Guy E, Jacquin-Joly E, Jaquiéry J, Nouhaud P, et al. Genomics of adaptation to host-plants in herbivorous insects. Brief Funct Genomics. 2015;14(6):413-423. https://doi.org/10.1093/bfgp/elv015
3. Vertacnik KL, Linnen CR. Evolutionary genetics of host shifts in herbivorous insects: insights from the age of genomics. Ann N Y Acad Sci. 2017;1389(1):186-212. https://doi.org/10.1111/nyas.13311
4. Birnbaum SSL, Abbot P. Gene expression and diet breadth in plant-feeding insects: summarizing trends. Trends Ecol Evol. 2020;35(3):259-277. https://doi.org/10.1016/j.tree.2019.10.014
5. Hardy NB, Kaczvinsky C, Bird G, Normark BB. What we don't know about diet-breadth evolution in herbivorous insects. Annu Rev Ecol Evol Syst. 2020;51:103-122. https://doi.org/10.1146/annurev-ecolsys-011720-023322
6. Allio R, Nabholz B, Wanke S, Chomicki G, Pérez-Escobar OA, Cotton AM, et al. Genome-wide macroevolutionary signatures of key innovations in butterflies colonizing new host plants. Nat Commun. 2021;12:354. https://doi.org/10.1038/s41467-020-20507-3
7. Engsontia P, Sangket U, Chotigeat W, Satasook C. Molecular evolution of the odorant and gustatory receptor genes in lepidopteran insects: implications for their adaptation and speciation. J Mol Evol. 2014;79:21-39. https://doi.org/10.1007/s00239-014-9633-0
8. Erb M, Reymond P. Molecular interactions between plants and insect herbivores. Annu Rev Plant Biol. 2019;70:527-557. https://doi.org/10.1146/annurev-arplant-050718-095910
9. Pelaez JN, Gloss AD, Goldman-Huertas B, Kim B, Lapoint RT, Pimentel-Solorio G, et al. Evolution of chemosensory and detoxification gene families across herbivorous Drosophilidae. G3 (Bethesda). 2023;13:jkad133. https://doi.org/10.1093/g3journal/jkad133
10. Shin NR, Okamura Y, Kirsch R, Pauchet Y. Genome sequencing provides insights into the evolution of gene families encoding plant cell wall-degrading enzymes in longhorned beetles. Insect Mol Biol. 2023;32(5):469-483. https://doi.org/10.1111/imb.12844
11. Despres L, David J-P, Gallet C. The evolutionary ecology of insect resistance to plant chemicals. Trends Ecol Evol. 2007;22(6):298-307. https://doi.org/10.1016/j.tree.2007.02.010
12. Sparks ME, Bansal R, Benoit JB, Blackburn MB, Chao H, Chen M, et al. Brown marmorated stink bug, Halyomorpha halys (Stål), genome: putative underpinnings of polyphagy, insecticide resistance potential and biology of a top worldwide pest. BMC Genomics. 2020;21(1):227. https://doi.org/10.1186/s12864-020-6510-7

13. Weng Y-M, Shashank PR, Godfrey RK, Plotkin D, Parker BM, Wist T, Kawahara AY. Evolutionary genomics of three agricultural pest moths reveals rapid evolution of host adaptation and immune-related genes. Gigascience. 2024;13:giad103. https://doi.org/10.1093/gigascience/giad103
14. Chevillotte H, Tavakilian G. Titan: database about longhorns or timber-beetles (Cerambycidae). http://titan.gbif.fr/. Accessed 16 Feb 2025.
15. Haack RA. Feeding biology of cerambycids. In: Wang Q, editor. Cerambycidae of the World: Biology and Pest Management. CRC Press; 2017. p. 105-132.
16. McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, et al. Genome of the Asian longhorned beetle (*Anoplophora glabripennis*), a globally significant invasive species, reveals key functional and evolutionary innovations at the beetle-plant interface. Genome Biol. 2016;17:227. https://doi.org/10.1186/s13059-016-1088-8
17. Fu N, Li J, Ren L, Li X, Wang M, Li F, et al. Chromosome-level genome assembly of *Monochamus saltuarius* reveals its adaptation and interaction mechanism with pine wood nematode. Int J Biol Macromol. 2022;222(A):325-336. https://doi.org/10.1016/j.ijbiomac.2022.09.108
18. Gao YF, Zhao JK, Li JQ, Ma X, Zong SX, Wei SJ. Chromosome-level genome assembly of the Japanese sawyer beetle *Monochamus alternatus*. Sci Data. 2024;11:199. https://doi.org/10.1038/s41597-024-03048-y
19. Sylvester T, Adams R, Mitchell RF, Ray AM, Shen R, Shin NR, et al. Insights into longhorn beetle (Cerambycidae) evolution from comparative analyses of the red-headed ash borer (*Neoclytus acuminatus acuminatus*) genome. J Hered. 2025;116(4):558-567. https://doi.org/10.1093/jhered/esaf016
20. Forcella F. Why twig-girdling beetles girdle twigs. Naturwissenschaften. 1982;69:398-400. https://doi.org/10.1007/BF00396699
21. Souza DS, French RLK, Júnior JOS, Nearns EH, Marinoni L, Swift IP, et al. Cutting the sap: First molecular phylogeny of twig-girdler longhorn beetles (Coleoptera: Cerambycidae: Lamiinae: Onciderini) suggests shifts in host plant attack behaviours contributed to morphological evolution. Syst Entomol. 2025;50(2):369-386. https://doi.org/10.1111/syen.12660
22. Rice ME. Branch girdling and oviposition biology of *Oncideres pustulatus* (Coleoptera: Cerambycidae) on *Acacia farnesiana*. Ann Entomol Soc Am. 1989;82(2):181-186. https://doi.org/10.1093/aesa/82.2.181
23. Germar EF. Insectorum species: novae aut minus cognitae, descriptionibus illustratae. Impensis J.C. Hendelii et Filii; 1823. https://doi.org/10.5962/bhl.title.130964
24. Machado DN, Costa E, Perini CR, Valmorbida I, Souza DS, Pedron L, et al. Genetic diversity, phylogenetic and phylogeographic analyses of *Oncideres impluviata* (Germar, 1823) (Coleoptera: Cerambycidae) in Rio Grande do Sul state, Brazil. Rev Bras Entomol. 2023;67(1):e20220092. https://doi.org/10.1590/1806-9665-RBENT-2022-0092
25. Schneider PR, Oesten G, Brill A, Mainardi GL. Determinação da produção de casca em acácia-negra, *Acacia mearnsii* De Wild. Cienc Florest. 1991;1:64-75. https://doi.org/10.5902/19805098257

26. Attias N, Siqueira MF, Bergallo HG. Acacias Australianas no Brasil: Histórico, Formas de Uso Potencial de Invasão. Biodivers Bras. 2013;3(2):74-96.
27. Ogawa S, Yazaki Y. Tannins from *Acacia mearnsii* De Wild. bark: tannin determination and biological activities. Molecules. 2018;23(4):837. https://doi.org/10.3390/molecules23040837
28. Monné ML. Catalogue of the Cerambycidae (Coleoptera) of the Neotropical Region. Part III. Subfamily Lamiinae. http://cerambycids.com/. Accessed 22 Feb 2026.
29. Paulino Neto HF. Lenhadores da natureza. Cienc Hoje. 2004;35:67-69.
30. Paulino Neto HF, Vasconcelos-Neto J, Carmello-Gerreiro SM. The biology of *Oncideres humeralis* Thorms (Coleoptera: Cerambycidae: Lamiinae) and new Cerambycidae–Melastomataceae host-plant associations. Stud Neotrop Fauna Environ. 2006;41(3):227-233. https://doi.org/10.1080/01650520600839680
31. Rice ME. Branch girdling by *Oncideres cingulata* (Coleoptera: Cerambycidae) and relative host quality of persimmon, hickory, and elm. Ann Entomol Soc Am. 1995;88:451-455. https://doi.org/10.1093/aesa/88.4.451
32. Shin NR, Shin S, Okamura Y, Kirsch R, Lombard V, Svacha P, et al. Larvae of longhorned beetles (Coleoptera: Cerambycidae) have evolved a diverse and phylogenetically conserved array of plant cell wall degrading enzymes. Syst Entomol. 2021;46(4):784-797. https://doi.org/10.1111/syen.12488
33. Schwander T, Libbrecht R, Keller L. Supergenes and complex phenotypes. Curr Biol. 2014;24:R288-R294. http://dx.doi.org/10.1016/j.cub.2014.01.056
34. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6):764-770. https://doi.org/10.1093/bioinformatics/btr011
35. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017;33(14):2202-2204. https://doi.org/10.1093/bioinformatics/btx153
36. Cheng H, Jarvis ED, Fedrigo O, Koeple K-P, Urban L, Gemmell NJ, et al. Haplotype-resolved assembly of diploid genomes without parental data. Nat Biotechnol. 2022;40:1332-1335. https://doi.org/10.1038/s41587-022-01261-x
37. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020;36(9):2896-2898. https://doi.org/10.1093/bioinformatics/btaa025
38. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013;29(8):1072-1075. https://doi.org/10.1093/bioinformatics/btt086
39. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210-3212. https://doi.org/10.1093/bioinformatics/btv351
40. Brown MR, de La Rosa PMG, Blaxter M. Tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. Bioinformatics. 2025;41(2):btaf049.

