Running Head: Pheromone Receptors of *Monochamus*

Submitted to:

Journal of Comparative Physiology A

Correspondence to:
Robert F. Mitchell
Department of Biology
University of Wisconsin Oshkosh
142 Halsey Science Center
800 Algoma Blvd.
Oshkosh, WI 54901
Ph. (920) 424-0930
mitchellr@uwosh.edu

PREDICTION OF A CONSERVED PHEROMONE RECEPTOR LINEAGE FROM ANTENNAL TRANSCRIPTOMES OF THE PINE SAWYER GENUS *MONOCHAMUS* (COLEOPTERA: CERAMBYCIDAE)

ROBERT F. MITCHELL[a*], DANIEL DOUCET[b], SUSAN BOWMAN[b], MARC C. BOUWER[b], AND JEREMY D. ALLISON[b]

[a] Department of Biology, University of Wisconsin Oshkosh, Oshkosh, WI 54901, USA

[b] Natural Resources Canada, Canadian Forest Service, Great Lakes Forestry Centre, Sault Ste. Marie, ON P6A2E5, Canada

*To whom correspondence should be addressed.

Full Address: 142 Halsey Science Center, 800 Algoma Blvd., Oshkosh, WI 54901

Email: mitchellr@uwosh.edu | Phone: 920-424-0930 | Fax: 920-424-1101

Abstract

Longhorned beetles (Cerambycidae) are a diverse family of wood-boring insects, many species of which produce volatile pheromones to attract mates over long distances. The composition and structure of the pheromones remain constant across many cerambycid species, and comparative studies of those groups could therefore reveal the chemoreceptors responsible for pheromone detection. Here, we use comparative transcriptomics to identify a candidate pheromone receptor in the large and economically important cerambycid genus *Monochamus*, males of which produce the aggregation-sex pheromone 2-(undecyloxy)-ethanol ("monochamol"). Antennal transcriptomes of the North American species *M. maculosus*, *M. notatus*, and *M. scutellatus* revealed 60–70 odorant receptors (ORs) in each species, including four lineages of simple orthologs that were highly conserved, highly expressed in both sexes, and upregulated in the flagellomeres where olfactory sensilla are localized. Two of these orthologous lineages, OR29 and OR59, remained highly expressed and conserved when we included a re-annotation of an antennal transcriptome of the Eurasian congener *M. alternatus*. OR29 is also orthologous to a characterized pheromone receptor in the cerambycid *Megacyllene caryae*, suggesting it as the most likely candidate for a monochamol receptor and highlighting its potential as a conserved lineage of pheromone receptors within one of the largest families of beetles.

Keywords: longhorned beetle, chemoreceptor, odorant receptor, 2-(undecyloxy)-ethanol, monochamol

Introduction

The genetics of pheromone detection in insects is of great interest because of its pivotal role in evolutionary biology and behavior (e.g., Leary et al. 2012) and its many potential applications in pest management (Anderson and Newcomb 2021). Odorant receptors (ORs) are the largest radiation of olfactory receptors in insects and thus the most likely genes to be involved in the detection of long-range volatile pheromones, with numerous such pheromone receptors (PRs) described from insects over the past two decades (Fleischer and Krieger 2021). The best studied PRs are those of the Lepidoptera, which form distinct radiations linked to the various sex pheromone structural motifs produced by moths (Montagné et al. 2021). However, ORs are not conserved across insect orders except for the co-receptor Orco (Hansson and Stensmyr 2011), and PRs of one insect order cannot inform their discovery elsewhere. Thus, PRs remain poorly described from most insects, and especially from the large and diverse Coleoptera (Mitchell and Andersson 2021). The few described beetle PRs include those sensitive to the pheromones ipsenol and ipsdienol produced by the bark beetle *Ips typographus* (L.) (Yuvaraj et al. 2021), as well as aggregation-sex pheromones produced by the palm weevil *Rhynchophorus ferrugineus* (Olivier) (Curculionidae; Antony et al. 2021) and hickory borer *Megacyllene caryae* (Gahan) (Cerambycidae; Mitchell et al. 2012).

We seek to identify additional ORs associated with the pheromone biology of the large family of longhorned beetles (Cerambycidae). Cerambycid larvae feed in the wood of trees and many species are notorious pests of forests (Solomon 1995), especially as invasive species (Haack et al. 2010). Thus, the attractive pheromones produced by adults are important tools to monitor populations of pests, and an extensive literature of pheromone structures has emerged over recent decades (Allison et al. 2004; Millar and Hanks 2017). This has revealed a peculiar

aspect of cerambycid pheromone biology: many species produce identical pheromone structures, whether in sympatry (Mitchell et al. 2015), across continents (Wickham et al. 2021), or across subfamilies (Ray et al. 2012; 2015). It is likely that conserved pheromone structures in closely related species stem from a common ancestor. Assuming that their detection mechanism is equally conserved, a simple comparison of ORs within a genus could therefore expedite the identification of pheromone receptors by revealing ORs that are orthologous and highly expressed across the multiple species that produce the pheromone. In fact, similar approaches have already been productive in identifying receptors for (E)-β-farnesene, an alarm pheromone produced by numerous species of aphids (Zhang et al. 2017).

We tested our hypothesis in the cerambycid genus *Monochamus*, which includes conifer-feeding species that damage their host trees (Cerezke and Volney 1995) and vector the destructive pine wood nematode *Bursaphelenchus xylophilus* (Steiner & Buhrer) (Morimoto and Iwasaki 1972). Males of many North American and Eurasian *Monochamus* spp. produce 2-(undecyloxy)-ethanol ("monochamol") as the single component of an aggregation-sex pheromone that attracts adult females and males (Pajares et al. 2010; Allison et al. 2012; Fierke et al. 2012; Ryall et al. 2015). Monochamol in turn is part of a broader motif of hydroxyether pheromones documented from other cerambycid species in the tribe Lamiini, including 2-(4-heptyloxy-1-butyloxy)-1-ethanol in the African congener *M. leuconotus* (of uncertain behavioral function; see Pajares et al. 2010) and 4-(heptyloxy)butan-1-ol and its aldehyde as aggregation-sex pheromones in the Asian longhorned beetle *Anoplophora glabripennis* (Motschulsky) and citrus longhorned beetle *A. chinensis* (Förster) (Zhang et al. 2002; Hansen et al. 2015). A receptor for monochamol would therefore be useful in understanding not only the pheromone

biology of *Monochamus*, but also the other pest genera of cerambycid beetles that produce hydroxyethers.

We predicted that an OR sensitive to monochamol would exhibit the following characteristics: 1) be present as an ortholog and under purifying selection in all species of *Monochamus* that produce monochamol; 2) be highly expressed relative to all other ORs, which is typical of beetle pheromone receptors (Mitchell et al. 2012; Yuvaraj et al. 2021; Antony et al. 2021); 3) be expressed equally in both sexes, because both sexes are highly attracted to the pheromone (Fierke et al. 2012; Ryall et al. 2015); and 4) be upregulated in the terminal antennal flagellomeres, because olfactory sensilla have been reported to increase in number across the length of the flagellum (Dyer and Seabrook, 1975). Here, we test these assumptions by sequencing, annotating, and measuring the expression of odorant receptors from antennal transcriptomes of three species of *Monochamus* endemic to North America that are attracted to monochamol: *M. maculosus* (syn. *mutator*) (Haldeman), *M. notatus* (Drury), and *M. scutellatus* (Say) (Fierke et al. 2012; Ryall et al. 2015). We then compare the resulting candidate genes to ORs of the Eurasian species *M. alternatus* and to characterized pheromone receptors from other beetle species to determine the most likely receptor for monochamol.

