



OPEN

Identification and analysis of odorant receptors expressed in the two main olfactory organs, antennae and palps, of *Schistocerca americana*

Alejandra Boronat-Garcia¹, James Iben², Eunice Dominguez-Martin³ & Mark Stopfer¹✉

Locusts depend upon their sense of smell and provide useful models for understanding olfaction. Extending this understanding requires knowledge of the molecular and structural organization of the olfactory system. Odor sensing begins with olfactory receptor neurons (ORNs), which express odorant receptors (ORs). In insects, ORNs are housed, in varying numbers, in olfactory sensilla. Because the organization of ORs within sensilla affects their function, it is essential to identify the ORs they contain. Here, using RNA sequencing, we identified 179 putative ORs in the transcriptomes of the two main olfactory organs, antenna and palp, of the locust *Schistocerca americana*. Quantitative expression analysis showed most putative ORs (140) are expressed in antennae while only 31 are in the palps. Further, our analysis identified one OR detected only in the palps and seven ORs that are expressed differentially by sex. An in situ analysis of OR expression suggested ORs are organized in non-random combinations within antennal sensilla. A phylogenetic comparison of OR predicted protein sequences revealed homologous relationships among two other *Acrididae* species. Our results provide a foundation for understanding the organization of the first stage of the olfactory system in *S. americana*, a well-studied model for olfactory processing.

Olfaction allows animals to detect, identify, and discriminate among hundreds of thousands of odor molecules present in the environment. This ability is essential for animals' survival and yet requires a complex process to generate the high dimensional neural representations needed to characterize odorant molecules, which have different sizes, shapes and electrical charges, and are often organized into chaotic and turbulent odor plumes¹. Understanding the anatomical organization of the olfactory system at the cellular and molecular levels has provided important insights into the coding mechanisms underlying olfaction^{2–6}, and studies performed in insects have contributed substantially to our knowledge of odor processing^{3,7–10}. Further, mechanisms that allow the olfactory system to generate representations for odors have been shown to be widely conserved among very divergent species^{11,12}.

Odor molecules are detected by transmembrane proteins including odorant receptors (ORs) expressed in the dendrites of olfactory receptor neurons (ORNs). Insect ORs are members of the seven transmembrane domain (TMD) superfamily and are highly divergent^{13–16}. Normally, a single ORN expresses the co-receptor Orco together with only one type of OR which imbues the neuron with its tuning properties and its complex and diverse response dynamics^{15,17–19}. In insects, ORNs are housed in specialized hair-like structures called olfactory sensilla which are classified into three morphological types: basiconic, trichoid, and coeloconic^{20,21}. Although the principles underlying the organization of ORNs within sensilla are poorly understood, there is evidence to suggest this organization plays important roles in olfactory coding. Sensilla, for example, can be organized into different functional classes based on the combination of ORN types housed within them⁵, and olfactory ephaptic

¹Section on Sensory Coding and Neural Ensembles, National Institutes of Health, Eunice Kennedy Shriver National Institute of Child and Human Development, Bethesda, MD, USA. ²Molecular and Genomics Core, National Institutes of Health, Eunice Kennedy Shriver National Institute of Child and Human Development, Bethesda, MD, USA. ³Biochemistry Section, National Institutes of Health, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA. ✉email: stopferm@nih.gov

interactions appear to require the close apposition of functionally related ORNs²². Understanding ORN organization at the sensilla level will help clarify the earliest stages of odor coding.

The locust *Schistocerca americana* is a well-studied model of olfactory processing, and electrophysiological studies in this species have improved fundamental understanding of information processing by neurons^{11,18,23,24}. However, little is known of the organization of ORNs in this organism in part because the ORs expressed in its odorant sensing organs have not yet been identified. Identifying these ORs will, among other benefits, facilitate research into olfactory systems and neural codes for sensory stimuli.

The locust has two main odor sensing organs: the antennae and the mouthparts (maxillary and labial palps). While each antenna houses ~100,000 ORNs, each of the four palps contains only ~150^{25–29}. These two organs have been suggested to play different roles in odor processing, with the antennae likely serving as general purpose odor sensors, and the palps processing food odors^{27,30,31}. However, a better understanding of these structures and their functions will require identifying the ORs they express.

To extend our knowledge of OR genes expressed in *S. americana* (Same) we analyzed the transcriptomes of the antennae and the palps. Because female and male locusts are sexually dimorphic and show some different behaviors^{32,33}, we separately analyzed tissue from adult females and males. All together, we identified 179 transcripts of putative ORs (named here as SameORs) expressed in the two main olfactory organs. Of the 159 sequences encoding proteins predicted to have at least two transmembrane domains (TMDs), 140 were present at medium to high abundances in the antenna, but only 31 were found in the palps. We found one putative OR (SameOR63) expressed only in the palp, suggesting this receptor plays a palp-specific role. Also, we found some putative SameORs expressed differentially in female and male locusts. In the antenna: four putative ORs (SameOR1, SameOR7, SameOR25, and SameOR152) were significantly more highly expressed in male than female and one (SameOR40) more expressed in females. In the palp: two putative ORs (SameOR33 and SameOR29) were significantly more highly expressed in male than female tissue, suggesting sexual dimorphisms of OR expression. Also, an in situ analysis of OR expression suggested ORs are organized in non-random combinations within antennal trichoid sensilla. Finally, we compared the sequences we obtained for *S. americana* to those of two other related *Acrididae* species. Our results provide a necessary foundation for understanding the organization of the peripheral olfactory system of the locust.

Results

Transcriptome sequencing and assembly. From female and male antennae, a total of 101.8 and 107.3 million raw reads, respectively, were produced using Illumina sequencing platforms, and from female and male palps, a total of 646 and 622.8 million raw read pairs, respectively, were produced (Supplementary Table 1 shows the number of read pairs produced for each tissue type sample). GC content ranged between 44 and 46%. After assembling the transcriptome, we obtained a total of 850,000 sequences, with an N50 of 1383 bp, a mean length of 726 bp, and a mapping percentage of reads to the assembly of 83% (Supplementary Table 2 shows additional assembly statistics.) To assess the quality of the assembly (i.e., accuracy and completeness), we performed two standard tests. First, we calculated the TransRate assembly score, which is appropriate for transcriptomes assembled de novo³⁴. Our TransRate score was 0.85 indicating high accuracy and completeness according to established standards³⁴. Second, we calculated BUSCO, which checks the transcriptome for the presence of a group of single-copy orthologous genes thought to be essential for life and occurring in more than 90% of species³⁵. BUSCO results (Supplementary Fig. 1) showed that of the 255 genes evaluated, 249 (97.6%) were complete— from which 89 (34.9%) were complete and single-copy and 160 (62.7%) were complete but duplicate—4 (1.5%) were fragmented, and 2 (0.9%) were missing. Overall, these metrics indicate a high-quality assembly in terms of completeness³⁶. The raw reads are deposited in the NCBI SRA database (BioProject ID PRJNA889432).

Identification of sequences encoding putative odorant receptors. We identified a total of 179 transcripts encoding putative ORs in the antennae and/or palps (See Methods and Supplementary Data 1–2). The lengths of the predicted coding sequences (CDS) ranged from 300 to 1509 bp (Fig. 1a). The majority of the CDS (127, or 70.9% of the total) was longer than 900 bp, while 28 (15.6%) were 501–900 bp and 24 (13.4%) were 300–500 bp. Most of the 179 putative ORs (115, 64.2%) had predicted secondary structures with a topology containing six or seven TMDs, 19 (10.6%) had four or five predicted TMDs, 35 (19.5%) had two or three TMDs, and 10 (5.6%) contained one or no TMDs (Fig. 1b, Supplementary Data 1–2). A total of 107 (59.7%) of the 179 sequences had full-length CDS (Fig. 1c) with mean length 1267 bp. Of these 107 full-length sequences, almost all (101) had six or seven predicted TMDs.

We performed BLASTx searches to compare transcripts to an *S. americana* genome recently deposited in NCBI (see Methods). We found most of our sequences (173 of 179) produced a hit to those sequences annotated to be open reading frames (ORFs) in the assembled genome, with an identity ranging from 69.7 to 100% (Supplementary material Excel file S1). Most of our sequences (n = 131) matched against loci predicted to code for ORs in the genome (e.g., SameOrco, SameOR1, SameOR2, SameOR7, SameOR40, SameOR25) while 40 matched uncharacterized loci (e.g., SameOR3, SameOR60, SameOR71). In addition, six of our sequences did not match any locus (e.g., SameOR63, SameOR94, SameOR155, SameOR133, SameOR165, SameOR188) and two matched against non-OR loci (i.e., SameOR80 and SameOR87) (Fig. 1d and Supplementary Data 1–2).