https://doi.org/10.1093/bioinformatics/btaf049

41. Blackmon H, Demuth JP. Coleoptera karyotype database. Coleopt Bull. 2015;69(1):174-175. https://doi.org/10.1649/0010-065X-69.1.174
42. He J, Bai T, Wan W, Dong Z, Wang Y, Zhang H, Li X. Chromosome-level genome assembly of stem borer *Batocera rufomaculata* using PacBio HiFi and Hi-C sequencing. Sci Data. 2025;12:1099. https://doi.org/10.1038/s41597-025-05463-1
43. Souza DS, Marinoni L, Monné ML, Gómez-Zurita J. Molecular phylogenetic assessment of the tribal classification of Lamiinae (Coleoptera: Cerambycidae). Mol Phylogenet Evol. 2020;145:106736. https://doi.org/10.1016/j.ympev.2020.106736
44. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009;19(9):1639-1645. https://doi.org/10.1101/gr.092759.109
45. Gu Z, Gu L, Eils R, Schlesner M, Brors B. circlize Implements and enhances circular visualization in R. Bioinformatics. 2014;30(19):2811-2812. https://doi.org/10.1093/bioinformatics/btu393
46. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34(17):i884-i890. https://doi.org/10.1093/bioinformatics/bty560
47. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat Biotechnol. 2011;29:644-652. https://doi.org/10.1038/nbt.1883
48. Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, et al. The Dfam database of repetitive DNA families. Nucleic Acids Res. 2016;44(D1):D81-D89. https://doi.org/10.1093/nar/gkv1272
49. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6:11. https://doi.org/10.1186/s13100-015-0041-9
50. Palmer JM, Stajich J. Funannotate v1.8.1: eukaryotic genome annotation. Zenodo; 2020. https://doi.org/10.5281/zenodo.1134477
51. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. 2006;34(Suppl 2):W435-W439. https://doi.org/10.1093/nar/gkl200
52. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 2005;33(20):6494-6506. https://doi.org/10.1093/nar/gki937
53. Li S, Ma L, Li H, Vang S, Hu Y, Bolund L, Wang J. SNAP: an integrated SNP annotation platform. Nucleic Acids Res. 2007;35:D707-D710. https://doi.org/10.1093/nar/gkl969
54. Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. Bioinformatics. 2004;20(16):2878-2879. https://doi.org/10.1093/bioinformatics/bth315

55. Testa AC, Hane JK, Ellwood SR, Oliver RP. CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. BMC Genomics. 2015;16:170. https://doi.org/10.1186/s12864-015-1344-4
56. Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. BMC Bioinformatics. 2005;6:31. https://doi.org/10.1186/1471-2105-6-31
57. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bairoch A. UniProtKB/Swiss-Prot. In: Edwards D, editor. Plant Bioinformatics. Methods in Molecular Biology, vol. 406. Humana Press; 2007. p. 89-112. https://doi.org/10.1007/978-1-59745-535-0_4
58. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019;37:907-915. https://doi.org/10.1038/s41587-019-0201-4
59. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290-295. https://doi.org/10.1038/nbt.3122
60. Wu TD, Watanabe CK. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. Bioinformatics. 2005;21(9):1859-1875. https://doi.org/10.1093/bioinformatics/bti310
61. Kent WJ. BLAT - the BLAST-like alignment tool. Genome Res. 2002;12:656-664. https://doi.org/10.1101/gr.229202
62. Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith Jr RK, Hannick LI, et al. Improving the *Arabidopsis* genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003;31(19):5654-5666. https://doi.org/10.1093/nar/gkg770
63. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. 2008;9:R7. https://doi.org/10.1186/gb-2008-9-1-r7
64. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, et al. Pfam: the protein families database in 2021. Nucleic Acids Res. 2021;49(D1):D412-D419. https://doi.org/10.1093/nar/gkaa913
65. Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, et al. InterPro: the integrative protein signature database. Nucleic Acids Res. 2009;37(1):D211-D215. https://doi.org/10.1093/nar/gkn785
66. Jensen LJ, Julien P, Kuhn M, von Mering C, Muller J, Doerks T, Bork P. eggNOG: automated construction and annotation of orthologous groups of genes. Nucleic Acids Res. 2008;36(1):D250-D254. https://doi.org/10.1093/nar/gkm796
67. Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 2000;28(1):33-36. https://doi.org/10.1093/nar/28.1.33
68. The Gene Ontology Consortium. The Gene Ontology knowledgebase in 2023. Genetics. 2023;224(1):iyad031. https://doi.org/10.1093/genetics/iyad031

69. Rawlings ND, Barrett AJ, Thomas PD, Huang X, Bateman A, Finn RD. The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. Nucleic Acids Res. 2018;46(D1):D624-D632. https://doi.org/10.1093/nar/gkx1134
70. Drula E, Garron M-L, Dogan S, Lombard V, Henrissat B, Terrapon N. The carbohydrate-active enzyme database: functions and literature. Nucleic Acids Res. 2022;50(D1):D571-D577. https://doi.org/10.1093/nar/gkab1045
71. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. 2023;51(D1):D523-D531. https://doi.org/10.1093/nar/gkac1052
72. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:238. https://doi.org/10.1186/s13059-019-1832-y
73. Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 2015;12:59-60. https://doi.org/10.1038/nmeth.3176
74. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-780. https://doi.org/10.1093/molbev/mst010
75. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972-1973. https://doi.org/10.1093/bioinformatics/btp348
76. Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ. 2016;4:e1660. https://doi.org/10.7717/peerj.1660
77. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530-1534. https://doi.org/10.1093/molbev/msaa015
78. Klopfenstein DV, Zhang L, Pedersen BS, Ramírez F, Vesztrocy AW, Naldi A, et al. GOATOOLS: a Python library for Gene Ontology analyses. Sci Rep. 2018;8:10872. https://doi.org/10.1038/s41598-018-28948-z
79. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8
80. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017;14:417-419. https://doi.org/10.1038/nmeth.4197
81. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. 2015;4:1521. https://doi.org/10.12688/f1000research.7563.1
82. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Methodol. 1995;57(1):289-300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