Materials and methods

Source of specimens

Adult specimens of *M. maculosus*, *M. notatus*, and *M. scutellatus* were collected from multi-funnel traps baited with monochamol and α-pinene and placed in July–August 2016 at a clear cut of coniferous forest near Aubrey Falls, ON, Canada. Antennal sections were then

removed from females and males of each species, homogenized in TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, USA), and stored at -80 °C until ready for RNA extraction. To determine which ORs were upregulated in the terminal flagellomeres, each antenna was sectioned into a "tip" that consisted of the four apical flagellomeres and a "base" consisting of the five (largely non-olfactory) basal flagellomeres.

Extraction and sequencing of RNA and transcriptome assembly

We prepared 36 tissue isolates for sequencing, each consisting of separated antennal tips or bases from females or males of the three *Monochamus* spp. (antennal region × sex × species, in triplicate). Each isolate was created from pooled antennal tips or bases of three individuals. RNA from TRIzol homogenates was purified and treated with RNAse-free DNAse (Qiagen, Hilden, Germany) to remove potential co-purifying genomic DNA. Sequencing libraries were created using the Illumina TruSeq Stranded mRNA kit (Illumina, San Diego, CA, USA). Paired-end sequence reads (150 nt) were generated on the Illumina HiSEQ2500 platform. Libraries and sequences were generated at the Toronto Center for Applied Genomics (TCAG, Hospital for Sick Children, Toronto, ON, Canada). Reference transcriptomes for each species were assembled from pooled female and male reads with CLC Genomics Workbench (Qiagen), using default parameters.

Annotation and phylogeny of OR genes

Models were identified from the CLC Bio assemblies through BLAST searches that initially used queries of OR protein sequences from the cerambycid *A. glabripennis* (Mitchell et al. 2017), followed by iterative searches with each new *Monochamus* model as a query. Once these de novo assemblies were exhausted of new models, we used the pooled cerambycid OR proteins as BLAST queries against the raw sequence reads and re-assembled the resulting

database of unique hits to identify any novel ORs. Finally, the pooled raw reads of each *Monochamus* sp. were mapped (Geneious Prime 2019.2 mapper, Biomatters Ltd., Auckland, New Zealand) to the OR models of the other two species to identify and assemble any remaining orthologs, and we manually extended incomplete ORFs via BLAST searches of the 5' and 3' ends versus the reads. Models that remained incomplete were denoted with suffixes that indicated missing N-terminal (NTE), C-terminal (CTE) or internal (INT) sequence data. If multiple suffixes applied to a model, they were abbreviated to single letters (e.g., MnotOR16NC is presumed to be missing N-terminal and C-terminal coding sequence). Models <300 bp in length were discarded from the analysis.

We assumed that candidate receptors for monochamol would present as simple orthologs, so we only considered ORs as pheromone receptors if they could be annotated as full-length 1:1 orthologs in all three *Monochamus* spp. ("orthologous triplets"). However, orthologous triplets containing partial genes were also considered if at least one full-length ortholog was present and the remaining genes were >95% of that predicted length. We determined evolutionary relationships by constructing a phylogenetic tree comparing the aligned protein sequences (MUSCLE, gap penalty -5; Edgar 2004) of the ORs of *A. glabripennis* and the three *Monochamus* spp. using PhyML 3.0 (Guindon et al. 2010), implementing a JTT+I+G+F model for gene evolution (selected using PartionFinder 2; Lanfear et al. 2016). Branch support was calculated using transfer bootstrap analysis, a modification of canonical bootstrapping that is more appropriate for large numbers of taxa (Lemoine et al. 2018). The tree also included all ORs that had been functionally characterized to date in beetles: three pheromone receptors for 2-methylbutan-1-ol, 2-phenylethanol, and 2,3-hexanediol in the cerambycid *M. caryae* (McarOR3, 5, and 20, respectively; Mitchell et al. 2012); two pheromone receptors for ipsenol and ipsdienol

(ItypOR46 and 49, respectively) and others to similar terpenoid structures in the bark beetle *I. typographus* (Hou et al. 2021; Yuvaraj et al. 2021); one pheromone receptor sensitive to both ferrugineol and ferrugineone in the palm weevil *R. ferrugineus* (RferOR1; Antony et al. 2021); a receptor for the host volatile α-pinene, again in *R. ferrugineus* (RferOR6; Ji et al. 2021); and a receptor for the host volatiles hexanal, lauric acid, and tetradecane in the scarab *H. parallela* (HparOR27; Wang et al. 2021).

ORs in novel species are usually named by progressive enumeration of each gene as it is annotated (e.g., OR1, OR2, OR3) because orthologous lineages are rarely maintained across beetle families (Mitchell et al. 2020). However, the genera *Monochamus* and *Anoplophora* are very closely related (Gorring 2019), and we were able to assign informative names to ORs from *Monochamus* spp. following the genomic annotation of *A. glabripennis* (McKenna et al. 2016; Mitchell et al. 2017). ORs from the three *Monochamus* spp. received the same gene number as their ortholog in *A. glabripennis* if they were monophyletic in the tree and in a 1:1 relationship. If *Monochamus* ORs had instead radiated as paralogs, they received numeric suffixes based on the homolog in *A. glabripennis* (e.g., MmacOR37.1 and MmacOR37.2 are paralogs of AglaOR37). If an OR lineage from *Monochamus* had no clear ortholog in *A. glabripennis*, it was provisionally named following the closest (non-pseudogenic) homolog, as judged by position in the tree and bit scores from BLAST hits.

Expression analysis

A candidate pheromone receptor would be highly expressed relative to other ORs in the antennae of all three *Monochamus* spp., so we calculated TPM (transcripts per million reads) of genes present in the antennal tips using Geneious. Assemblies of each species were used as the template for mapping, although modified by first removing any models of ORs native to the

assembly, and then merging the assembly with that species' final annotated OR set. Although TPM is intended to normalize for comparisons, it is still not appropriate to directly compare numeric values across different transcriptomes (Zhao et al. 2020). This can be addressed by instead comparing the relative TPM ranks (Zhao et al. 2021); thus, we ranked our pool of candidate pheromone receptors by their TPM within each transcriptome, resulting in 43 ranked ORs per species (see Results). We then averaged the rank of each orthologous triplet across all antennal tip samples and across species to quantify its expression. For example, a triplet with a mean TPM rank of 5 indicates an OR that was expressed highly in all three species, while a mean rank of 30 indicates an OR that was expressed at low levels or inconsistently across species. The ten triplets with the highest mean rank (representing the upper quartile) were considered as the most highly expressed ORs in the transcriptomes, and thus as candidate pheromone receptors.

Candidate pheromone receptors were furthermore assumed to be upregulated in the tip of the antenna relative to the base and expressed similarly in both sexes. We therefore used DESeq2 (parametric model; Love et al. 2014) to measure differential expression of ORs in each of the three *Monochamus* spp. The same transcriptome assemblies as above were used for mapping. We then compared expression of ORs within each species in 1) antennal tips vs. bases (N=6), and 2) antennal tips of females vs. males (N=3). Upregulation of an orthologous triplet was predicted only if a significant DE ($P < 0.05$, FDR adjusted) was obtained from the gene in all three species.