To evaluate the success of our sequence assembly with an independent approach, we designed primers for some of our putative ORs and performed RT-PCR followed by sequencing to check for expression of the identified putative SameOR transcripts in antenna and palp tissues. Among the analyzed sequences, we included some with no matches to the NCBI *S. americana* genome (e.g., SameOR63), some that matched to uncharacterized loci in the genome but with short alignments (e.g., SameOR71 with only 81 bp aligned), and some that matched to a predicted OR loci (e.g., SameOrco). For example, for SameOR63, we amplified ~800 bp of the sequence and

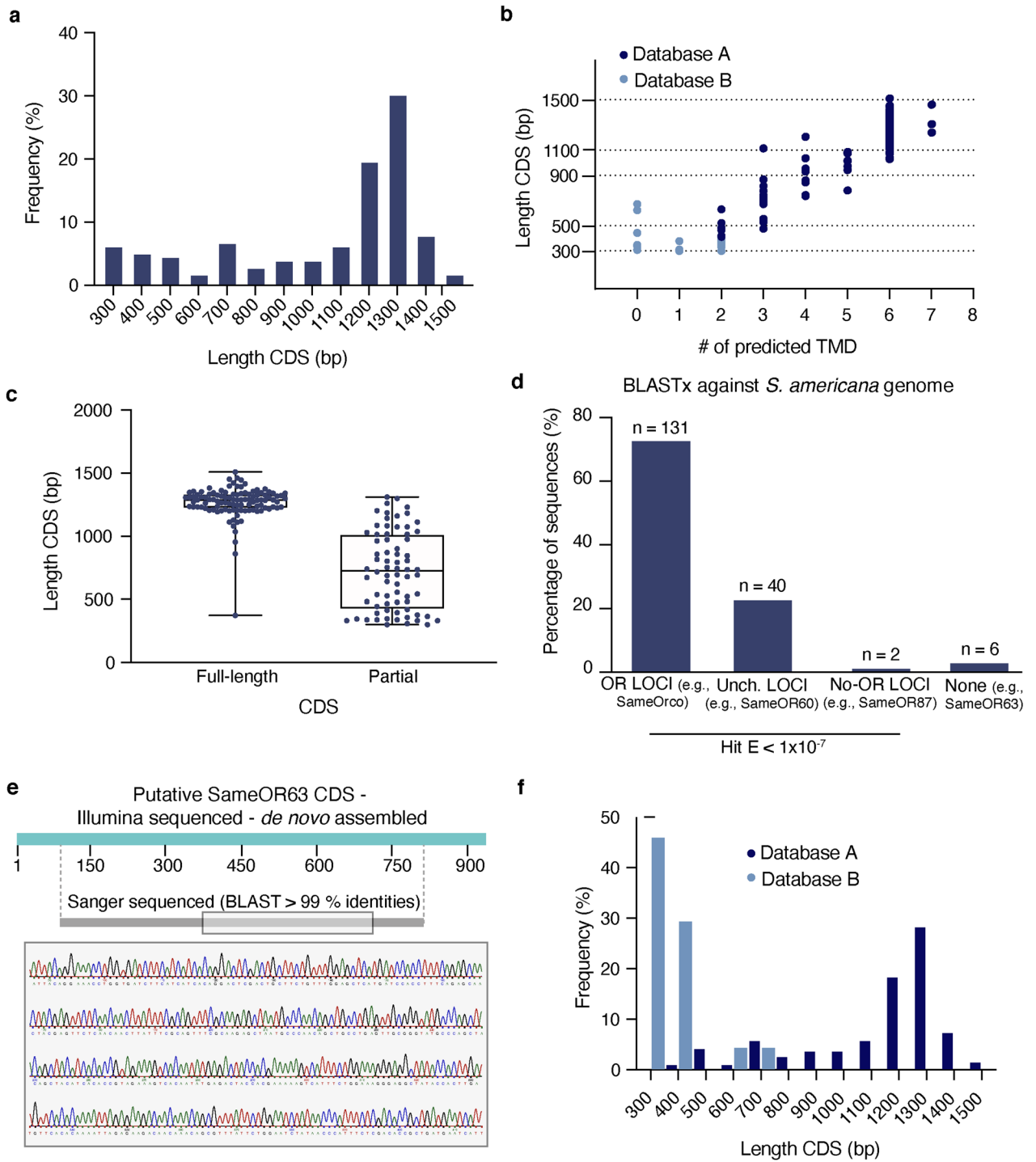


Figure 1. Statistical characterization of SameORs identified in the unified transcriptome assembly. (a) Frequency distribution of CDS length (bp). (b) Relation between CDS length and number of TMDs predicted from the deduced amino acid sequences of the 179 putative ORs. Putative OR sequences were divided into database A (dark blue) and database B (light blue) for further analysis. (c) Length of full and partial CDSs. (d) Blastx results of putative SameOR CDSs against a recently posted *S. americana* genome assembly. Shown are the percentage of sequences with a significant match ($E < 1E^{-7}$) to either a predicted OR, an uncharacterized (Unch.) or a none-OR loci, and those that did not match any genomic region (None). (e) Schematic highlighting of SameOR63 CDS region amplified by RT-PCR and sequenced using Sanger. The Sanger chromatogram at the bottom is an example of the sequencing results. (f) Frequency distributions of CDS length in Database A (dark blue) and Database B (light blue).

aligned it to our de novo assembled sequences (Fig. 1e). Our results confirmed the successful assembly of all the tested sequences and showed that all of them code for genes that are expressed in antennae and/or palp tissues (Fig. 1e, Supplementary Fig. 2).

Because ORs are predicted to have seven TMDs^{13–16}, we divided our collection of identified transcript sequences in two databases based upon the structural topology predicted from the derived amino acid sequences (Fig. 1b) and CDS quality (see Methods). The higher quality Database A contained 159 putative ORs with a predicted topology ranging from two to seven TMDs, including 115 sequences with six or seven TMDs. Most of Database A's transcripts (127, 79.9%) were longer than 900 bp, and of these, 105 were full-length sequences, while 26 (16.3%) were 501–900 bp, and only six (3.8%) were 400–500 bp long (Fig. 1f). Database B contained the remaining 20 putative ORs. Most of these were 300–500 bp (18, 90%) and only two (10%) were 501–700 bp (Fig. 1f). The number of predicted TMDs in these sequences ranged from zero to two (Fig. 1b). Because ORs are predicted to have seven TMDs, only sequences in Database A were used for further analyses including the quantification of transcript abundances, differential expression and spatial distribution analyses, and the evolutionary relationships among related species.

Tissue specific expression and sex differences of putative ORs in the antennae and palps. To determine whether the identified putative ORs were expressed differentially in antenna and palp, we quantified in both tissues the number of reads of the 159 putative ORs in Database A. Because the palp contains many fewer ORNs than the antenna, to increase sensitivity we performed additional reads from the palp (antenna: ~ 100 M reads; palp: ~ 650 M reads; see Methods). In the antenna tissue samples, 140 out of the 159 putative ORs were present in medium to high abundance (above 10 DESeq2 normalized counts on average per tissue, Fig. 2a black color labels left panel, Supplementary Figs. 3–4), one putative OR was not detected (Fig. 2a red labels left panel, Supplementary Figs. 3–4), and the rest were present at very low levels (0.1–10 DESeq2 normalized counts on average per tissue) (Fig. 2a orange and green labels left panel, Supplementary Figs. 3–4). Notably, palp tissue showed a very different expression pattern: only 31 out of 159 putative ORs were found from medium to high abundances, 56 were not detected, and 72 were present at very low levels (61 of them from 0.1 to 5 DESeq2 normalized counts on average per tissue) (Fig. 2a right panel, Supplementary Figs. 3 and 5). SameOR63 was detected only in the palps, suggesting this receptor plays a palp specific role.