83. Mitchell R, Poloni R, Natural History Museum Genome Acquisition Lab, et al. The genome sequence of the four-banded longhorn beetle, *Leptura quadrifasciata* Linnaeus, 1758. Wellcome Open Res. 2024;9:386. https://doi.org/10.12688/wellcomeopenres.22611.1
84. Gao YF, Yang F, Chen JC, Song W, Wang JX, Zong SX, Wei SJ. Chromosome-level genome assembly of the longhorn beetle *Arhopalus rusticus* (Coleoptera: Cerambycidae). Sci Data. 2025;12:1171. https://doi.org/10.1038/s41597-025-05523-6
85. Sylvester T, Adams R, Mitchell RF, Ray AM, Shen R, Shin NR, et al. Comparative analyses of the banded alder borer (*Rosalia funebris*) and Asian longhorned beetle (*Anoplophora glabripennis*) genomes reveal significant differences in genome architecture and gene content among these and other Cerambycidae. J Hered. 2024;115(5):516-523. https://doi.org/10.1093/jhered/esae021
86. Sivell O, Sivell D, Barclay MVL, Crowley LM, Natural History Museum Genome Acquisition Lab, et al. The genome sequence of a longhorn beetle, *Rutpela maculata* (Poda, 1769). Wellcome Open Res. 2023;8:579. https://doi.org/10.12688/wellcomeopenres.20500.1
87. Gilbert C, Peccoud J, Cordaux R. Transposable elements and the evolution of insects. Annu Rev Entomol. 2021;66:355-372. https://doi.org/10.1146/annurev-ento-070720-074650
88. Petersen M, Armisen D, Gibbs RA, Hering L, Khila A, Mayer G, et al. Diversity and evolution of the transposable element repertoire in arthropods with particular reference to insects. BMC Evol Biol. 2019;19:11. https://doi.org/10.1186/s12862-018-1324-9
89. Peccoud J, Loiseau V, Cordaux R, Gilbert C. Massive horizontal transfer of transposable elements in insects. Proc Natl Acad Sci U S A. 2017;114(18):4721-4726. https://doi.org/10.1073/pnas.1621178114
90. Rozek M, Lachowska D, Petitpierre E, Holecova M. C-bands on chromosomes of 32 beetle species (Coleoptera: Elateridae, Cantharidae, Oedemeridae, Cerambycidae, Anthicidae, Chrysomelidae, Attelabidae and Curculionidae). Hereditas. 2004;140(3):161-170. https://doi.org/10.1111/j.1601-5223.2004.01810.x
91. Cesari M, Marescalchi O, Francardi V, Mantovani B. Taxonomy and phylogeny of European *Monochamus* species: first molecular and karyological data. J Zool Syst Evol Res. 2005;43:1-7. https://doi.org/10.1111/j.1439-0469.2004.00279.x
92. Hanrahan SJ, Johnston JS. New genome size estimates of 134 species of arthropods. Chromosome Res. 2011;19:809-823. https://doi.org/10.1007/s10577-011-9231-6
93. Dutrillaux AM, Dutrillaux B. Chromosome evolution of some Palaearctic and Nearctic species of *Monochamus* (Coleoptera: Cerambycidae: Lamiinae). Ann Soc Entomol Fr (NS). 2014;50(2):213-218. https://doi.org/10.1080/00379271.2014.937599
94. Dutrillaux AM, Dutrillaux B. Sex chromosome rearrangements in Polyphaga beetles. Sex Dev. 2009;3(1):43-54. https://doi.org/10.1159/000200081
95. Van Dam MH, Cabras AA, Henderson JB, Rominger AJ, Pérez Estrada C, Omer AD, et al. The Easter Egg Weevil (*Pachyrhynchus*) genome reveals syntenic patterns in Coleoptera across 200 million

years of evolution. PLoS Genet. 2021;17(8):e1009745. https://doi.org/10.1371/journal.pgen.1009745

96. Andersson MN, Keeling CI, Mitchell RF. Genomic content of chemosensory genes correlates with host range in wood-boring beetles (*Dendroctonus ponderosae*, *Agrilus planipennis*, and *Anoplophora glabripennis*). BMC Genomics. 2019;20:690. https://doi.org/10.1186/s12864-019-6054-x
97. Mitchell RF, Andersson MN. Olfactory genomics of the Coleoptera. In: Blomquist GJ, Vogt RG, editors. Insect Pheromone Biochemistry and Molecular Biology. 2nd ed. Elsevier; 2021. p. 547-590. https://doi.org/10.1016/B978-0-12-819628-1.00017-1
98. Johnson KS, Rabosky D. Phylogenetic distribution of cysteine proteinases in beetles: evidence for an evolutionary shift to an alkaline digestive strategy in Cerambycidae. Comp Biochem Physiol B Biochem Mol Biol. 2000;126(4):609-619. https://doi.org/10.1016/S0305-0491(00)00232-7
99. Nei M, Niimura Y, Nozawa M. The evolution of animal chemosensory receptor gene repertoires: roles of chance and necessity. Nat Rev Genet. 2008;9:951-963. https://doi.org/10.1038/nrg2480
100. Sanchez-Gracia A, Vieira FG, Rozas J. Molecular evolution of the major chemosensory gene families in insects. Heredity (Edinb). 2009;103:208-216. https://doi.org/10.1038/hdy.2009.55
101. Hu P, Wang J, Cui M, Tao J, Luo Y. Antennal transcriptome analysis of the Asian longhorned beetle *Anoplophora glabripennis*. Sci Rep. 2016;6:26652. https://doi.org/10.1038/srep26652
102. Li DZ, Huang XF, Yang RN, Chen JY, Wang MQ. Functional Analysis of Two Odorant-Binding Proteins, MaltOBP9 and MaltOBP10, in *Monochamus alternatus* Hope. Front Physiol. 2020;11:317. https://doi.org/10.3389/fphys.2020.00317
103. Farooqui T. Octopamine-mediated neuromodulation of insect senses. Neurochem Res. 2007;32:1511-1529. https://doi.org/10.1007/s11064-007-9344-7
104. Menzel R. The honeybee as a model for understanding the basis of cognition. Nat Rev Neurosci. 2012;13:758-768. https://doi.org/10.1038/nrn3357
105. Butterwick JA, del Marmol J, Kim KH, Kahlson MA, Rogow JA, Walz T, et al. Cryo-EM structure of the insect olfactory receptor Orco. Nature. 2018;560:447-452. https://doi.org/10.1038/s41586-018-0420-8
106. Tribolium Genome Sequencing Consortium. The genome of the model beetle and pest *Tribolium castaneum*. Nature. 2008;452:949-955. https://doi.org/10.1038/nature06784
107. Keeling CI, Yuen MM, Liao NY, Docking TR, Chan SK, Taylor GA, et al. Draft genome of the mountain pine beetle, *Dendroctonus ponderosae* Hopkins, a major forest pest. Genome Biol. 2013;14:R27. https://doi.org/10.1186/gb-2013-14-3-r27
108. Barbehenn RV, Constabel CP. Tannins in plant-herbivore interactions. Phytochemistry. 2011;72(13):1551-1565. https://doi.org/10.1016/j.phytochem.2011.01.040
109. Scully ED, Geib SM, Hoover K, Tien M, Tringe SG, Barry KW, et al. Metagenomic profiling reveals lignocellulose degrading system in a microbial community associated with a wood-feeding beetle. PLoS One. 2013;8:e73827. https://doi.org/10.1371/journal.pone.0073827

110. Zhu-Salzman K, Zeng R. Insect Response to Plant Defensive Protease Inhibitors. Annu Rev Entomol. 2015;60:233-252. https://doi.org/10.1146/annurev-ento-010814-020816

111. Babu SR, Subrahmanyam B. Bio-potency of serine proteinase inhibitors from *Acacia senegal* seeds on digestive proteinases, larval growth and development of Helicoverpa armigera (Hubner). Pestic Biochem Physiol. 2010;98(3):349-358. https://doi.org/10.1016/j.pestbp.2010.07.008

112. Gruden K, Kuipers AGJ, Guncar G, Slapar N, Strukelj B, Jongsma MA. Molecular basis of Colorado potato beetle adaptation to potato plant defence at the level of digestive cysteine proteinases. Insect Biochem Mol Biol. 2004;34(4):365-375. https://doi.org/10.1016/j.ibmb.2004.01.003

113. Truman JW, Riddiford LM. The origins of insect metamorphosis. Nature. 1999;401:447-452. https://doi.org/10.1038/46737

114. Millar JG, Hanks LM. Chemical ecology of cerambycids. In: Wang Q, editor. Cerambycidae of the World: Biology and Pest Management. CRC Press; 2017. p. 161-208.