Evolutionary analysis

Candidate monochamol receptors should maintain a high sequence similarity due to their conserved function, so we measured the total number of sites across each orthologous triplet

(aligned in MAFFT 7.450, default settings; Katoh and Standley 2013) that included nonsynonymous substitutions. We also measured the dN/dS ratio (ω), which quantifies the ratio of substitution rates at non-synonymous (dN) and synonymous (dS) sites. Selection is assumed to be negative (purifying) as ω approaches 0, and positive when ω exceeds 1. The dN/dS for each orthologous triplet was calculated using the codeml suite of PAML (codon-based analysis, NSsites=0, default settings) in PAMLX (Yang 2007; Xu and Yang 2013), using individual gene trees generated from FastTree 2.1.11 at its default settings (Price et al. 2010). Low total numbers of substitutions in tandem with $\omega < 1$ were considered as additional evidence of pheromone sensitivity.

Reannotation of olfactory receptors from Monochamus alternatus

The three *Monochamus* spp. in our study have been diverging for ~2 million years (Gorring 2019), but we wished to extend our predictions to more distant species. *Monochamus alternatus*, the Japanese pine sawyer, is a Eurasian species that produces and responds to monochamol (Teale et al. 2011), but it last shared a common ancestor with North American congeners 5.33 mya (Gorring 2019). A previously published antennal transcriptome of this species (derived from pooled antennae of females and males) included an initial annotation of only nine ORs (Wang et al. 2014), so we reanalyzed this dataset to determine if additional ORs were present, and if these ORs were orthologous to and consistent with our candidate pheromone receptors. We reannotated ORs of *M. alternatus* by mapping the published transcriptome reads (NCBI SRA Accession SRP033376) to sequences of each OR lineage from the three North American *Monochamus* spp. MaltORs were named following their orthologs in the North American species, which were immediately evident from mapping results and sequence similarity. Since these data were sequenced from a single transcriptome, the relative expression

of each MaltOR was calculated by FPKM (mapped fragments per effective kilobase of exon, per million mapped fragments), based on a draft assembly generated from the Geneious Prime 2021.2 assembler at its default settings. The ten most highly expressed ORs were compared to pheromone receptor candidates from North American *Monochamus* spp. As above, we counted nonsynonymous substitutions present in those genes and we recalculated dN/dS for the regions with sequence coverage in *M. alternatus*.

Results

Annotation and phylogeny of ORs

We annotated 73 ORs from *M. maculosus*, 68 ORs from *M. notatus*, and 60 ORs from *M. scutellatus* in addition to a single Orco gene from each species. The majority of the ORs were complete ORFs, but we included 28 partial models in *M. maculosus*, 24 partial models in *M. notatus*, and 18 partial models in *M. scutellatus*. Notes on ORFs and sequence data for all genes are included in Online Resource 1, and transcriptome reads are deposited in the NCBI Short Read Archive (BioProject ID PRJNA821225, accession numbers SRX15155624–SRX15155659).

The phylogeny (Fig. 1) sorted the ORs into the expected eight subfamilies documented to date from the Cerambycidae (all but Group 6; Mitchell et al. 2020), save for an apparent paraphyly of Group 2A and one divergent lineage of 2B genes (OR69+70), both cases likely arising because of the limited diversity of beetle species in this tree (Mitchell et al. 2020). Each of the OR subfamilies was prominently represented in *Monochamus* except for 5A, a subfamily of coleopteran ORs whose members are rarely expressed in the antennae (Mitchell and

Andersson 2021). Seven lineages of *Monochamus* ORs had radiated relative to genes of the related cerambycid *A*. *glabripennis* (OR13, 21, 37, 80, 85, 100, and 108), but radiations of OR21 and OR100 could not be cleanly assembled from all three species, and thus may have incorporated allelic variation or omitted cryptic paralogs. Additionally, two ORs may be pseudogenes, although they were included here as functional ORs. MscuOR16 was assembled with a stop codon in the ORF, but it was not consistent in the raw transcriptome reads, so the model is presented as the functional allele. However, MscuOR110 and MnotOR110 both include a stop codon in their penultimate exon that is invariant in the reads. The expected terminal exon is transcribed in both genes, although it is now out of frame. Since the termination signal was conserved across species, we have annotated it as a true stop codon, although the resulting protein has lost its final transmembrane domain and may have novel function. The apparent ortholog AglaOR111 (Fig. 1) is also pseudogenized in *A*. *glabripennis*, but much more clearly so as the model in the genome is highly fragmented (Mitchell et al. 2017).

Most of the remaining ORs were organized into 55 lineages of simple orthologs maintained across all three *Monochamus* spp. ("orthologous triplets"; Fig. 1). Of these, 43 were annotated to full-length (or nearly so) in all three species and considered further in our analyses. Thirty-eight of the orthologs also extended to the related species *A*. *glabripennis*, and two such lineages included potential orthologs of characterized receptors in other beetle species: OR29 included McarOR3 (sensitive to the pheromone component 2-methylbutan-1-ol of the cerambycid *M*. *caryae*) and OR54 included HparOR27 (sensitive to several host volatiles relevant to the scarab *Holotrichia parallela* (Motschulsky); Fig. 1).