The most abundant putative receptor in both tissue types and in all samples was SameOrco (Fig. 2a,b), consistent with Orco's role as a coreceptor in all ORNs expressing an OR¹⁷. The 10 most expressed putative ORs varied depending on the tissue type: on the antenna, SameORs 1–10 were the most expressed, from ~ 20 to 1% relative to Orco abundance (Fig. 2b, left panel). For the palp, SameORs 2, 4, 10, 21, 24, 32, 36, 49, 63, and 153 were the most present, from ~ 49 to 8% relative to Orco (Fig. 2b, right panel).

We also tested whether the 159 putative ORs in Database A are expressed differentially in tissues from female and male animals. In the antenna, we observed that only five putative ORs were differentially expressed based on sex (SameOR1, SameOR7, SameOR25, SameOR40, and SameOR152) (Fig. 2c, left panel). Of these five ORs, only the expression of SameOR40 was higher in females than in males (Fig. 2c, left panel). In the palps, two putative ORs (SameOR33 and SameOR29) were significantly more highly expressed in male than female palp tissue (Fig. 2c, right panel).

Sensilla-specific expression of putative ORs. In insects, ORNs are housed in olfactory sensilla. A subset of these neurons expresses ORs, all of which are accompanied by the OR co-receptor Orco. Normally, a single ORN expresses Orco together with only one type of OR^{15,17–19}. To characterize the expression patterns of putative OR sequences generated by our de novo assembly, we next used RNAscope to perform in situ hybridization in longitudinal sections of adult antenna. In two species closely related to *S. americana*, *S. gregaria* and *L. migratoria*, Orco is expressed only in ORNs in trichoid and basiconic sensilla³⁷. Here, as expected, RNAscope revealed highly abundant SameOrco expression in clusters of cells along the antenna (Fig. 3a,b). The expression patterns of SameOrco allowed us to identify trichoid and basiconic sensilla based on the number of SameOrco-expressing somata in each of the clusters: trichoid were identified as containing very few ORNs (from one to three ORNs) and basiconic, many more (Fig. 3a,b)²¹. Further, clusters of SameOrco⁺ cells were observed in sections of palp tissue but with fewer cells than in the antenna samples (Fig. 3c). These results establish that SameOrco is expressed in patterns expected for the co-receptor Orco.

In addition to characterizing SameOrco expression, we determined the spatial distributions of the second two most-expressed putative ORs in each of the tissues (Fig. 2). Using SameOrco to identify ORNs, we tested SameOR2 in the palp, and SameOR1 in the antenna. The results showed expression of these putative ORs exclusively in Orco⁺ cells in multiple sensilla in each analyzed section, and only in basiconic sensilla (Fig. 3d,e). Also, in some cases we found expression in multiple ORNs within the same cluster of Orco⁺ cells (Supplementary Fig. 6). These expression patterns are consistent with the initial transcript annotations as ORs.

Sensilla organization at the molecular level is thought to play important roles in olfactory coding⁵. Individual ORNs express only one OR type in addition to Orco¹⁹, however whether multiple ORNs housed in the same sensillum express the same or different OR types is unknown for *S. americana*. To start an analysis of this organization in the locust we focused on trichoid sensilla in the antenna because they contain relatively few ORNs, making the analysis tractable. We tested three putative ORs: SameOR58, SameOR71, and SameOR160. These candidates were selected because of their homologies to putative ORs previously described as present only in trichoid sensilla in the desert locust (SgreOR3–SameOR58 97% identity, SgreOR102–SameOR160 97% identity, and SgreOR111–SameOR71 98% identity)³⁸. We found that, as in the desert locust, these putative receptors were expressed in SameOrco⁺ ORNs in trichoid sensilla (Fig. 4a–c). However, in rare instances we

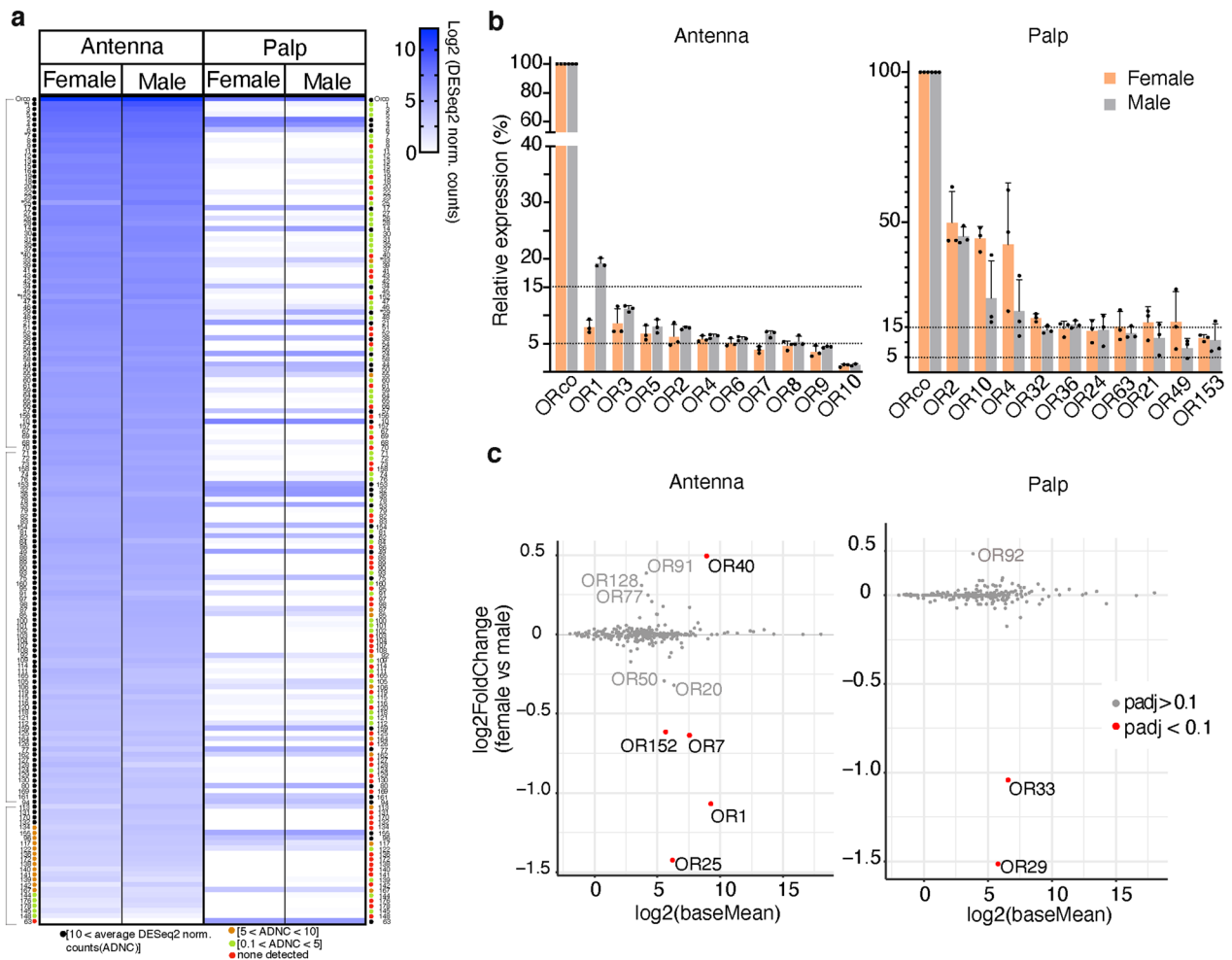


Figure 2. Expression of putative SameORs by tissue type and sex. RNAseq reads were aligned against identified putative SameORs and quantified. All data shown are based on DESeq2 normalized read counts. Quantification was performed in three replicates per condition. (a) Expression heat map of the 159 putative SameORs contained in Database A. Data are sorted according to antenna expression from highest to lowest normalized counts. Each row corresponds to the same putative SameOR in all columns. Colored circles next to each named SameOR indicate expression level: black, $10 < \text{average} < \sim 4300$; orange, $5 < \text{average} < 10$; green, $0 < \text{average} < 5$ normalized counts. Red: none detected. Light gray lines indicate grouped OR labels; see Supplementary Fig. 3 for an enlarged version. Asterisks: statistically significant differences in expression between female and male in antenna or palp tissues from panel (c). (b) Expression of the 10 most expressed putative SameORs in female (orange bars) and male (gray bars) tissue. Expression level is given relative to Orco. Error bars: standard deviation. (c) Differential expression between female and male in antennae (left panel) and palp (right panel) tissues. Red: differentially expressed putative SameORs $\text{padj} < 0.1$.

also found SameOR160 expressed in ORNs in basiconic sensilla (only two cases among hundreds of examined sensilla; Supplementary Fig. 7).