115. Mason C, Scully E, Geib S, Hoover K. Contrasting diets reveal metabolic plasticity in the tree-killing beetle, *Anoplophora glabripennis* (Cerambycidae: Lamiinae). Sci Rep. 2016;6:33813. https://doi.org/10.1038/srep33813

116. Dai L, Ma M, Wang C, Shi Q, Zhang R, Chen H. Cytochrome P450s from the Chinese white pine beetle, *Dendroctonus armandi*: expression profiles of different stages and responses to host allelochemicals. Insect Biochem Mol Biol. 2015;65:35-46. https://doi.org/10.1016/j.ibmb.2015.08.004

117. Dawson SP, Arnold JE, Mayer NJ, Reynolds SE, Billett MA, Gordon C, et al. Developmental changes of the 26 S proteasome in abdominal intersegmental muscles of *Manduca sexta* during programmed cell death. J Biol Chem. 1995;270(4):1850-1858. https://doi.org/10.1074/jbc.270.4.1850

118. Pintér M, Jékely G, Szepesi RJ, Farkas A, Theopold U, Meyer HE, et al. TER94, a *Drosophila* homolog of the membrane fusion protein CDC48/p97, is accumulated in nonproliferating cells: in the reproductive organs and in the brain of the imago. Insect Biochem Mol Biol. 1998;28(2):91-98. https://doi.org/10.1016/S0965-1748(97)00095-7

119. Feyereisen R. Insect CYP genes and P450 enzymes. In: Gilbert LI, editor. Insect Molecular Biology and Biochemistry. Academic Press; 2012. p. 236-316. https://doi.org/10.1016/B978-0-12-384747-8.10008-X

120. Dippel S, Oberhofer G, Kahnt J, Gerischer L, Opitz L, Schachtner J, et al. Tissue-specific transcriptomics, chromosomal localization, and phylogeny of chemosensory and odorant binding proteins from the red flour beetle *Tribolium castaneum* reveal subgroup specificities for olfaction or more general functions. BMC Genomics. 2014;15:1141. https://doi.org/10.1186/1471-2164-15-1141

121. Dippel S, Kollmann M, Oberhofer G, Montino A, Knoll C, Krala M, et al. Morphological and Transcriptomic Analysis of a Beetle Chemosensory System Reveals a Gnathal Olfactory Center. BMC Biol. 2016;14(1):90. https://doi.org/10.1186/s12915-016-0304-z

122. Rondoni G, Roman A, Meslin C, Montagné N, Conti E, Jacquin-Joly E. Antennal Transcriptome Analysis and Identification of Candidate Chemosensory Genes of the Harlequin Ladybird Beetle,

Harmonia axyridis (Pallas) (Coleoptera: Coccinellidae). Insects. 2021;12(3):209. https://doi.org/10.3390/insects12030209

123. Johny J, Große-Wilde E, Kalinová B, Roy A. Antennal Transcriptome Screening and Identification of Chemosensory Proteins in the Double-Spine European Spruce Bark Beetle, *Ips duplicatus* (Coleoptera: Scolytinae). Int J Mol Sci. 2024;25(17):9513. https://doi.org/10.3390/ijms25179513
124. Datta SR, Vasconcelos ML, Ruta V, Luo S, Wong A, Demir E, et al. The *Drosophila* pheromone cVA activates a sexually dimorphic neural circuit. Nature. 2008;452:473-477. https://doi.org/10.1038/nature06808
125. Ruta V, Datta SR, Vasconcelos ML, Freeland J, Looger LL, Bhatt AS. A dimorphic pheromone circuit in *Drosophila* from sensory input to descending output. Nature. 2010;468:686-690. https://doi.org/10.1038/nature09554
126. Kohl J, Ostrovsky AD, Frechter S, Jefferis GSXE. A bidirectional circuit switch reroutes pheromone signals in male and female brains. Cell. 2013;155(7):1610-1623. https://doi.org/10.1016/j.cell.2013.11.025
127. Mitchell RF, Hughes DT, Luetje CW, Millar JG, Soriano-Agaton F, Hanks LM, et al. Sequencing and characterizing odorant receptors of the cerambycid beetle *Megacyllene caryae*. Insect Biochem Mol Biol. 2012;141(7):499-505. https://doi.org/10.1016/j.ibmb.2012.03.007
128. Mitchell RF, Schneider TM, Schwartz AM, Andersson MN, McKenna DD. The diversity and evolution of odorant receptors in beetles (Coleoptera). Insect Mol Biol. 2020;29(1):77-91. https://doi.org/10.1111/imb.12611
129. Hanks LM, Millar JG. Sex and aggregation-sex pheromones of cerambycid beetles: basic science and practical applications. J Chem Ecol. 2016;42:631-654. https://doi.org/10.1007/s10886-016-0733-8
130. Fierke MK, Skabeikis DD, Millar JG, Teale SA, McElfresh JS, Hanks LM. Identification of a male-produced aggregation pheromone for *Monochamus scutellatus scutellatus* and an attractant for the congener *Monochamus notatus* (Coleoptera: Cerambycidae). J Econ Entomol. 2012;105(6):2029-2034. https://doi.org/10.1603/EC12101
131. Pelosi P, Iovinella I, Zhu J, Wang G, Dani FR. Beyond chemoreception: diverse tasks of soluble olfactory proteins in insects. Biol Rev. 2018;93(1):184-200. https://doi.org/10.1111/brv.12339
132. Li Z, Dai L, Chu H, Fu D, Sun Y, Chen H. Identification, expression patterns, and functional characterization of chemosensory proteins in *Dendroctonus armandi* (Coleoptera: Curculionidae: Scolytinae). Front Physiol. 2018;9:291. https://doi.org/10.3389/fphys.2018.00291
133. Chertemps T, Francois A, Durand N, Rosell G, Dekker T, Lucas P, et al. A carboxylesterase, Esterase-6, modulates sensory physiological and behavioral response dynamics in *Drosophila*. BMC Biol. 2012;10:56. https://doi.org/10.1186/1741-7007-10-56
134. Chertemps T, Younus F, Steiner C, Durand N, Coppin CW, Pandey G, et al. An antennal carboxylesterase from *Drosophila melanogaster*, Esterase 6, is a candidate odorant-degrading

enzyme toward food odorants. Front Physiol. 2015;6:315. https://doi.org/10.3389/fphys.2015.00315

135. Ishida Y, Leal WS. Chiral discrimination of the Japanese beetle sex pheromone and a behavioral antagonist by a pheromone-degrading enzyme. Proc Natl Acad Sci U S A. 2008;105(26):9076-9080. https://doi.org/10.1073/pnas.0802610105
136. Broehan G, Krober T, Lorezen M, Merzendorfer H. Functional analysis of the ATP-binding cassette (ABC) transporter gene family of *Tribolium castaneum*. BMC Genomics. 2013;14:6. https://doi.org/10.1186/1471-2164-14-6
137. Dermauw W, Van Leeuwen T. The ABC gene family in arthropods: comparative genomics and role in insecticide transport and resistance. Insect Biochem Mol Biol. 2014;45:89-110. https://doi.org/10.1016/j.ibmb.2013.11.001
138. Guo Z, Kang S, Chen D, Wu Q, Wang S, Xie W, et al. MAPK signaling pathway alters expression of midgut ALP and ABCC genes and causes resistance to *Bacillus thuringiensis* Cry1Ac toxin in diamondback moth. PLoS Genet. 2015;11:e1005124. https://doi.org/10.1371/journal.pgen.1005124
139. Leal WS. Odorant reception in insects: roles of receptors, binding proteins, and degrading enzymes. Annu Rev Entomol. 2013;58:373-391. https://doi.org/10.1146/annurev-ento-120811-153635
140. Zhu YC, Zeng F, Oppert B. Molecular cloning of trypsin-like cDNAs and comparison of proteinase activities in the salivary glands and gut of the tarnished plant bug *Lygus lineolaris* (Heteroptera: Miridae). Insect Biochem Mol Biol. 2008;33(9):889-899. https://doi.org/10.1016/j.ibmb.2008.04.001
141. Noh MY, Muthukrishnan S, Kramer KJ, Arakane Y. A chitinase with two catalytic domains is required for organization of the cuticular extracellular matrix of a beetle. PLoS Genet. 2018;14:e1007307. https://doi.org/10.1371/journal.pgen.1007307
142. Lee Y, Muthukrishnan S, Kramer KJ, Sakamoto T, Tabunoki H, Arakane Y, et al. Functional importance of groups I and II chitinases in cuticle chitin turnover during molting in a wood-boring beetle, *Monochamus alternatus*. Pestic Biochem Physiol. 2023;194:105496. https://doi.org/10.1016/j.pestbp.2023.105496
143. Kim D, Muthukrishnan S, Kramer KJ, Arakane Y, Noh MY. Requirement of group I lytic polysaccharide monooxygenase for turnover of chitinous cuticle during moulting in two forest pest beetles, *Monochamus alternatus* and *Psacothea hilaris*. Insect Mol Biol. 2026;35(2):115-125. https://doi.org/10.1111/imb.70013
144. Yeaman S. Genomic rearrangements and the evolution of clusters of locally adaptive loci. Proc Natl Acad Sci U S A. 2013;110(19):E1743-E1751. https://doi.org/10.1073/pnas.1219381110
145. Lawrence JG, Roth JR. Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics. 1996;143(4):1843-1860. https://doi.org/10.1093/genetics/143.4.1843
146. Kondrashov FA. Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc Biol Sci. 2012;279(1749):5048-5057. https://doi.org/10.1098/rspb.2012.1108