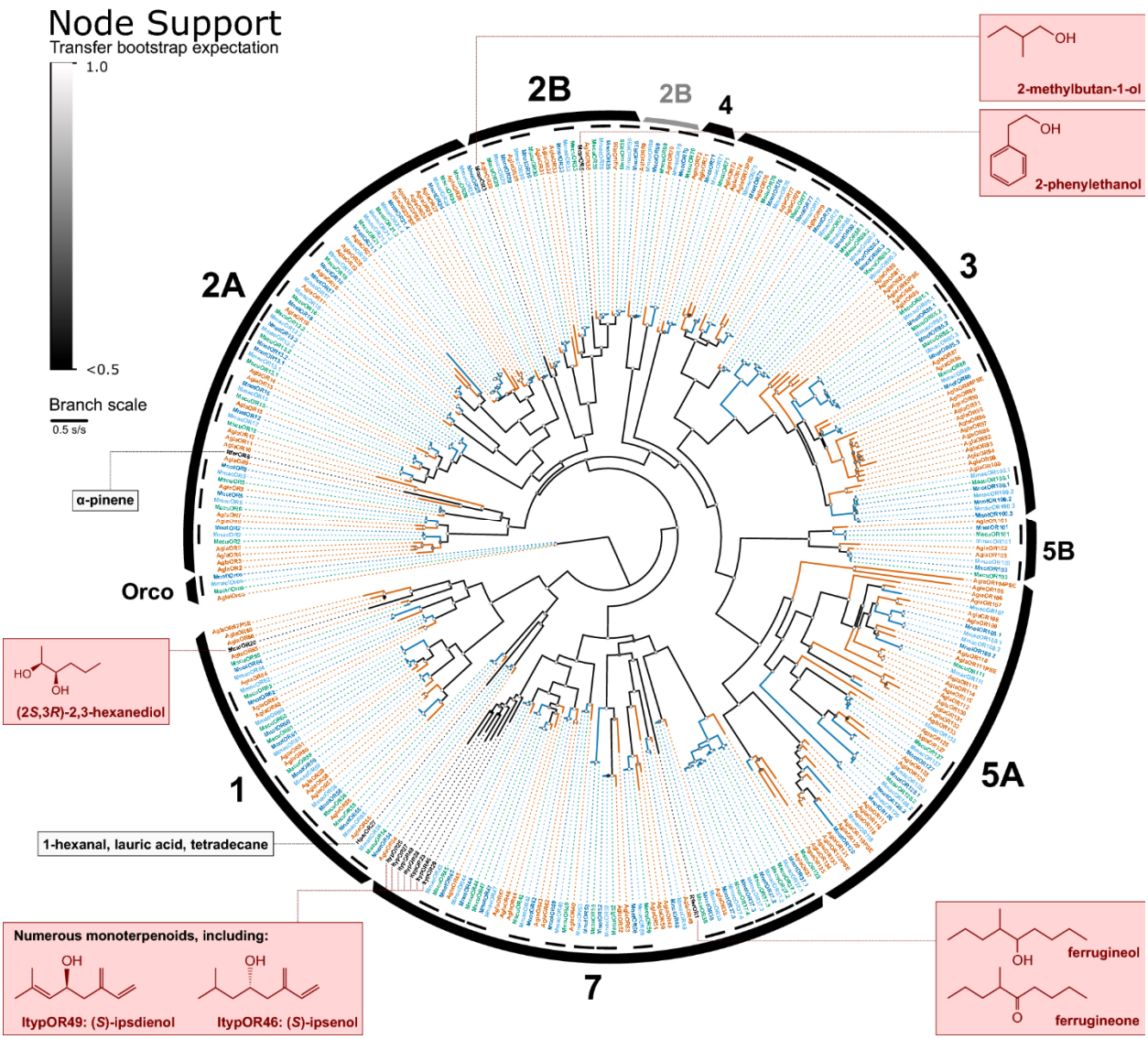

Figure 1. Phylogeny illustrating the relationships between odorant receptors (ORs) of *Anoplophora glabripennis* (Agla) and three species of *Monochamus*: *M. maculosus* (Mmac), *M. notatus* (Mnot), and *M. scutellatus* (Mscu). Functionally characterized ORs are also included from the species *Holotrichia parallela* (Hpar), *Ips typographus* (Ityp), *Megacyllene caryae* (Mcar), and *Rhyncophorus ferrugineus* (Rfer), with the confirmed ligands and pheromone structures inset. Thick black arcs delineate the recognized subfamilies of coleopteran ORs, with a gray arc indicating divergent ORs of subfamily 2B (see text) and thin black arcs indicating lineages of orthologs present in all three species of *Monochamus* ("orthologous triplets"). Shaded circles on nodes represent branch support by transfer bootstrap analysis (removed for clarity within orthologous triplets of *Monochamus* ORs)

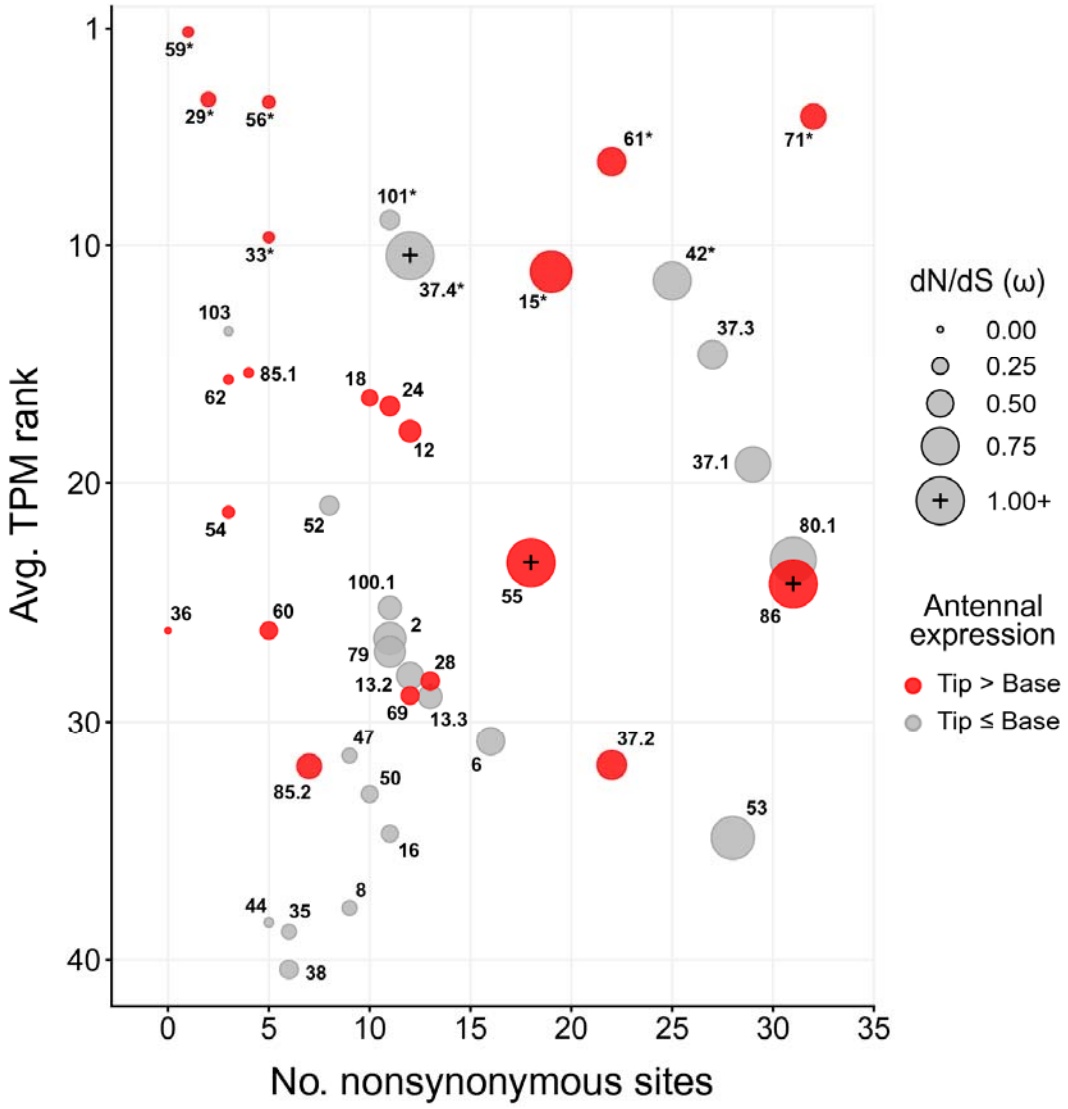

Figure 2. Bubble plot comparing all 43 orthologous lineages of odorant receptors (ORs) that were sequenced to full-length in *Monochamus maculosus*, *M. notatus*, and *M. scutellatus*. Each lineage is positioned based on the number of sites that have experienced nonsynonymous mutations (x-axis) and its expression within transcriptomes (y axis; averaged rank across species of Transcripts per Million Reads, TPM), with a higher expression as its rank approaches 1. Data points increase in size with the ratio of synonymous to nonsynonymous mutations (dN/dS, ω), and are marked with a '+' sign if the ratio exceeds 1 and suggests positive selection. Data points are shaded as light gray if any of the orthologs were not significantly ($P < 0.05$) expressed in the tip (four apical flagellomeres) of the antenna relative to the base (five basal flagellomeres). Asterisks next to OR numbers indicate the ten most highly expressed ORs

Expression analysis

Ranking the orthologous triplets by their expression revealed ten lineages (OR15, 29, 33, 37.4, 42, 56, 59, 61, 71, and 101) that were most consistently and highly expressed in the antennae of all three *Monochamus* spp. (Fig. 2), and thus consistent with our assumptions of a pheromone receptor. Expression of OR30 was also high (Online Resource 2), yet this gene could not be recovered to full length in either *M. maculosus* (92% coverage) or *M. notatus* (78%) and so was not considered as part of our analysis.