After confirming that these three SameORs were expressed in trichoid sensilla, we characterized their structural and molecular organization with an in situ RNAscope analysis of their expression patterns. Trichoid sensilla can contain one, two, or three ORNs. We used in situ RNAscope for SameOrco plus SameOR58, SameOR71, and SameOR160 to assess which combinations were present. Unlike the case of SameOR1 or SameOR2 that were found in one or more ORNs within the same basiconic sensillum, none of the trichoid sensilla we examined (of ~ 400 – 600 sensilla per condition across animals) contained more than one ORN expressing the same OR type (SameOR58, SameOR71, or SameOR160). In addition, we regularly observed the following combinations: sensilla housing one ORN expressed SameOR58, or SameOR71, or SameOR160; sensilla housing two ORNs expressed SameOR58 and SameOR71 or, SameOR58 and SameOR160; sensilla housing three ORNs expressed SameOR58, SameOR71, and SameOR160. However, we never observed the co-expression of SameOR160 and SameOR71 in a sensillum housing two ORNs. Further, we found sensilla containing three ORNs (Orco⁺ cells) that did not express any of the three tested ORs (Fig. 4b, red arrows), indicating that ORs we did not test with RNAscope are also expressed in trichoid sensilla. Together, these results suggest the organization of ORs within sensilla is not random and likely reflects an underlying logic.

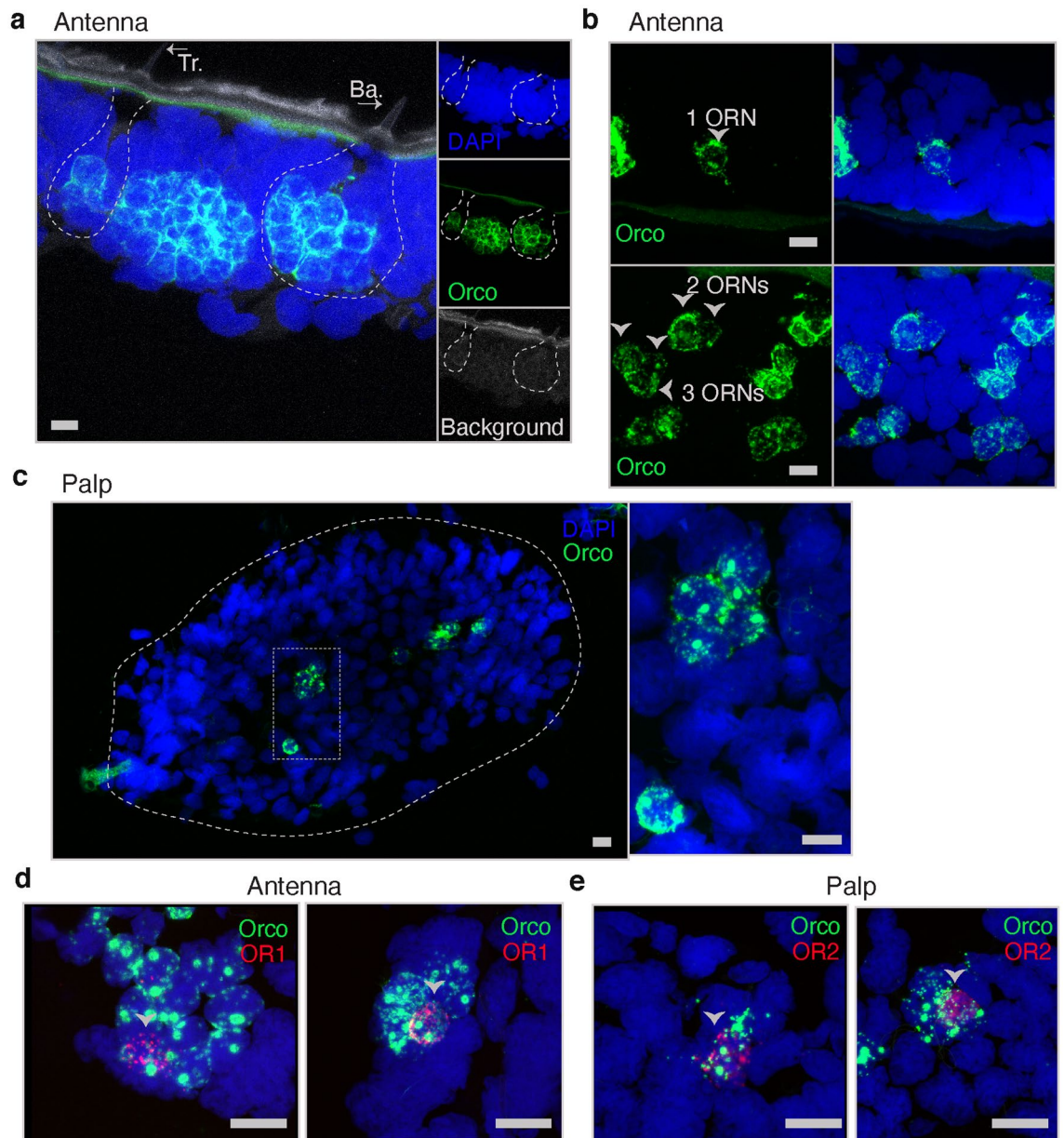


Figure 3. Expression of SameOrco, SameOR1, and SameOR2 in antenna and palp tissue sections, respectively. RNAscope in situ hybridization was used to characterize the spatial distribution of these putative ORs. (a–c) Probes against SameOrco (green) in antenna sections (a, b) and palp sections (c). Orco⁺ cells in trichoid (Tr) and basiconic (Ba) sensilla are highlighted by arrows in panel a. Trichoid sensilla containing one, two, and three ORNs are depicted by arrows in panel (b). (d, e) SameOrco probe with the probe against the second most expressed putative SameOR in each tissue type were used. SameOR1 (antenna, panel d) and SameOR2 (palp, panel e) are present only in Orco⁺ cells. Blue label in all images: DAPI. Scale bars: 10 μm in panels a–c; 15 μm in panels d–e.

Phylogenetic comparison of three orthopteran putative ORs. To further characterize and identify homologous sequences of our set of putative SameORs transcripts in other orthopteran, we performed a phylogenetic analysis. Using the 159 amino acid sequences deduced from the sequences in Database A, we generated a maximum likelihood-based phylogenetic tree by aligning them against previously reported OR sequences from two other members of the *Acrididae* family (Orthoptera order): *L. migratoria* and *S. gregaria*^{39,40}. The phylogenetic tree was rooted in the highly conserved Orco protein (Fig. 5a) and included Orco proteins from the two orthopteran and insects from different orders: *D. melanogaster* (Dme), *A. gambiae* (Agam), *M. sexta* (Mse), *S. gregaria* (Sgre) and *L. migratoria* (Lmig). Putative SameOrco exhibited the characteristic sequence conservation of Orco proteins as it shared above 60% sequence identity with Orco orthologs spanning the different insect orders⁴¹. The highest similarity of putative SameOrco was with those from the two orthopteran Lmig and Sgre—96.63 and 99.11% identity respectively; while it shared ~60–61% identity with the Dme, Agam and Mse