Figures

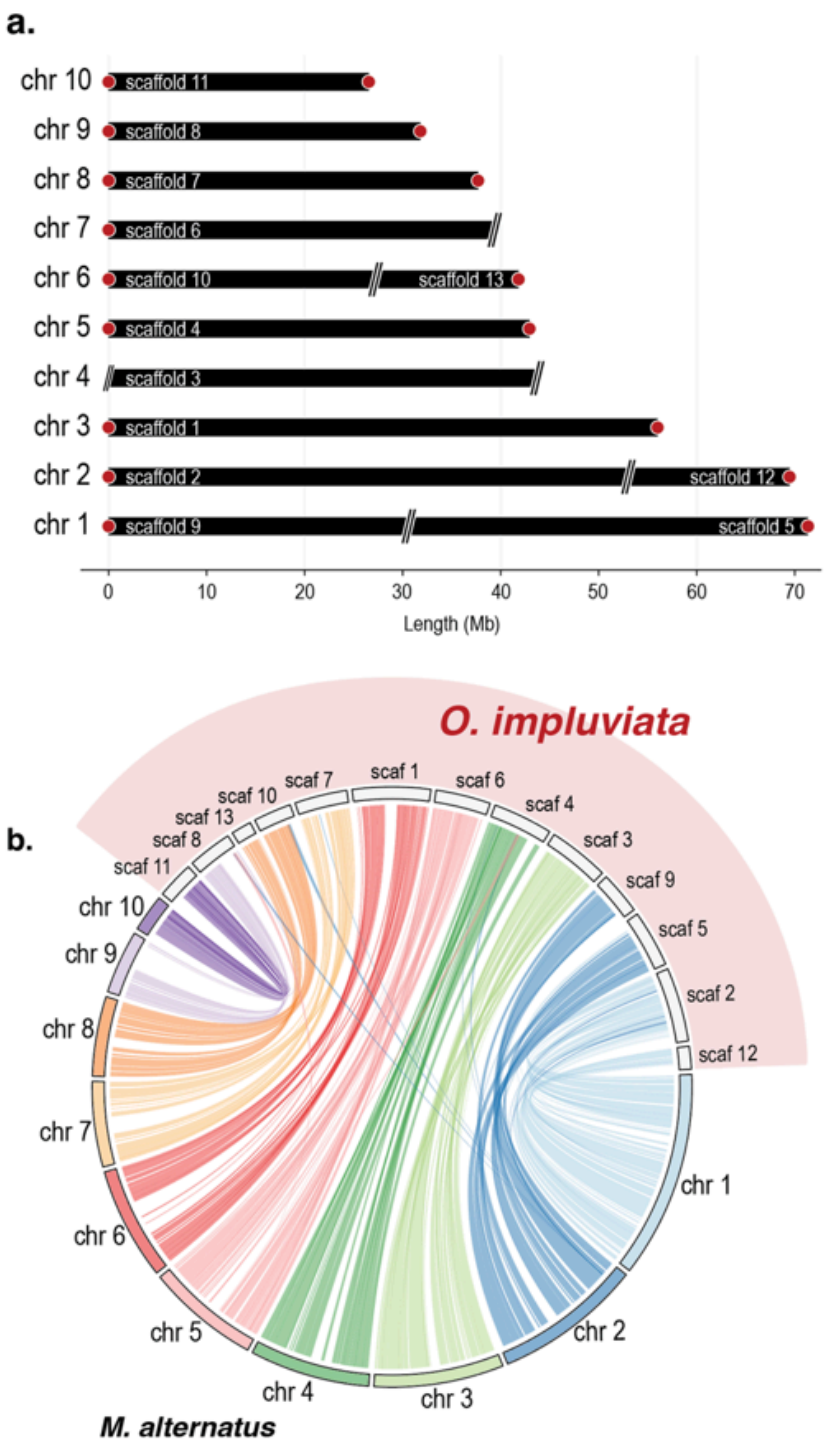


Figure 1

Chromosome-scale inference and macrosynteny of the *Oncideres impluviata* genome.
(a) Chromosome-scale organization inferred from telomeric repeat localization and scaffold assignment. Black bars represent the major *O. impluviata* scaffolds assigned to ten putative chromosomes. Red circles mark detected terminal telomeric repeat arrays. Diagonal breaks indicate inferred scaffold joins or truncated scale intervals used for visualization. Scaffold labels indicate the

assembly scaffold contributing to each inferred chromosome. **(b)** Whole-genome synteny between *O. impluviata* scaffolds and chromosomes of the Japanese pine sawyer *Monochamus alternatus*. Outer ideograms represent *O. impluviata* scaffolds and *M. alternatus*chromosomes, and internal links represent filtered syntenic alignments. Link colors correspond to *M. alternatus* chromosomes.