We further compared differential expression between the tip of the antenna and its base to determine which ORs were most likely to be involved in olfaction. Roughly half (21) of the orthologous triplets were consistently upregulated in the antennal tip of all three *Monochamus* spp., but excluding OR101, 37.4, and 42 of the highly expressed ORs noted above (Fig. 2). A comparison of females and males indicated that none of the triplets were consistently upregulated in the antennal tips of one sex, and only the single genes MscuOR71 and MmacOR108.1 were significantly sex biased (in females and males, respectively; Online Resource 2).

Evolutionary analysis

Sequence conservation was high within all orthologous triplets (Fig. 1), and they rarely exhibited more than 30 nonsynonymous sites (Fig. 2). Most triplets exhibited signatures of purifying selection based on a dN/dS ratio of $\omega < 1.0$, and OR36 maintained a singular amino acid identity across the three *Monochamus* spp. (Fig. 2). Four of the highly expressed triplets (OR29, 33, 56, 59) also experienced a notably decreased mutation rate, with <5 cumulative amino acid differences across each lineage.

Selection of candidate monochamol receptors and comparison to Monochamus alternatus

We predicted that a receptor for monochamol would be at the union of ORs that were highly expressed, biased in expression toward the antennal tip, and highly conserved across all *Monochamus* spp. Applying those criteria to our data yielded four candidates: OR29, 33, 56, and 59 (Fig. 3). We then tested this hypothesis by reannotating a published antennal transcriptome of the Eurasian congener *M. alternatus* (Wang et al. 2014).

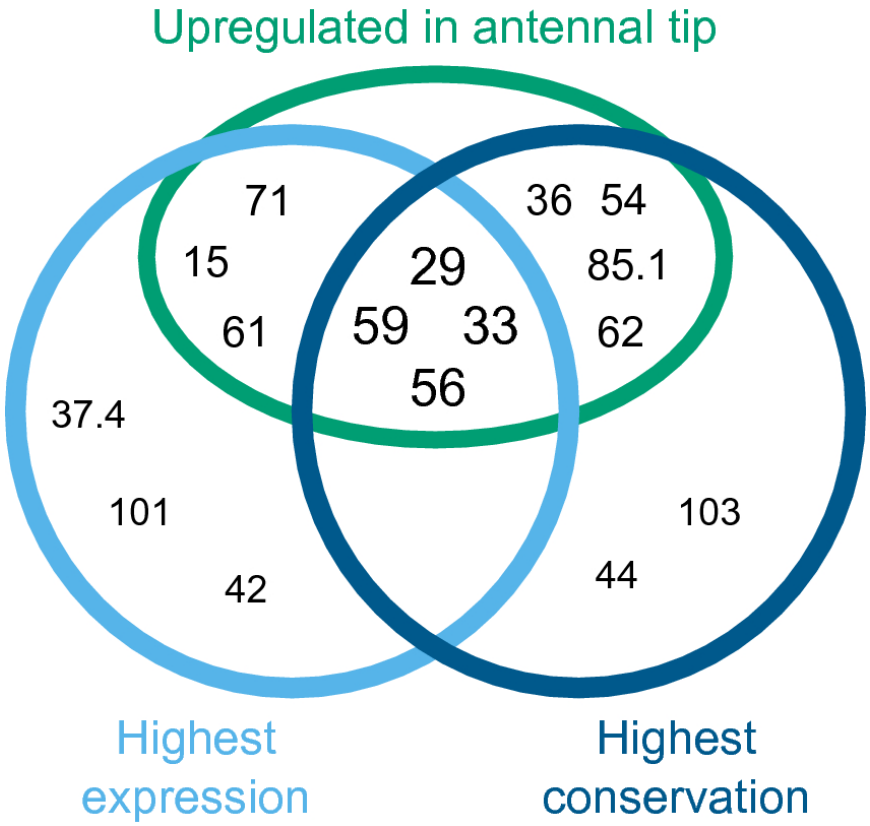

Figure 3. Venn diagram illustrating the ten orthologous lineages of odorant receptors (ORs) in *Monochamus maculosus*, *M. notatus*, and *M. scutellatus* that were superlative in antennal transcriptome expression or sequence conservation. Lineages are further separated by whether they were consistently upregulated in the antennal tip (four apical flagellomeres) relative to the base (five basal flagellomeres). Lineages at the union of the diagram are candidate receptors for the sex-aggregation pheromone 2-(undecyloxy)-ethanol ("monochamol"), produced by adult males of all three *Monochamus* spp

The previous annotation effort (Wang et al. 2014) had identified nine putative OR models, which we have clarified here as fragments of six ORs and a partial model of Malt\Orco. Our re-annotation recovered a complete Orco gene and orthologs of 49 of the lineages identified from the North American species. Only the Orco model was full length, but several orthologs were sequenced to at least 90% of their predicted length (Table 1). Correspondence with gene names from Wang et al. (2014) and the nucleotide and protein models of *M. alternatus* ORs are included in Online Resource 3 (unsequenced sites indicated by the letters N or X, respectively).

Table 1. Sequence coverage, expression, and divergence of Orco and the ten most highly expressed ligand-binding ORs in an antennal transcriptome of pooled males and females of the Japanese pine sawyer, *Monochamus alternatus*. Rows in bold indicate that the orthologs of these genes are also among the most highly expressed in antennae of the three North American species *M. maculosus*, *M. notatus*, and *M. scutellatus*.

Gene	Sequence Coverage	FPKM	Nonsynonymous Sites[a]	dN/dS (ω)[a]
Malt\Orco	100.0%	36.78	1	0.011
MaltOR18	76.8%	4.02	17	0.268
MaltOR101	**95.5%**	**3.90**	**31**	**0.281**
MaltOR29	**95.4%**	**2.76**	**11**	**0.137**
MaltOR42	**96.4%**	**2.37**	**41**	**0.205**
MaltOR103	90.7%	2.22	12	0.073
MaltOR15	**86.4%**	**1.76**	**34**	**0.485**
MaltOR59	**96.5%**	**1.57**	**9**	**0.268**
MaltOR85.1	80.1%	1.07	16	0.163
MaltOR86	82.2%	0.99	51	0.614
MaltOR13.3	84.8%	0.88	24	0.235

[a]Measured from alignments of orthologs from all four *Monochamus* spp.

Many of the most highly expressed ORs in *M. alternatus* overlapped with those of the three North American species, including our pheromone receptor candidates OR29 and 59 (Table 1). However, the remaining two pheromone receptor candidates in the North American species, OR33 and 56, were only modestly expressed in *M. alternatus*. Notably, OR37.4, 61, and 71 were highly expressed in North American species but fragmentary and recovered at low levels in *M. alternatus* (Online Resource 3). All the highly expressed ORs exhibited signatures of purifying selection ($\omega < 1$), but OR29, 59, and 103 were especially highly conserved across all four *Monochamus* spp. and presented very few nonsynonymous sites in >90% sequence coverage (Table 1). Unfortunately, the most highly conserved OR among the North American species, OR36, could only be recovered from 3 reads constituting 27% of its sequence length in *M. alternatus*, so further analysis in this species was not possible.