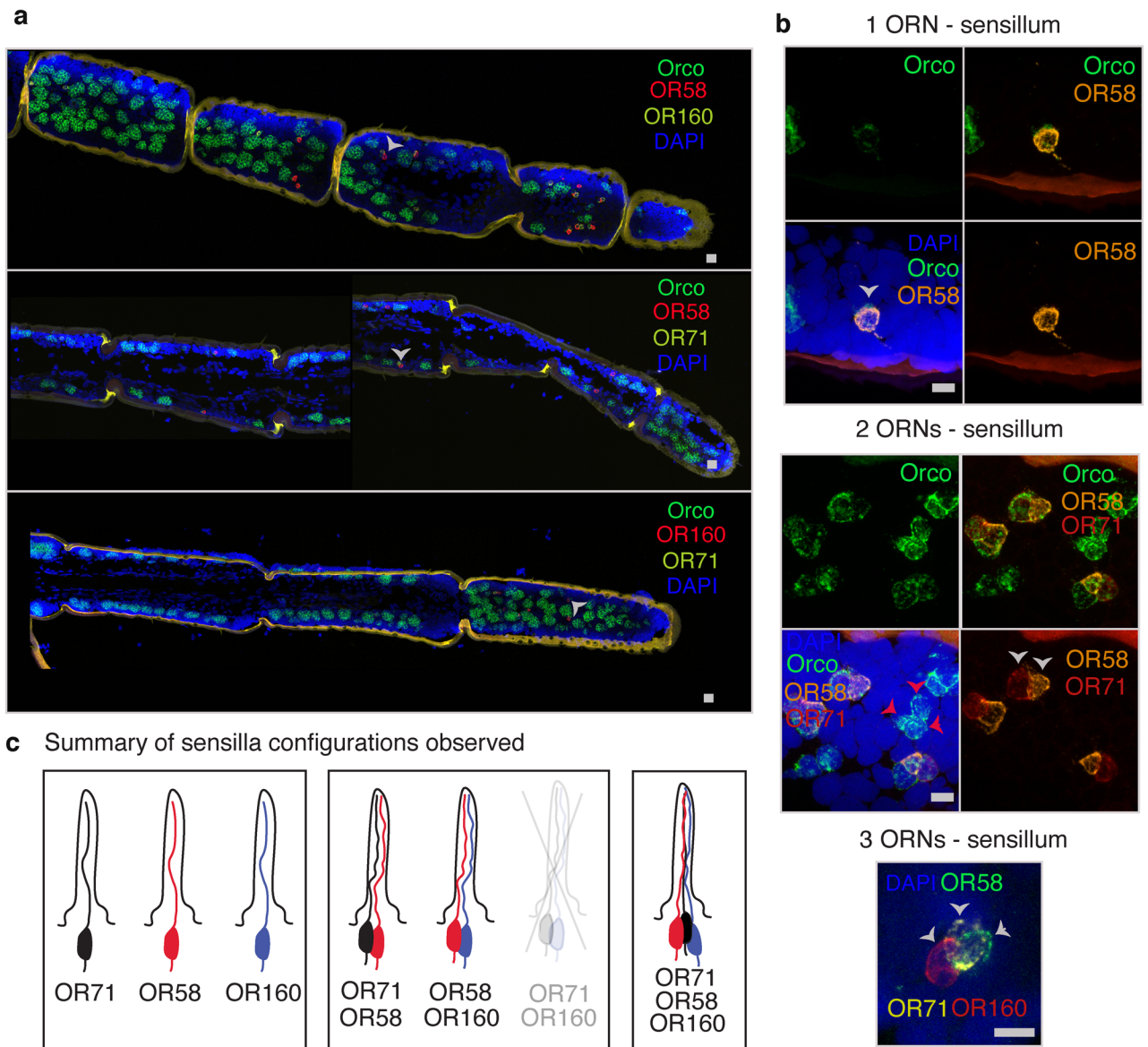


Figure 4. Characterization of SameOR71, SameOR58, and SameOR160 expression in trichoid sensilla. RNAscope in situ hybridization was used to determine the spatial distribution of these putative ORs. (a) Examples of antenna sections showing different combinations of probes against SameOrco and SameOR58 + SameOR160 (top), SameOR58 + SameOR71 (middle), and SameOR160 + SameOR71 (bottom), respectively. (b) Examples of individual trichoid sensilla containing one, two, or three ORNs and different combinations of the SameORs tested and indicated by gray arrows. Note that the images showing one and two ORNs are the same as in Fig. 3b but with additional color channels. Middle panel: red arrows: a cluster of three Orco⁺ cells that do not express either of the two tested putative SameORs. (c) Summary of the configurations observed in trichoid sensilla. Scale bars: 20 and 10 μ m in panel a and b, respectively.

Orcos. These results support our identification of SameOrco as encoding the Orco protein. Further, we found homologous sequences for almost all putative SameORs in either of the two-orthopteran species' ORs included in our analysis (Fig. 5b). Putative ORs in *S. americana* and *S. gregaria* were mapped more closely together in our phylogenetic analysis than those in *L. migratoria*, as shown in Fig. 5—most of them sharing >80% sequence identity—consistent with their closer evolutionary relationship. Together, these results provide additional evidence that the transcripts we identified as putative SameORs represent genuine ORs sequences.

Previous phylogenetic analysis using of either LmigORs, or LmigORs against SgreORs^{39,40} have shown clusters of ORs forming multiple families including two major groups with larger expansions. Following a similar methodological approach, we observed that the putative SameORs together with Sgre and Lmig ORs form similar phylogenetical clusters supported by high confidence values (>80), as previously observed^{39,40}. Based on the first seven nodes present in the tree, we divided the ORs into eight groups. Groups 1–6 radiated relatively early from the Orco-family and contain 11 putative SameORs. The small number of ORs present in Groups

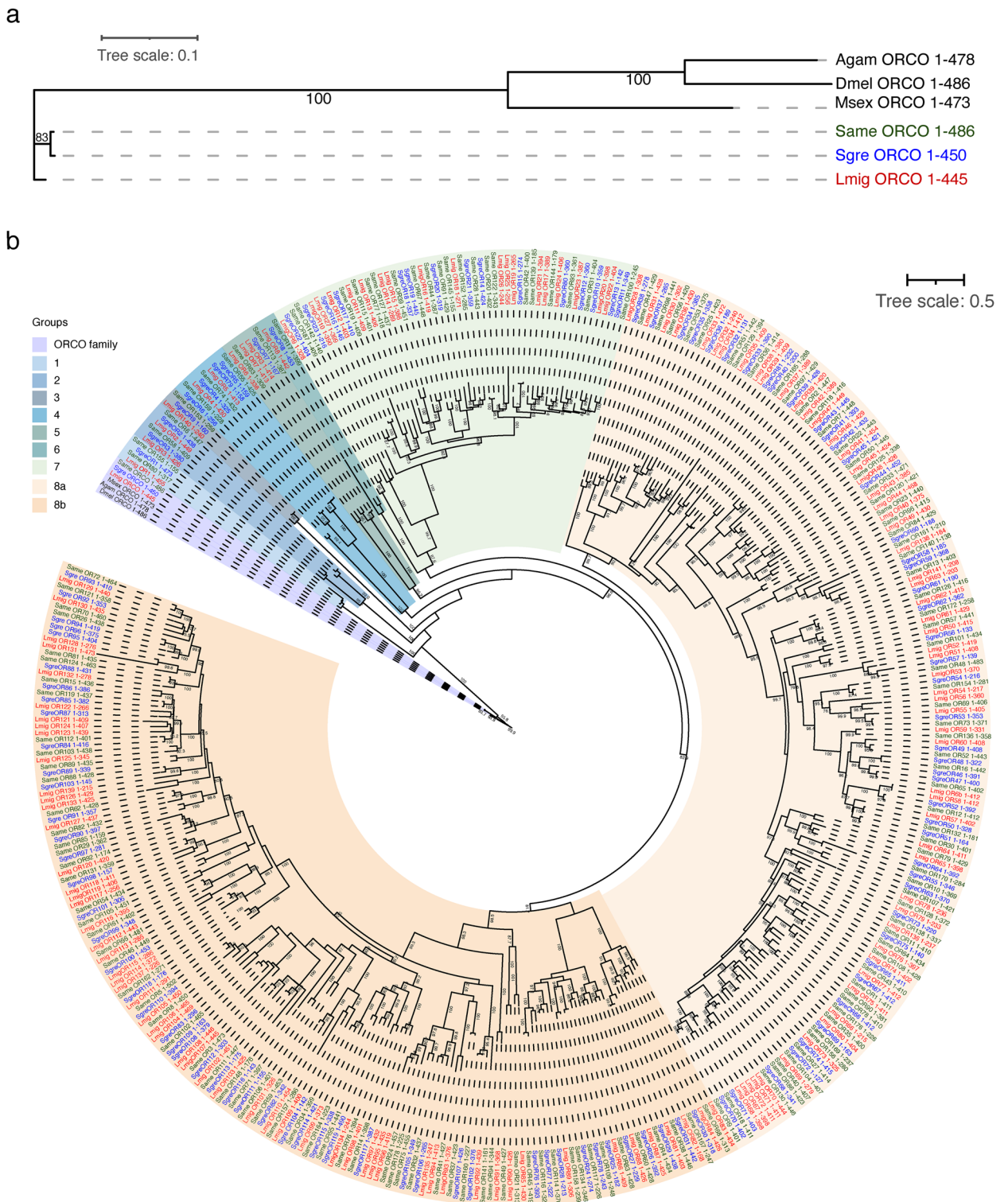


Figure 5. OR protein family phylogeny including putative SameORs and two other members of the *Acaridae* family (*S. gregaria*—Sgre and *L. migratoria*—Lmig). The maximum likelihood tree was generated from amino acid sequences deduced from the 159 putative ORs contained in Database A, and published ORs from *S. gregaria* and *L. migratoria* (see Methods). **(a)** Enlarged view of the co-receptor Orco subgroup shown in panel **(b)**. Confidence bootstrapping values are indicated. **(b)** The tree was rooted using Orco protein family members, including *D. melanogaster* (Dmel), *A. gambiae* (Agam), *M. sexta* (Msex), Sgre, Lmig, and putative SameOrco (shown in panel **a**) protein's amino acid sequences. Confidence bootstrapping values above 80 are shown (1000 iterations and a correlation coefficient of 0.99 were used). For descriptive purposes the ORs were divided into eight groups based on the first seven nodes. The length of each predicted protein's amino acid sequence is included in the labels.