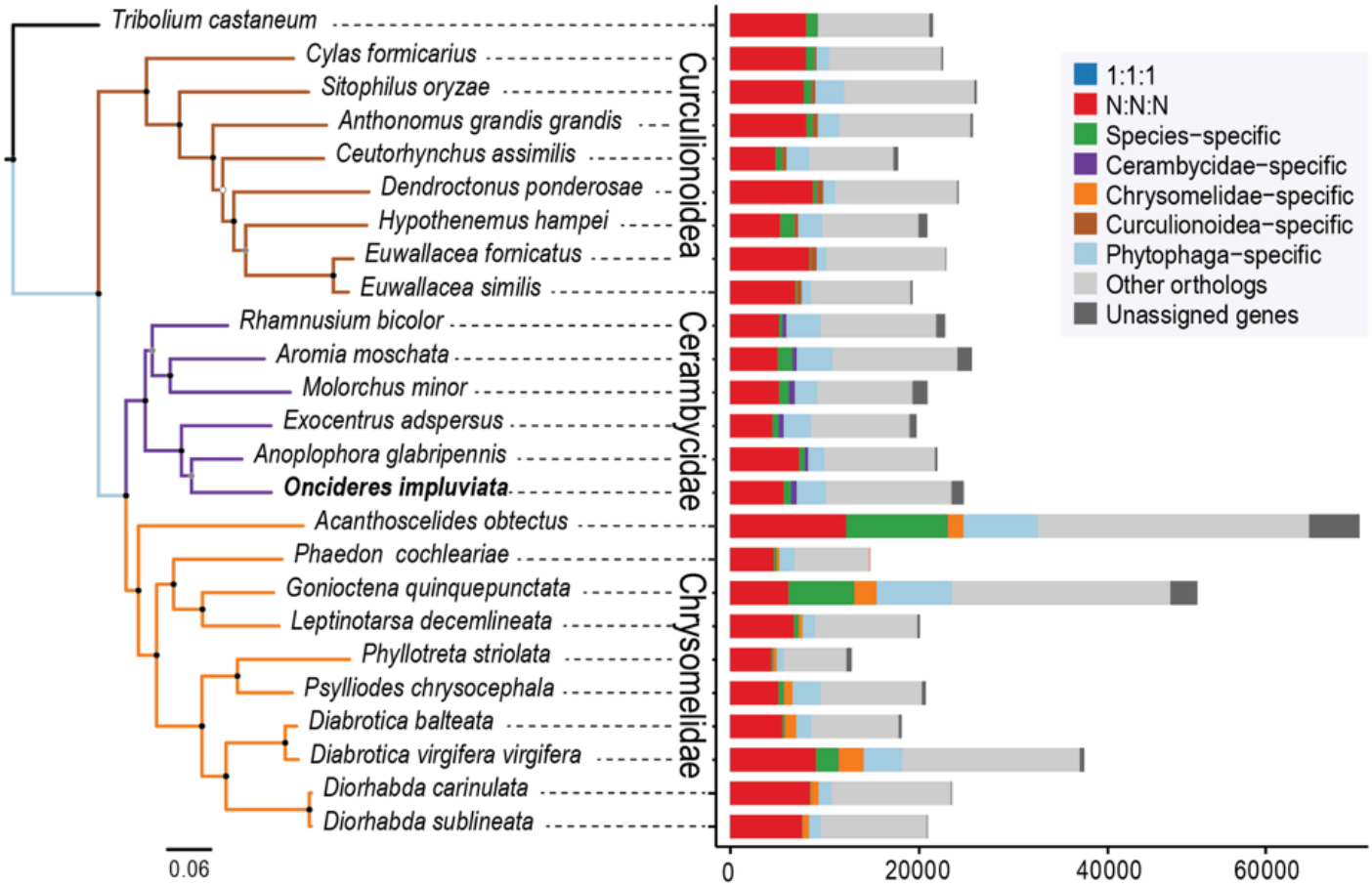


Figure 2

Orthology structure (right) across phytophagous beetles and phylogenetic placement of *Oncideres impluviata* within Cerambycidae (left). Maximum-likelihood phylogeny inferred from 47 single-copy orthologs shared across the sampled Coleoptera, rooted with *Tribolium castaneum*. Branch colors denote major phytophagous beetle lineages: Curculionoidea, Cerambycidae, and Chrysomelidae. Stacked bars summarize orthogroup membership categories for its annotated protein-coding genes for each taxon. Orthology inference was performed with OrthoFinder, and visualization was generated in R with ggplot2.

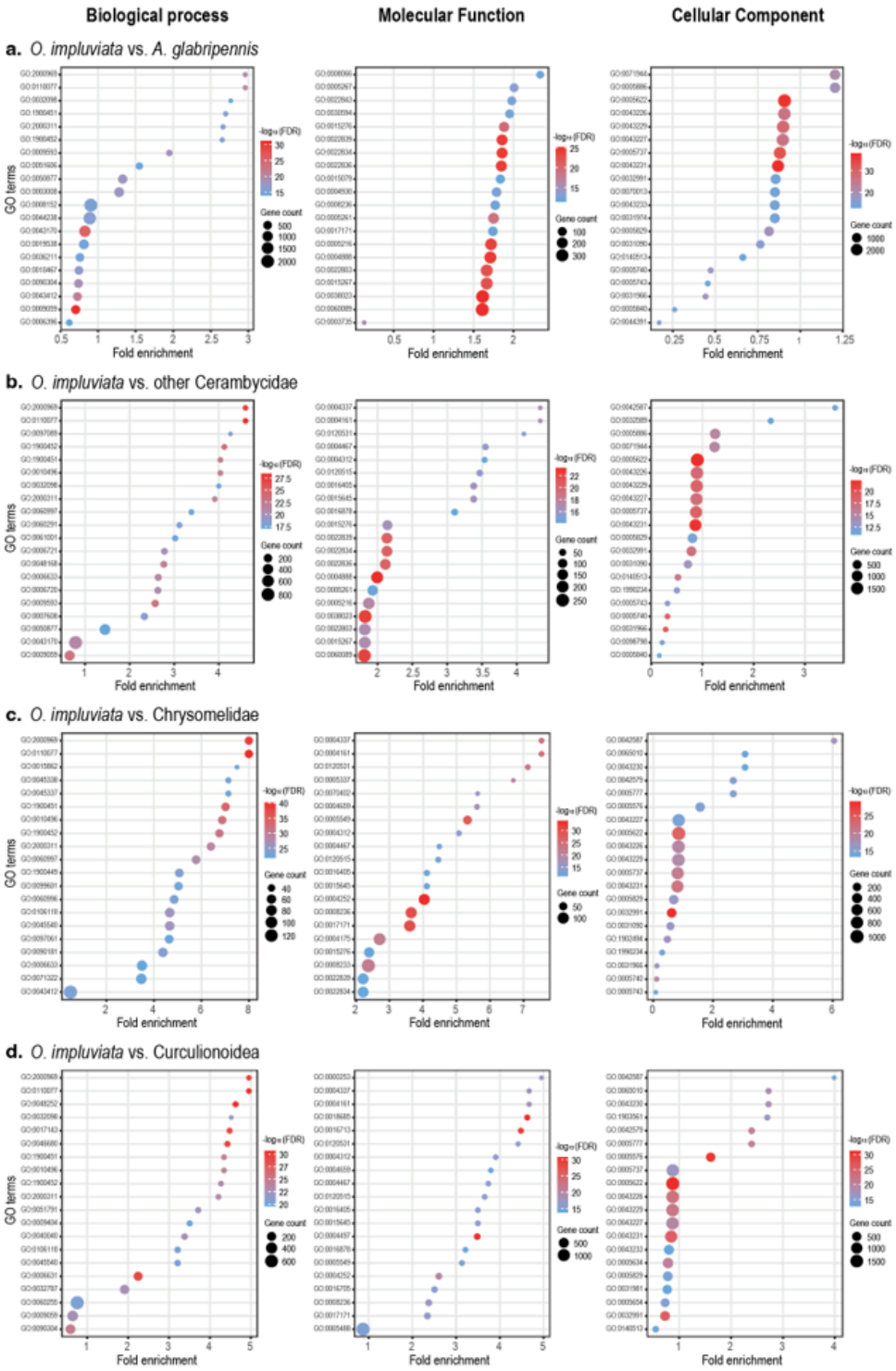


Figure 3

Gene Ontology enrichment of expanded orthogroups in *Oncideres impluviata*.
GO enrichment results for genes in orthogroups expanded in *O. impluviata*relative to four comparative baselines: **(a)** *Anoplophora glabripennis*, **(b)** other Cerambycidae, **(c)** Chrysomelidae, and **(d)**Curculionoidea. Columns show the three GO namespaces: Biological Process, Molecular Function, and Cellular Component. Each point represents an enriched GO term. Enrichment was assessed with

Fisher's exact tests in GOATOOLS using Benjamini-Hochberg correction and propagated GO counts. Terms shown represent the top enriched categories used for visualization; full GO enrichment results, including terms descriptions, are provided in Table S11 and Supplementary file 2.

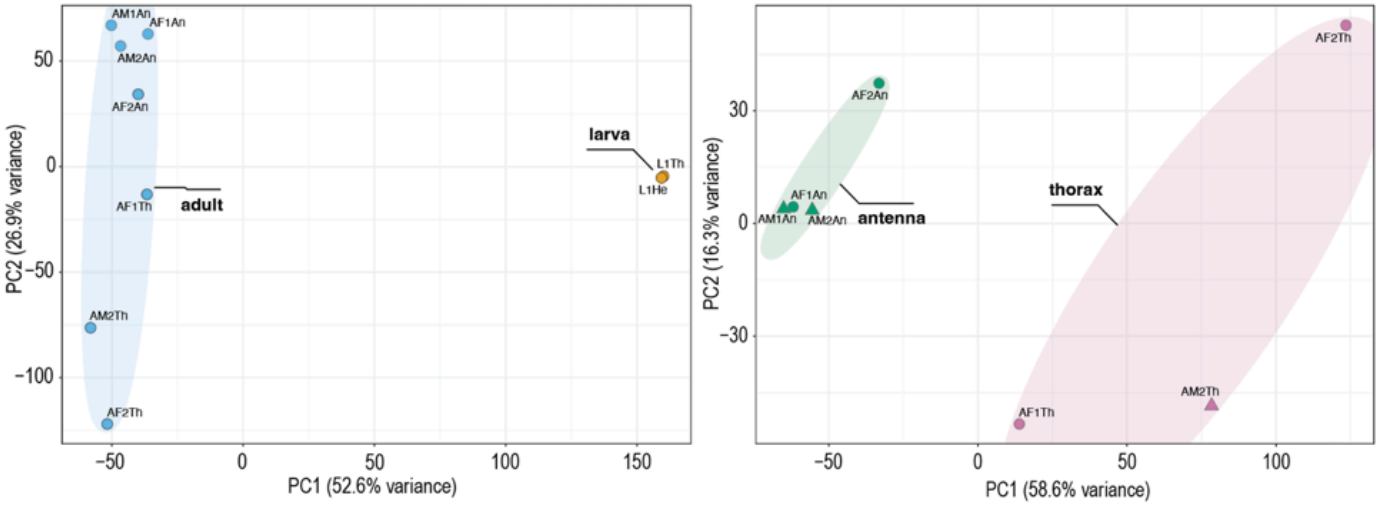


Figure 4

Transcriptome expression structure across developmental stages and adult tissues in *Oncideres impluviata*. **(a)** PCA including larval and adult samples, showing developmental stage as the major axis of transcriptomic variation. **(b)**PCA restricted to adult samples, showing stronger separation by tissue than by sex. Sample labels indicate developmental stage, sex, and tissue: A, adult; L, larva; M, male; F, female; An, antennae; Th, thorax + gut; He, head.