Discussion

We analyzed the phylogeny, total and differential expression, and divergence of ORs across four species of *Monochamus*, which have themselves diverged for over five million years yet remain attracted to the same pheromone, monochamol (Teale et al. 2011; Fierke et al. 2012; Ryall et al. 2015). Our results clearly highlight OR29 and OR59 as the most conserved and highly expressed lineages and therefore most likely to be sensitive to monochamol. OR29 is especially encouraging because it is orthologous to the confirmed pheromone receptor McarOR3 (Fig. 1). This receptor is sensitive to 2-methylbutan-1-ol, which is one of the several pheromone components produced by males of the hickory borer *M. caryae* (Mitchell et al. 2012). The genera *Megacyllene* and *Monochamus* descend from different cerambycid subfamilies (Cerambycinae and Lamiinae, respectively), hinting that OR29 may be a conserved pheromone

receptor throughout the Cerambycidae. Indeed, orthologs of OR29 have been documented from antennal transcriptomes of cerambycids in both subfamilies (e.g., *Aromia bungii* OR5, Wu et al. 2022; *Apriona germari* OR3, Qian et al. 2020), where they may be among the most highly expressed ORs (*Rhaphuma horsfieldi* OR29, Zhao et al. 2020). It is also auspicious that OR29 is among the very few lineages of ORs that are conserved across families of Coleoptera (Group 2B.ii; Mitchell et al. 2020). Though not universal in beetles, homologs of OR29 have been sequenced from some of the most ancient extant groups (viz. Cupedidae; Mitchell et al. 2020) and may be highly expressed relative to other ORs (e.g., RdomOR17; Oppert et al. 2022).

OR59 is also a compelling candidate due to its universally high expression and sequence conservation. This lineage is unrelated to any characterized PRs, but this could paradoxically offer an equal measure of support. Few beetle pheromone receptors characterized to date share an evolutionary history: the curculionid pheromone receptors hail from Group 7, but from separate sections of that OR subfamily, while cerambycid receptors are split across Groups 1 and 2B (Fig. 1). Functional inference of ORs via phylogeny is an uncertain task even when the clades of PRs are well-defined (e.g., Bengtsson et al. 2014), and the scattered emergence of pheromone receptors in beetles only complicates this further. This point is illustrated in our present study by OR36, a close paralog of the pheromone receptor McarOR5 (sensitive to 2-phenylethanol; Mitchell et al. 2012). The protein sequence of OR36 is identical across the three North American *Monochamus* spp., but it is expressed at low levels in the antennae (Fig. 2), and nearly absent from *M. alternatus* (Online Resource 3); this OR is unlikely to be a pheromone receptor in *Monochamus*, despite the phylogenetic support. Thus, a receptor for monochamol – a hydroxyether motif quite distinct from the pheromones bound by any characterized ORs – could

easily have been co-opted from a lineage such as OR59 that is not associated with pheromones of other species.

Whichever the case, our analysis was not solely predictive of pheromone receptors, and should also yield ORs sensitive to any key odors pertinent to both sexes of adult *Monochamus*. This would include odors associated with the host plants (in this case, conifers) necessary for adult maturation feeding and oviposition (Allison et al. 2004). Thus, while at least one of our candidate ORs may be sensitive to monochamol, we predict that many or all of our remaining candidates will instead detect volatiles produced by host plants, or kairomones produced by other wood-boring insects. Prominent targets for these ORs include ethanol and the cyclic monoterpene α-pinene, both of which are produced in abundance by softwoods and are powerful synergists for pheromone activity (Pajeres et al. 2010; Teale et al. 2011; Allison et al. 2012), as well as effective attractants in their own right (Miller 2006). Monoterpenoid pheromones produced by sympatric bark beetles, such as ipsdienol and ipsenol, are also broadly attractive across *Monochamus* spp. (Allison et al. 2001; Miller et al. 2013). Receptors for these compounds and other terpenoids have been characterized from the bark beetle *I. typographus* (Hou et al. 2021; Yuvaraj et al. 2021), but they are not clearly related to any cerambycid genes (Fig. 1). The sole characterized beetle receptor for α-pinene, also from a weevil (RferOR6; Ji et al. 2021), similarly lacked an expressed ortholog in *Monochamus* antennae (Fig. 1). Thus, sensitivity to these compounds must be mediated by other OR lineages in *Monochamus*, which is consistent with predictions of functional convergence in coleopteran ORs (Mitchell et al. 2020). However, one potential link to host volatiles in our study is a characterized receptor from the polyphagous scarab *H. parallela* (HparOR27), which is sensitive to volatiles produced by its numerous host plants (Wang et al. 2021). HparOR27 is related to the cerambycid OR54 (Fig. 1),

which was only modestly expressed in *Monochamus* (Fig. 2), but both are members of another OR lineage that is maintained across numerous beetle families (Group 1.ii; Mitchell et al. 2020).

Notably, half of the ORs in our analysis were not significantly upregulated in the antennal tip relative to the base. Much of this is an artifact of their very low overall expression (Fig. 2; Online Resource 2); nevertheless, some ORs were robustly expressed in the antennal base in the North American species, and this may indicate a non-olfactory function. OR101 and 103 were especially prominent in this regard because they were also highly conserved across *Monochamus* spp. These two genes are members of the small OR subfamily 5B, which has been found to be expressed throughout the body of larvae and adults (Engsontia et al. 2008; Dippel et al. 2016; Mitchell et al. 2017). Our results provide additional evidence that the 5B subfamily plays an important and conserved biological role, but one that may be functionally distinct from a typical olfactory receptor.

Overall, our data have provided a clear path forward for functional characterization of a novel and high value pheromone receptor. From a narrow perspective, a receptor for monochamol can be adapted into biosensors to survey for *Monochamus* (Bohbot and Vernick 2020), or it could be tested in heterologous systems to select more potent or more economically viable alternatives to the pheromone (Franco et al. 2018). More broadly, a confirmed pheromone receptor lineage in the Lamiini could open the door to pheromone identification throughout this economically important tribe. Pheromone data are entirely lacking for many pest species (e.g., *Batocera* spp.), while the hydroxyether pheromones of arguably the most prominent pest genus, *Anoplophora*, have seen relatively weak attraction in field settings (e.g., Nehme et al. 2014). Both of our candidate receptors included orthologs in *A. glabripennis*, and we predict a similar relationship across the tribe Lamiini. Thus, orthologs of a monochamol receptor would be

powerful tools for identifying or designing attractive blends for numerous species of cerambycid pests worldwide. Furthermore, if the OR29 lineage is indeed sensitive to monochamol in *Monochamus* and involved in pheromone detection in other longhorned beetles, then its many homologs in other beetle species (Mitchell et al 2020) may also be pheromone receptors. Such a lineage of conserved pheromone receptors would be a cornerstone for future comparative and functional studies of pheromone biology in the largest order of insects.

Acknowledgments

We thank Patrick Gorring for comments on an early draft of this manuscript and Nick Boyonoski for assistance collecting insects.