1–6 represent relatively little evolutionary expansion with only one or two *Acrididae* OR members per clade. The remaining OR sequences were mapped into Groups 7 and 8, more distant from the Orco-family groups, as proposed previously^{39,40}. Group 7 contains 21 putative SameORs, while group 8 contains 126, with 63 and 63 in each of its two subgroups (8a and 8b respectively). These results support earlier observations that the *Acrididae* OR superfamily expanded from a common ancestor, which, after some possible initial duplications, gave rise to a set of groups with different extents of expansion. Our phylogenetic results, together with the previous reports from *L. migratoria* and *S. gregaria*, help characterize the evolution and conservation of the *Acrididae* OR protein superfamily. However, further genome sequencing of organisms within this family, together with complementary genetic and functional studies, will shed light on the details of its evolutionary history and functional relationships.

Discussion

In insects, olfaction begins when odorants bind with odor-sensitive proteins including ORs, ionotropic receptors (Irs), and some gustatory receptors⁴². Here, we identified a total of 179 transcripts of putative ORs in the antenna and palps, the two main olfactory organs of *S. americana*, and further characterized 159 putative ORs in terms of their differential expression patterns in the antennae and palps of female and male locusts. We also examined their phylogenetic relationships with closely related species, and homologies between SameORs to ORs of *S. gregaria* and *L. migratoria*. We also determined the spatial locations of a subset of these putative ORs within different types of sensilla, and, based upon their expression patterns, obtained evidence suggesting the organization of ORs within sensilla is not random but rather reflects an underlying logic. We also identified a single putative OR (SameOR63) expressed in the palps but not the antennae. Together, these results provide a new and precise view of OR expression in the locust, and useful tools for further exploration of an animal that is both a valuable model system for studies of sensory processing, and of economic interest since it can represent an agricultural pest.

Assembly and identification of putative SameORs. We used a de novo method using the Trinity platform to assemble the transcriptome of antennal and palp tissues (see “Methods”). De novo methods to assemble RNAseq data do not require the use of a genome for reference and are widely used^{36,43}. When it became available during our work, we then used the recently posted genome assembly of *S. americana* (GenBank GCA_021461395.2) as an additional tool to validate our approach and to help annotate the putative ORs we had identified.

Our high quality sequencing and assembly allowed us to identify 179 putative ORs in *S. americana*, more than had been reported in other orthopterans including the 142 and 119 putative ORs previously identified in *L. migratoria*³⁹ and *S. gregaria*⁴⁰, respectively. Our search for ORs included data pooled from antenna and palp transcriptomes to create a unified assembly, while the other studies combined antenna transcriptomes with locust genome data³⁹, or used the antenna transcriptome only⁴⁰. The latter approach would exclude putative ORs that are palp specific—for example, SameOR63. Our study also used relatively deep sequencing to identify potentially low-abundance transcripts; we generated 1.4724×10^9 paired end reads, compared to 9.5674882×10^7 reads generated in *L. migratoria*³⁹ and 5.1151235×10^7 paired end reads generated in *S. gregaria*⁴⁰. Our annotation process also benefited from comparisons to published putative OR sequences from *L. migratoria* and *S. gregaria* together with putative ORs from other members of *Acrididae* family (see Methods).

Insect ORs, other than Orco, are highly divergent with little sequence homology within insect species and, in some cases, even across species of the same order⁴⁴, complicating the process of identifying and annotating candidate OR-encoding sequences. Our sequencing and assembly methods revealed 107 sequences had full-length coding regions, equivalent to 59.7% of the full set of 179, more than previously reported for related species: 54% (77 putative ORs) and 15% (18 putative ORs) in *migratoria*³⁹ and *gregaria*⁴⁰, respectively. Moreover, the majority of the protein sequences of the putative OR CDSs we identified (64.2%) had predicted structural topologies containing at least six TMDs, as expected for OR proteins. It is possible that shorter sequences in our dataset with incomplete CDSs are poorly seeded assembly fragments (*i.e.*, noise) derived from the assembler, pseudogenes, or even pseudo-pseudo genes as is the case for some OR genes in *Drosophila melanogaster*⁴⁵. We divided our full set of putative SameORs into two databases. Database A includes sequences most likely to encode genuine SameOR proteins: those with multiple predicted TMDs^{13–16} and high homologies with those of closely related species, as revealed by our phylogenetic analysis. We limited our expression and phylogenetic analyses to Database A. Database B includes shorter sequences—some possibly sharing identities with putative SameORs contained in Database A—and some lacking multiple predicted TMDs. We provide Database B because future analyses may reveal it also contains some genuine ORs. Functional studies are now needed to determine whether these sequences encode fully functional ORs.

Tissue specific expression of putative SameORs. *S. americana*, like other insects, has two main olfactory organs: the antennae and the palps. We found these two organs expressed different but mainly overlapping sets of putative ORs. First, the majority (141 of 159) of the putative ORs were present in the antenna while less than a fifth of that number (only 31) were observed in the palp in moderate to high abundance. This difference is not surprising given that the palps contain three orders of magnitude fewer ORNs than the antenna (~200 vs ~100,000^{25–29}) and each ORN expresses only one type of OR together with Orco. Similar organ-specific differences have been observed in other insects. For example, in *L. migratoria*, of 149 putative OR genes identified, only 11 have been found in the palps⁴⁶; in *A. gambiae*, of ~80 putative ORs, only 13 are in the palps⁴⁷; in *D. melanogaster*, of 62 ORs, only seven are present in the palps⁴⁸; in *M. sexta*, of 70 OR genes tested, 17 are found in the mouthparts⁴⁹.

Second, we found that all but one of the putative ORs identified in the palps were also observed in the antenna. This overlap is different from the case of *Drosophila*, in which ORs expressed in antenna and palp are mutually exclusive⁴⁸. However, observations from other species including the dipteran *A. gambiæ*⁴⁷, the lepidopteran *M. sexta*⁴⁹, the orthopterans *L. migratoria*⁴⁶ and *S. gregaria*⁵⁰, and our results, show substantial overlap of OR types expressed in both structures, suggesting that OR mutually exclusive expression of ORs in antenna and palp tissues of *Drosophila* appears to be exceptional.

This partitioning of OR types to distinct olfactory structures may in some ways be comparable to the mammalian olfactory system, in which the molecular identities of ORNs in the main olfactory epithelium and the vomeronasal organ also overlap to some degree⁵¹. In mammals, ORNs from these two peripheral olfactory areas project to different target regions in the brain: ORNs from the main olfactory epithelium project into the main olfactory bulb, while the axons from vomeronasal ORNs expressing the same OR type project to the accessory olfactory bulb⁵¹. Similarly, in locust, antennal ORNs project to the antennal lobe, but palp ORNs project to the lobus glomerulatus²⁹. Functional studies are needed to determine whether odor information traveling through these pathways is processed differently or plays distinct roles.