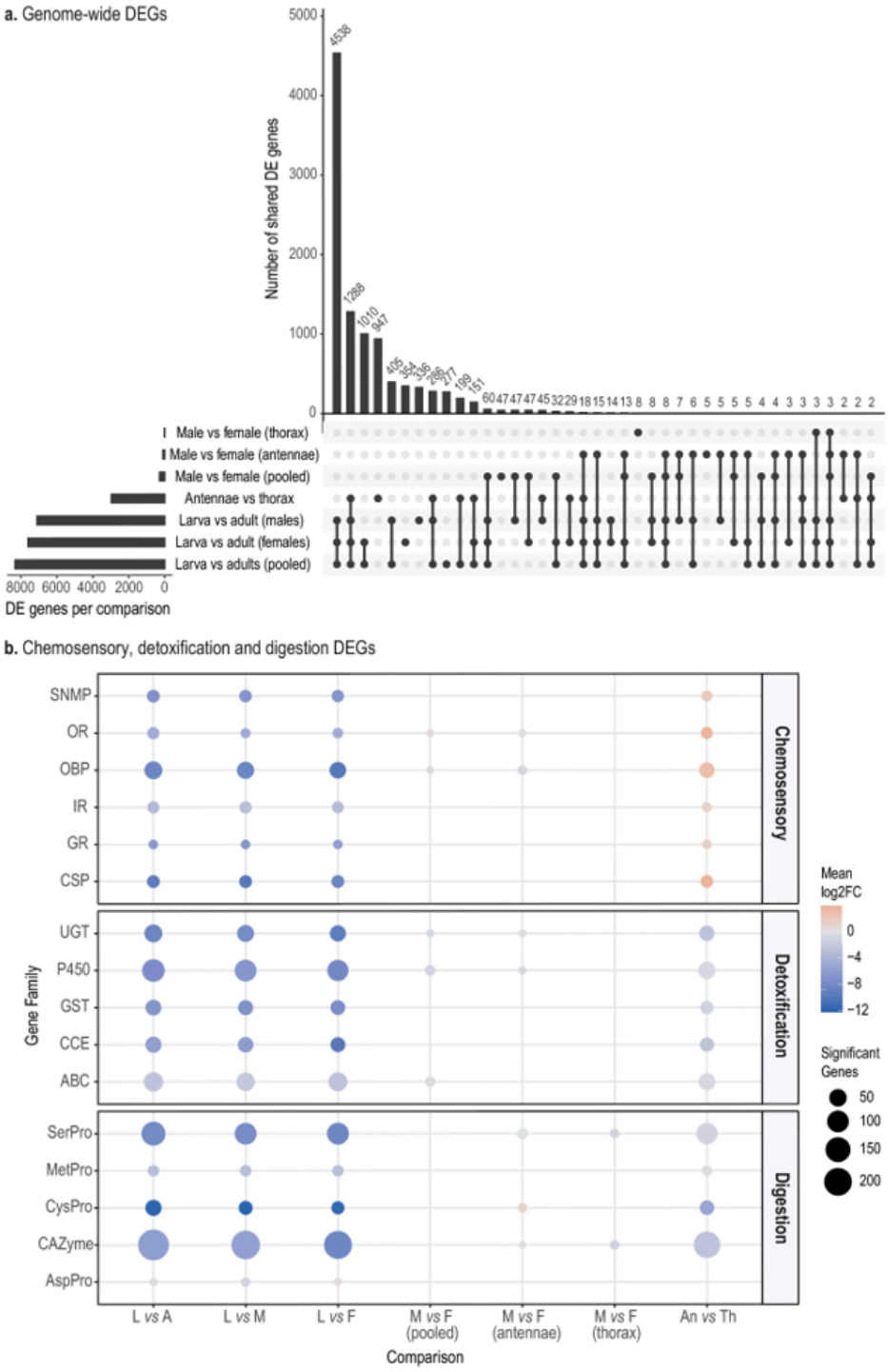


Figure 5

Overlap of differentially expressed genes across six transcriptomic contrasts in *O. impluviata*: larva *vs* adults pooled, larva *vs* adult males, larva *vs* adult females, male *vs* female pooled adults, male *vs*female antennae, male *vs* female thorax, and antennae *vs* thorax. Vertical bars show the number of genes in each intersection, and connected black points indicate which comparisons contribute to each intersection. Horizontal bars show the total number of differentially expressed genes detected in each

comparison. Differential expression was assessed with DESeq2 using Benjamini-Hochberg correction. Gene family abbreviations: chemosensory: CSP, chemosensory protein; GR, gustatory receptor; IR, ionotropic receptor; OBP, odorant-binding protein; OR, odorant receptor; SNMP, sensory neuron membrane protein; detoxification: ABC, ATP-binding cassette transporter; CCE, carboxyl/cholinesterase; GST, glutathione S-transferase; P450, cytochrome P450; UGT, UDP-glucuronosyltransferase; digestion: AspPro, aspartic protease; CAZyme, carbohydrate-active enzyme; CysPro, cysteine protease; MetPro, metalloprotease; SerPro, serine protease.