Declarations

Contributions

RFM, DD, and JDA designed the study; MCB and JDA oversaw field collection of samples; DD and SB performed molecular work and sequencing; RFM analyzed data and wrote the manuscript; all authors reviewed, edited, and approved the final manuscript.

Funding

This work was supported by the University of Wisconsin Oshkosh (to RFM), a Genomics Research and Development Initiative (GRDI) grant (to DD), and by the Canadian Forest Service (to DD and JDA).

Competing Interests

The authors have no competing interests to declare.

Data Availability Statement

The data that support the findings of this study are openly available in the NCBI Sequence Read Archive (SRA) at https://www.ncbi.nlm.nih.gov/sra, BioProject ID PRJNA821225 and in the supplementary material of this article.

References

Allison JD, Borden JH, McIntosh RL, De Groot P, Gries R (2001) Kairomonal response by four *Monochamus* species. (Coleoptera: Cerambycidae) to bark beetle pheromones. J Chem Ecol 27:633–646

Allison JD, Borden JH, Seybold SJ (2004) A review of the chemical ecology of the Cerambycidae (Coleoptera). Chemoecology 14:123–150

Allison JD, McKenney JL, Millar JG, McElfresh JS, Mitchell RF, Hanks LM (2012) Response of the woodborers *Monochamus carolinensis* and *Monochamus titillator* (Coleoptera: Cerambycidae) to known cerambycid pheromones in the presence and absence of the host plant volatile α-pinene. Environ Entomol 41:1587–1596

Anderson AR, Newcomb RD (2021) Olfactory genomics and biotechnology in insect control. In: Blomquist G, Vogt R (eds) Insect Pheromone Biochemistry and Molecular Biology, 2nd Ed. Academic Press, Cambridge, MA, USA, pp 645–674

Antony B, Johny J, Montagné N, Jacquin-Joly E, Capoduro R, Cali K, Persaud K, Al-Saleh MA, Pain A (2021) Pheromone receptor of the globally invasive quarantine pest of the palm tree, the red palm weevil (*Rhynchophorus ferrugineus*). Mol Ecol 30:2025–2039

Bengtsson JM, Gonzalez F, Cattaneo AM, Montagné N, Walker WB, Bengtsson M, Anfora G, Ignell R, Jacquin-Joly E, Witzgall P (2014) A predicted sex pheromone receptor of codling moth *Cydia pomonella* detects the plant volatile pear ester. Front Ecol Evol 2:33

Bohbot JD, Vernick S (2020) The emergence of insect odorant receptor-based biosensors. Biosensors 10:26

Cerezke HF, Volney WJA (1995) Forest insect pests in the Northwest region. In: Armstrong JA, Ives WGH (eds) Forest insect pests in Canada. Natural Resources Canada, Ottawa, ON, Canada, pp 59–72

Dyer LJ, Seabrook WD (1975) Sensilla on the antennal flagellum of the sawyer beetles *Monochamus notatus* (Drury) and *Monochamus scutellatus* (Say) (Coleoptera: Cerambycidae). J Morphol 146:513–531

Engsontia P, Sangket U, Chotigeat W, Satasook C (2014) Molecular evolution of the odorant and gustatory receptor genes in lepidopteran insects: implications for their adaptation and speciation. J Mol Evol 79:21–39

Fierke MK, Skabeikis DD, Millar JG, Teale SA, McElfresh, JS, Hanks LM (2012) Identification of a male-produced aggregation pheromone for *Monochamus scutellatus scutellatus* and an attractant for the congener *Monochamus notatus* (Coleoptera: Cerambycidae). J Econ Entomol 105:2029–2034

Fleischer J, Krieger J (2021) Molecular mechanisms of pheromone detection. In: Blomquist G, Vogt R (eds) Insect Pheromone Biochemistry and Molecular Biology, 2nd Ed. Academic Press, Cambridge, MA, USA, pp 355–413

Franco TA, Xu P, Brito NF, Oliveira DS, Wen X, Moreira MF, Unelius CR, Leal WS, Melo AC (2018) Reverse chemical ecology-based approach leading to the accidental discovery of repellents for *Rhodnius prolixus*, a vector of Chagas diseases refractory to DEET. Insect Biochem Mol Biol 103:46–52

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321

Gorring PS (2019) Gene to Genus: Systematics and Population Dynamics in Lamiini Beetles (Coleoptera: Cerambycidae) With Focus on *Monochamus* Dejean. Dissertation, Harvard University

Haack RA, Hérard F, Sun J, Turgeon JJ (2010) Managing invasive populations of Asian longhorned beetle and citrus longhorned beetle: a worldwide perspective. Annu Rev Entomol 55:521–546

Hansen L, Xu T, Wickham J, Chen Y, Hao D, Hanks LM, Millar JG, Teale SA (2015) Identification of a male-produced pheromone component of the citrus longhorned beetle, *Anoplophora chinensis*. PLoS ONE 10:e0134358

Hansson BS, Stensmyr MC (2011) Evolution of insect olfaction. Neuron 72:698–711

Hou XQ, Yuvaraj JK, Roberts RE, Zhang DD, Unelius CR, Löfstedt C, Andersson MN (2021) Functional evolution of a bark beetle odorant receptor clade detecting monoterpenoids of different ecological origins. Mol Biol Evol 38:4934–4947

Ji T, Xu Z, Jia Q, Wang G, Hou Y (2021) Non-palm Plant Volatile α-Pinene Is Detected by Antenna-Biased Expressed Odorant Receptor 6 in the *Rhynchophorus ferrugineus* (Olivier) (Coleoptera: Curculionidae). Front Physiol 12:701545

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2016) PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol Biol Evol 34:772–773

Leary GP, Allen JE, Bunger PL, Luginbill JB, Linn CE, Macallister IE, Kavanaugh MP, Wanner KW (2012) Single mutation to a sex pheromone receptor provides adaptive specificity between closely related moth species. Proc Natl Acad Sci USA 109:14081–14086

Lemoine F, Domelevo Entfellner JB, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, Gascuel O (2018) Renewing Felsenstein's phylogenetic bootstrap in the era of big data. Nature 556:452–456

Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550

Millar JG, Hanks LM (2017) Chemical ecology of cerambycids. In: Wang Q (ed) Cerambycidae of the world: biology and pest management. CRC Press/Taylor & Francis, Boca Raton, FL, USA, pp 161–208

Miller DR (2006) Ethanol and (−)-α-pinene: attractant kairomones for some large wood-boring beetles in southeastern USA. J Chem Ecol 32:779–794

Miller DR, Dodds KJ, Eglitis A, Fettig CJ, Hofstetter RW, Langor DW, Mayfield III AE, Munson AS, Poland TM, Raffa KF (2013) Trap lure blend of pine volatiles and bark beetle pheromones for *Monochamus* species. (Coleoptera: Cerambycidae) in pine forests of Canada and the United States. J Econ Entomol 106:1684–1692

McKenna DD, Scully ED, Pauchet Y, Hoover K, Kirsch R, Geib SM, Mitchell RF, Waterhouse RM, Ahn S-J, Arsala D et al (2016) 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 17:227

Mitchell RF, Hughes DT, Luetje CW, Millar JG, Soriano-Agatón F, Hanks LM, Robertson HM (2012) Sequencing and characterizing odorant receptors of the cerambycid beetle *Megacyllene caryae*. Insect Biochem Mol Biol 42:499–505