Only one of the putative ORs we identified—SameOR63—was expressed exclusively in the palps, raising the possibility that it plays a specialized role in feeding. A homologue for SameOR63 was not identified in *S. gregaria*, but in *L. migratoria*, LmigOR6³⁹—also known as LmigOR12⁴⁶—shares 76.3% identity with SameOR63. A previous report showed that LmigOR12 was also found in the palps but not in the antennae of adult animals⁴⁶. Prior electrophysiological studies in *L. migratoria* have shown that certain odorants (e.g., (E,E)-2,4-heptadienal, hexanal and E-2-hexenal) elicited stronger responses in the palps than in the antennae, and that these responses were diminished when LmigOR12 expression was knocked down via RNAi⁴⁶. Several other putative ORs we identified in *S. americana*, in addition to SameOR63, are differentially expressed in the antennae and palps. These results at the molecular level between these two olfactory organs may reflect differences in their functions.

Sex differences in putative SameOR expression. In holometabolous insects (those undergoing complete metamorphosis during development) it is common to observe sexual dimorphic expression of some ORs. In Diptera^{47,52} and Lepidoptera^{53–55}, for example, receptors that detect pheromones are most often expressed differentially in females and males. However, less is known about pheromone detection or differences in OR expression based on sex in hemimetabolous insects, including orthopterans like the locust. Furthermore, there are no obvious differences between female and male antennal sensilla arrays in orthopterans *S. gregaria*²¹, *L. migratoria*⁵⁶, or *S. americana*⁵⁷, suggesting that, if olfactory sexual dimorphism exists, it would be present only at the molecular level. In agreement with this hypothesis, our results showed that, in *S. americana*, at least five putative ORs in the antenna and two in the palps were differentially expressed by sex. A previous report in *S. gregaria*⁵⁰ did not find evidence for sexual differences in the expression of putative ORs in either antennae or palps. However, the evidence provided for *S. gregaria* may not be sufficient to rule out the possibility of sex differences in the expression of OR proteins, since only nine putative ORs were analyzed in the study⁵⁰. Thus, the apparent discrepancy between our results and those from *S. gregaria* could be explained because none of the putative ORs we found to be differentially expressed were analyzed in *S. gregaria*⁵⁰. Whether these putative ORs detect pheromones remains to be tested.

Further, some pheromone receptors have already been identified in locust^{40,50,58}. For example, in *L. migratoria*, the aggregation pheromone 4-vinylanisole (4VA) was shown to be detected by LmigOR35⁵⁸; knocking down this receptor impaired behavioral responses normally elicited by this odorant. In our database, SameOR36, which is expressed on antennae and palps, comes closest to matching LmigOR35 (75% identity) and would be a good candidate for future analysis.

Organization of SameORs in sensilla. ORNs are housed in olfactory sensilla. Three types of sensilla, basiconic, trichoid and coeloconic, are found in the antenna, but only basiconic sensilla have been observed in the palps^{25,26}. In locust, basiconic and trichoid sensilla contain ORNs that co-express ORs with Orco. In *Drosophila* as well as some other species, sensillar organization in the antenna and the palps at the molecular and functional levels has been shown to be stereotypical⁵. However, in the locust this organization is unknown. The molecular organization of sensilla likely has functional consequences because ephaptic communication—nonsynaptic neural interactions—among ORNs closely apposed within a sensillum can influence their mutual responses to odors²². Therefore, it is important to understand sensillar organization as a contributor to early stages of odor coding.

To make our initial analysis of the organization of ORs into sensilla tractable, we focused mainly on trichoid sensilla, which each house only one to three ORNs. We identified trichoid sensilla as those containing small numbers of SameOrco expressing ORNs, and then analyzed them based on the patterns of putative ORs expressed by their ORNs. Our analysis confirmed the expression of at least three putative OR types in these sensilla. These results are consistent with a study performed in the trichoid sensilla of *S. gregaria* that identified the expression of ORs homologous to the three we tested⁴⁰, and to another study performed in *L. migratoria* that identified the expression of an OR homologous to SameOR58⁵⁹, suggesting the conservation of the spatial organization of these ORs as orthopterans evolved. Further, we observed that some combinations of these ORs were often expressed together in a sensillum, but other combinations were never expressed together. Our sample of three ORs is too small to draw specific conclusions, but we broadly interpret these results to suggest the organization of ORs within sensilla is not random and likely reflects an underlying logic²². An exciting direction for future work will be to explore this logic by examining more ORs and their functions.

Finally, we found evidence that additional combinations of ORs within trichoid sensilla are present in the antenna. For example, we observed cases in which none of the three Orco⁺ neurons housed in a trichoid sensillum

expressed any of the putative ORs tested. Each type of OR appears to be expressed in only one ORN per trichoid sensillum. Because we tested at least two different putative ORs in each experiment, it is likely that at least two other OR types that were not tested by RNAscope in this study are also expressed in trichoid sensilla. In addition, our observation that at least one of the putative ORs present in trichoid sensilla is also expressed in basiconic sensilla points to the possibility that additional OR types could be present in trichoid sensilla. However, the co-expression of this OR in both trichoid and basiconic sensilla was unusual because we observed it only once. In *L. migratoria*, 16 functional classes of trichoid sensilla have been described⁶⁰ which could be achieved by the presence of ~ five OR types.

Basiconic sensilla outnumber others on the locust antennae (<2000, compared to ~ 500 trichoid sensilla), each contain many ORNs (20–50, compared to one to three in trichoid sensilla), and are the only sensillar type found in the palps^{21,26–28,57}. Thus, one might expect the great majority of ORs to be expressed within basiconic sensilla. It will be important to explore the functional organization of basiconic sensilla in future studies.

Methods

Animal rearing. *Schistocerca americana* used in this study were raised in our large, crowded colony, in screened cages (45 × 45 × 45 cm) at a density of 300–400 per cage. The colony was maintained in a 12 h light 12 h dark cycle at ~ 30 °C and fed fresh wheat grass and oat bran.

RNA extraction and sequencing. We performed RNA sequencing to compare the molecular expression profiles of antenna and palp tissues in male and female adult *S. americana*. Three biological replicates were analyzed per condition, each consisting of left and right antennae or a mix of maxillary and labial palps from 5 adult females or males. Samples were collected and stored in RNAlater reagent (Qiagen) to prevent degradation, and then total RNA was extracted from the samples using Rneasy plus mini kit (Qiagen) and gDNA eliminator columns in accordance with the manufacturer's protocol. RNA samples were then assessed for concentration, purity, and integrity using Nanodrop and Bioanalyzer. Only samples with high purity and minimal degradation (absorbance ratio at 260/280 of ~ 2.0 and RNA integrity number—RIN—above 8) were used for subsequent steps. cDNA library preparation and sequencing were performed in the Molecular Genomics Core (NICHD-NIH, Bethesda MD). To prepare libraries, 1 to 3 µg of RNA was used and mature poly-A transcripts were enriched using a TruSeq RNA library preparation kit (Illumina). Subsequently, samples were sequenced using Illumina HiSeq 2500 and NovaSeq 6000 platforms (100 bp paired end reads). A total of 1,472.4 million clean (*i.e.*, post-trimmed) read pairs were obtained from all 12 tissue samples with a mapped ratio, on average, of 83% (Supplementary Table 1).

Assembly of RNA sequences. Before assembly, data were cleaned and trimmed using cutadapt software v2.5 (–a AGATCGGAAGAGCACACGTCTGAACTCCAGTCA –A AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT—overlap 6 –q 20—minimum-length 25). Sequences generated were assembled in Trinity v2.12.0 using a de novo assembly pipeline as previously described⁶¹. Briefly, reads from both tissue types were pooled and passed into the processing pipeline to create a unified de novo transcriptome assembly with software default parameters using 32 CPUs and 350 Gb of memory. This process generated a total of 850,291 contigs of which 60,863 contained ORFs (Supplementary Table 2). The completeness of the assembly was assessed with the TransRate assembly score and the benchmarking sets of universal single-copy orthologs (BUSCO) v3.0.2 completeness assessment tool (Supplementary Table 2, Supplementary Fig. 1).

OR annotation. To identify putative OR sequences, contigs were analyzed with BLASTn searches using Blast v2.13.0 software against a custom database containing annotated putative ORs from several related species, *Schistocerca gregaria*, *Locusta migratoria*, *Ceracris nigricornis*, and *Oedaleus infernalis* (Supplementary Fasta 1) and all labeled odorant receptors from the National Center for Biotechnology Information (NCBI) nucleotide database (Supplementary Fasta 2). From the extracted transcripts, we predicted all possible ORFs using TransDecoder software and analyzed with protein BLASTp searches against the NCBI and UniProtKB/Swiss-Prot databases. Only those sequences producing *e*-values below 1E⁻⁷ were defined as hits. This procedure allowed us to identify chimeric assembly transcripts and extract regions of the transcript that aligned to OR sequences, and to confirm that the predicted coding sequences (CDSs) matched to only ORs and not to a non-OR protein. Predicted CDSs aligning only to an OR were extracted, and those > 300 bp in length were defined as putative SameOR sequences.