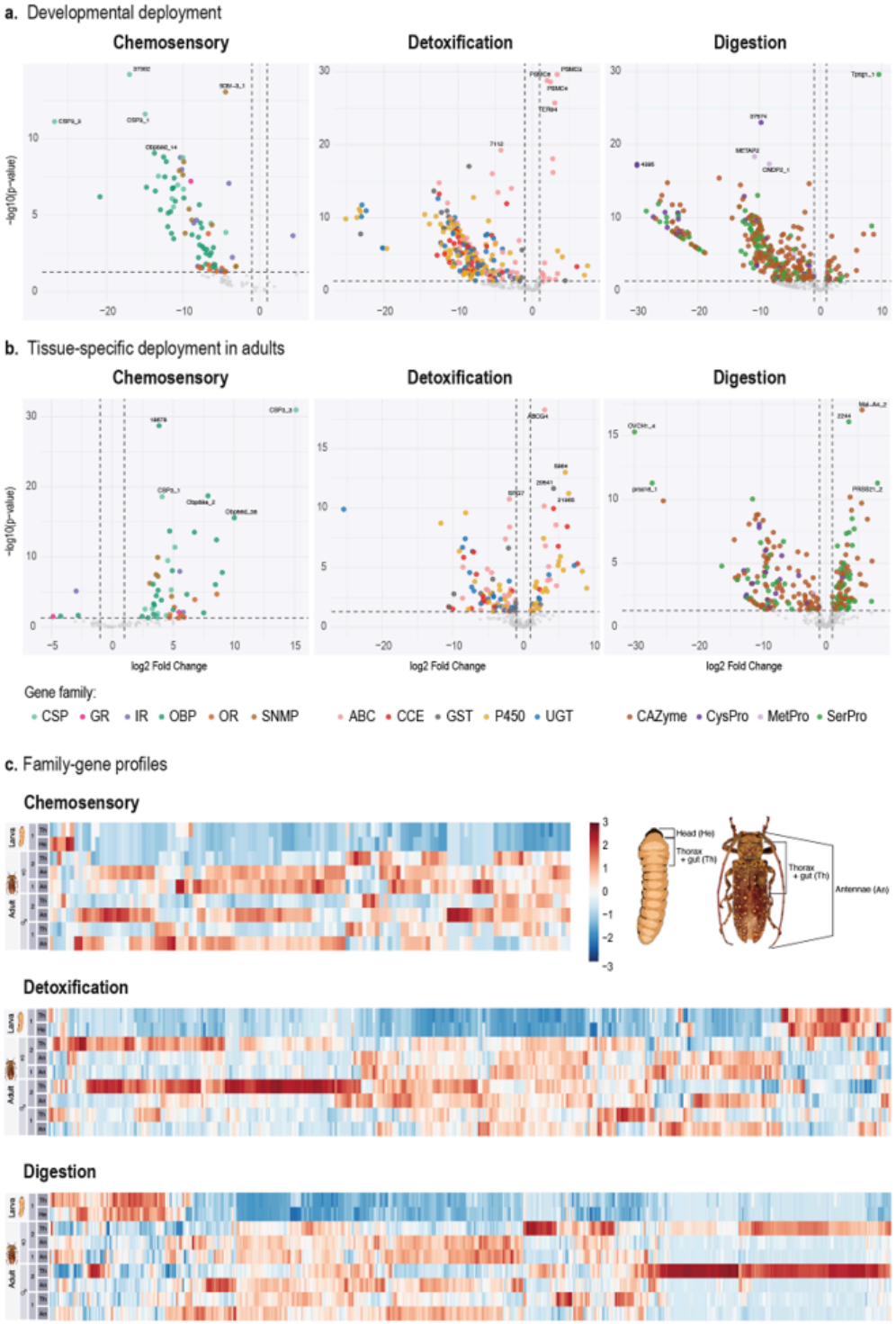


Figure 6

Developmental and tissue-specific deployment of host-interface gene families in *Oncideres impluviata*. **(a)** Volcano plots of differential expression between larvae and adults for chemosensory, detoxification, and digestion gene families. Each point represents a gene assigned to one of the curated host-interface families. Positive $\log_2$ fold-change values (log2FC) indicate larva-biased expression and negative values indicate adult-biased expression. **(b)** Volcano plots of differential expression between adult antennae and thorax for the same three functional groups. Positive $\log_2$FC values indicate antenna-biased expression and negative values indicate thorax-biased expression. In panels (a) and (b), horizontal dashed lines mark the significance threshold (adjusted $p < 0.05$), vertical dashed lines mark the effect-size cutoff ($|\text{log2FC}| > 1$), significant genes are colored by gene family, non-significant genes are shown in grey, and selected highly significant genes are labeled. **(c)** Heatmaps showing scaled expression profiles for genes in chemosensory, detoxification, and digestion families across larval and adult tissues. Rows represent samples and columns represent genes, with expression scaled by gene. Differential expression was assessed with DESeq2 using Benjamini-Hochberg correction. Numbers correspond to uncharacterized genes with the “FUN_” prefix removed. Putative BLAST-based identifications of uncharacterized genes are provided in Supplementary file 5. Gene family abbreviations are as defined in Figure 5.

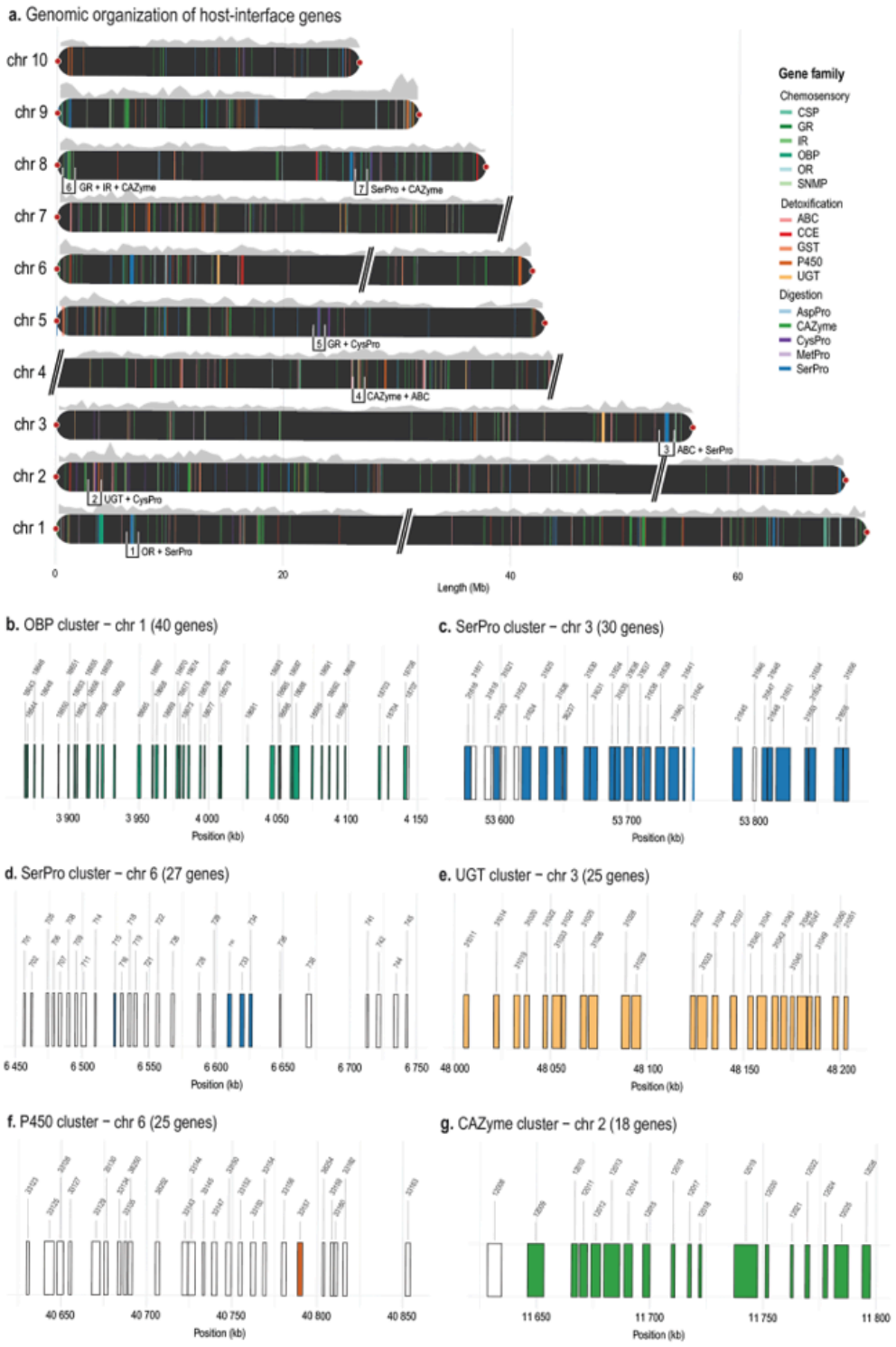


Figure 7

Chromosomal organization of host-interface genes and tandem gene clusters in *Oncideres impluviata*. **(a)** Distribution of chemosensory, detoxification, and digestion genes across the ten inferred chromosomes of *O. impluviata*. Colored vertical marks indicate genes assigned to curated host-interface families, grouped as chemosensory genes, detoxification genes, and digestion-related genes. Grey profiles above chromosomes represent gene density calculated in sliding windows. Red circles mark

terminal telomeric repeat arrays, and diagonal breaks indicate scale discontinuities used to display long chromosomes. Numbered callouts identify multi-family superloci where tandem clusters from different gene families co-localize within 50 kb. **(b–g)**Detailed views of major tandem arrays: **(b)** an OBP cluster on chromosome 1, **(c)** a serine protease cluster on chromosome 3, **(d)** a serine protease cluster on chromosome 6, **(e)** a UGT cluster on chromosome 3, **(f)**a P450 cluster on chromosome 6, and **(g)** a CAZyme cluster on chromosome 2. Colored boxes indicate genes on the sense strand (+), white boxes indicate genes on the antisense strand (−), and numbers above boxes correspond to gene identifiers with the “FUN_” prefix removed. Tandem clusters were defined as two or more genes from the same family within 100 kb on the same chromosome. Superloci were defined as regions in which clusters from different gene families co-localized within 50 kb. Chromosomal ideograms were generated in R using ggplot2.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryfile4DETargetfamilies.xlsx
- Supplementaryfile5BLAST.xlsx
- Supplementaryfile1GenomeScopeccsreads.pdf
- SupplementaryTablesS1S18.xlsx
- SupplementaryFigsS1S3.docx
- Supplementaryfile2GOenrichment.xlsx
- Supplementaryfile3DEall.xlsx