Mitchell RF, Reagel PF, Wong JC, Meier LR, Silva WD, Mongold-Diers J, Millar JG, Hanks LM (2015) Cerambycid beetle species with similar pheromones are segregated by phenology and minor pheromone components. J Chem Ecol 41:431–440

Mitchell RF, Hall LP, Reagel PF, McKenna DD, Baker TC, Hildebrand, JG (2017) Odorant receptors and antennal lobe morphology offer a new approach to understanding olfaction in the Asian longhorned beetle. J Comp Physiol A 203:99–109

Mitchell RF, Schneider TM, Schwartz AM, Andersson MN, McKenna DD (2020) The diversity and evolution of odorant receptors in beetles (Coleoptera). Insect Mol Biol 29:77–91

Mitchell RF, Andersson MN (2021) Olfactory genomics of the Coleoptera. In: Blomquist G, Vogt R (eds) Insect Pheromone Biochemistry and Molecular Biology, 2nd Ed. Academic Press, Cambridge, MA, USA, pp 547–590

Morimoto K, Iwasaki A (1972) Role of *Monochamus alternatus* (Coleoptera: Cerambycidae) as a vector of *Bursaphelenchus lignicolus* (Nematoda: Aphelenchoididae). J Jap For Soc 54:177–183

Montagné N, Wanner K, Jacquin-Joly E (2021) Olfactory genomics within the Lepidoptera. In: Blomquist G, Vogt R (eds) Insect Pheromone Biochemistry and Molecular Biology, 2nd Ed. Academic Press, Cambridge, MA, USA, pp 469–505

Nehme ME, Trotter RT, Keena MA, McFarland C, Coop J, Hull-Sanders HM, Meng P, De Moraes CM, Mescher MC, Hoover K (2014) Development and evaluation of a trapping system for *Anoplophora glabripennis* (Coleoptera: Cerambycidae) in the United States. Environ Entomol 43:1034–1044

Oppert B, Muszewska A, Steczkiewicz K, Šatović-Vukšić E, Plohl M, Fabrick JA, Vinokurov KS, Koloniuk I, Johnston JS, Smith TP et al (2022) The genome of *Rhyzopertha dominica* (Fab.) (Coleoptera: Bostrichidae): adaptation for success. Genes 13:446

Pajares JA, Álvarez G, Ibeas F, Gallego D, Hall DR, Farman DI (2010) Identification and field activity of a male-produced aggregation pheromone in the pine sawyer beetle, *Monochamus galloprovincialis*. J Chem Ecol 36:570–583

Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – Approximately maximum-likelihood trees for large alignments. PloS ONE 5:e9490

Qian JL, Mang DZ, Lv GC, Ye J, Li ZQ, Chu B, Sun L, Liu YJ, Zhang LW (2020) Identification and expression profile of olfactory receptor genes based on *Apriona germari* (Hope) antennal transcriptome. Front Physiol 11:807

Ray AM, Barbour JD, McElfresh JS, Moreira JA, Swift I, Wright IM, Žunič A, Mitchell RF, Graham EE, Alten RL, Millar JG (2012) 2, 3-Hexanediols as sex attractants and a female-produced sex pheromone for cerambycid beetles in the prionine genus *Tragosoma*. J Chem Ecol 38:1151–1158

Ray AM, Millar JG, Moreira JA, McElfresh JS, Mitchell RF, Barbour JD, Hanks LM (2015) North American species of cerambycid beetles in the genus *Neoclytus* share a common hydroxyhexanone-hexanediol pheromone structural motif. J Econ Entomol 108:1860–1868

Ryall K, Silk P, Webster RP, Gutowski JM, Meng Q, Li Y, Gao W, Fidgen J, Kimoto T, Scarr T, Mastro V (2015) Further evidence that monochamol is attractive to *Monochamus* (Coleoptera: Cerambycidae) species, with attraction synergised by host plant volatiles and bark beetle (Coleoptera: Curculionidae) pheromones. Can Entomol 147: 564–579

Solomon JD (1995) Guide to insect borers in North American broadleaf trees and shrubs (No. 706). USDA Forest Service, Washington, D.C., USA

Teale SA, Wickham JD, Zhang F, Su J, Chen Y, Xiao W, Hanks LM, Millar JG (2011) A male-produced aggregation pheromone of *Monochamus alternatus* (Coleoptera: Cerambycidae), a major vector of pine wood nematode. J Econ Entomol 104:1592–1598

Wang J, Li DZ, Min SF, Mi F, Zhou SS, Wang MQ (2014) Analysis of chemosensory gene families in the beetle *Monochamus alternatus* and its parasitoid *Dastarcus helophoroides*. Comp Biochem Physiol D 11:1–8

Wu Z, Ye J, Qian J, Purba ER, Zhang Q, Zhang L, Mang D (2022) Identification and expression profile of chemosensory receptor genes in *Aromia bungii* (Faldermann) antennal transcriptome. Insects 13:96

Xu B, Yang Z (2013) PAMLX: a graphical user interface for PAML. Mol Biol Evol 30:2723–2724

Wickham JD, Harrison RD, Lu W, Chen Y, Hanks LM, Millar JG (2021) Rapid assessment of cerambycid beetle biodiversity in a tropical rainforest in Yunnan Province, China, using a multicomponent pheromone lure. Insects 12:277

Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yuvaraj JK, Roberts RE, Sonntag Y, Hou XQ, Grosse-Wilde E, Machara A, Zhang DD, Hansson BS, Johanson U, Löfstedt C, Andersson MN (2021) Putative ligand binding sites of two functionally characterized bark beetle odorant receptors. BMC Biol 19:1–21

Zhang A, Oliver JE, Aldrich JR, Wang B, Mastro VC (2002) Stimulatory beetle volatiles for the Asian longhorned beetle, *Anoplophora glabripennis* (Motschulsky). Z Naturforsch C 57:553–558

Zhang R, Wang B, Grossi G, Falabella P, Liu Y, Yan S, Lu J, Xi J, Wang G (2017) Molecular basis of alarm pheromone detection in aphids. Curr Biol 27:55–61

Zhao S, Ye Z, Stanton R (2020) Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. RNA 26:903–909

Zhao YJ, Li GC, Zhu JY, Liu NY (2020) Genome-based analysis reveals a novel SNMP group of the Coleoptera and chemosensory receptors in *Rhaphuma horsfieldi*. Genomics 112:2713–2728

Zhao Y, Li MC, Konaté MM, Chen L, Das B, Karlovich C, Williams PM, Evrard YA, Doroshow JH, McShane LM (2021) TPM, FPKM, or normalized counts? A comparative study of quantification measures for the analysis of RNA-seq data from the NCI Patient-Derived Models Repository. J Transl Med 19:1–15

Electronic Supplementary Material Captions

Online Resource 1 (ESM_1.xslx) Names, description, and sequences of odorant receptor genes recovered from the antennal transcriptomes of *Monochamus maculosus*, *M. notatus*, and *M. scutellatus*

Online Resource 2 (ESM_2.xslx) Expression levels and differential expression of odorant receptors in transcriptomes of antennal tips and bases from *Monochamus maculosus*, *M. notatus*, and *M. scutellatus*. Species are separated into three tabs for clarity (Tables S2A, B, and C, respectively)

Online Resource 3 (ESM_3.xslx) Sequences and expression of odorant receptor genes recovered from a published antennal transcriptome of *Monochamus alternatus*