To help validate our approach, we performed a BLASTx search of our putative OR sequences against a genome assembly for *S. americana* recently deposited in NCBI (GenBank GCA_021461395.2) and identified their loci (Supplementary Data 1). Only those sequences producing *e*-value scores below 1E⁻⁷ were considered hits.

To increase the likelihood that the extracted sequences encoded ORs, we applied two additional analyses. First, because ORs are expected to have multiple TMDs^{13–16}, we used deepTMHMM software⁶² to check for them in the amino acid sequences predicted from the CDSs (Supplementary Fasta 3 for CDSs and 4 for the predicted protein sequences). Second, we aligned the putative SameOR derived amino acid sequences using the clustal Omega Multiple Sequence Alignment web tool (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) and checked for predicted amino acid sequence lengths and redundancy. Based on these results we collected those non-redundant (*i.e.*, a shared identity above 90%) transcripts with two or more TMDs and lengths above 400 bp (~ 133 amino acid sequence length) in Database A (Supplementary Data 1), and the rest of the putative ORs in Database B (Supplementary Data 2). In addition, we determined the completeness of the CDSs by checking for the presence of start and stop codons using TransDecoder (Supplemental Data 1 and 2).

Reverse transcription polymerase chain reaction (RT-PCR). To further validate our putative ORs, we checked for their presence in the antennae and palps of female and male locusts. Briefly, total RNA was extracted from antennal and palp tissue following the protocol described above. Using Oligo-dT primers and the GoScript Reverse Transcription System (Promega) the first strand of cDNA was synthesized from 1 µg of RNA in a reaction volume of 20 µl. Synthesis was conducted at 42 °C for 60 min and was followed by 15 min incubation at 70 °C. Gene-specific primers were designed with Primer-BLAST (Supplementary Table 3) spanning an exon junction when possible. In addition, a negative control condition without reverse transcription for each sample was processed for Orco amplification which allowed us to rule out the possibility of genomic DNA contamination. PCRs were performed using GoTaq Green Master Mix (Promega). For all samples, 1 µl of cDNA was used as template in a reaction volume of 25 µl with the following conditions: 95 °C for 2 min, followed by 35 cycles with 95 °C for 30 s, 65 °C for 40 s, and 72 °C for 20 s, and, after the last cycle, a final incubation at 72 °C for 5 min. All PCR products in a given experiment were run in the same gel (2% agarose) and visualized by staining with SYBR green. Images of the gels were acquired with the SmartDoc Imaging System for smartphones and processed in Image J (i.e., 8-bit image conversion, rotation, and LUT inversion). The lengths of the amplified PCR products were 150–764 bp. PCR products were purified (Nucleospin gel and PCR clean-up, Macherey–Nagel) and sequenced using Sanger for verification (Eurofins Genomics).

Phylogenetic analysis. A phylogenetic analysis of *S. americana* putative ORs was performed using the 159 sequences contained in Database A. We used the 159 aligned amino acid sequences of SameOR against the previously identified *S. gregaria* and *L. migratoria* OR amino acid sequences^{39,40} (Supplementary Data 3 and 4). Alignments obtained from CLUSTALO were used to generate a maximum likelihood-based tree with the IQ-Tree software package (<http://www.iqtree.org>)⁶³ using the “auto” substitution model and default values for Perturbation Strength and IQ-tree stopping rule. The tree was generated using the ultrafast bootstrap approximation approach (UFBoot2)⁶⁴ with 1000 iterations and a 0.99 correlation coefficient. The phylogenetic tree was rooted using the Odorant Receptor CO-receptor (Orco) sequences of five insect species (*D. melanogaster*, *A. gambiae*, *M. sexta*, *S. gregaria*, *L. migratoria*) and the deduced amino acid sequence of the *S. americana* Orco identified in this work. Also, these six Orco sequences were aligned with CLUSTALO (Supplementary Data 3) and a maximum likelihood-based tree was generated with the IQ-tree software using the same parameters previously described for the OR sequences. iTOL (<https://itol.embl.de/>)⁶⁵ was used to visualize and annotate the dendrograms.

Differential expression analysis. Differential expression analysis of predicted putative ORs was performed by aligning trimmed RNAseq reads against the indexed full putative OR CDS subset using kallisto v0.46.0⁶⁶. The resulting quantification was imported into DESeq2 for differential expression testing⁶⁷ across separate female and male tissue samples using tximport as described in DESeq2 vignettes using a two-factor tissue and gender model. Gene expression in female and male tissue was considered significantly different when characterized by an adjusted p-value (padj) of under 0.1.

RNAscope multiplex fluorescent assay. Fresh frozen sections of antenna and maxillary palp tissues were processed following a standard protocol (Advance Cell Diagnostics Inc., Newark, CA). Briefly, antennal distal segments 1–14 and maxillary palp dome were cut in longitudinal and coronal serial sections, respectively, of 14–16 µm. These regions were selected because they contain most olfactory sensilla in the adult: the antenna segments 1–14 contain >80% of basiconic and trichoid sensilla⁵⁷ while the palp’s dome contains all of the basiconic sensilla^{25,26,31}. Serial sections were mounted onto SuperFrost plus slides, dried at room temperature, and stored in air-tight slide boxes at –80 °C until further processing. Sections were fixed with cold 4% PFA for 15 min at 4 °C followed by ethanol dehydration. For pre-treatment, slides were incubated with RNAscope Protease III for ~25 min at 40 °C using a HybEZ oven. Slides were incubated with customized probes for SameOrco (871191-C1) plus one or a combination of the following: SameOR1 (1201441-C2), SameOR2 (1201451-C3), SameOR58 (previously named SameOR3, 871201-C2), SameOR71 (previously named SameOR111, 872061-C4), SameOR160 (SameOR102, 871221-C3) for 2 h at 40 °C, followed by signal amplification and detection using the RNAscope Multiplex Fluorescent Reagent Kit version 2. For positive and negative control conditions we used a customized probe for SameGAPDH (872871) and DapB (of *Bacillus subtilis* strain) provided by the manufacturer (321831). For signal detection, Opal 520, 570, 620, and 690 fluorophores (AKOYA Biosciences, Marlborough, MA) were prepared at 1:1000 dilution in TSA buffer. Sections were counterstained with DAPI and mounted with ProLong Diamond Antifade Mountant (Thermo Fisher Scientific). Samples were visualized and image were acquired with a Zeiss LSM 880 Airy confocal microscope. Each condition was replicated in at least two animals. The number of sections analyzed per antenna segment for each condition ranged from 3 to 6 per animal. All confocal images shown are maximum projections unless otherwise stated.

Data availability

The data generated and analyzed during the current study is available in SRA repository under the BioProject ID PRJNA889432 [<https://www.ncbi.nlm.nih.gov/sra/PRJNA889432>] and in GenBank repository with the accession numbers: OP777696–OP777865, OP777867, OP777871–72, OP777874, OP777877–79, OP777883, OP777892. In addition, data is included in this manuscript and its supplementary information files.

Received: 1 November 2022; Accepted: 28 December 2022

Published online: 31 December 2022

References

- Murlis, J., Willis, M. A. & Carde, R. T. Spatial and temporal structures of pheromone plumes in fields and forests. *Physiol. Entomol.* **25**(3), 211–222 (2000).
- Carlson, J. R. Olfaction in *Drosophila*: From odor to behavior. *Trends Genet.* **12**(5), 175–180 (1996).
- Buck, L. B. Information coding in the vertebrate olfactory system. *Annu. Rev. Neurosci.* **19**, 517–544 (1996).
- Mombaerts, P. Seven-transmembrane proteins as odorant and chemosensory receptors. *Science* **286**(5440), 707–711 (1999).
- de Bruyne, M., Foster, K. & Carlson, J. R. Odor coding in the *Drosophila* antenna. *Neuron* **30**, 537–552 (2001).
- Vosshall, L. B. Olfaction in *Drosophila*. *Curr. Opin. Neurobiol.* **10**(4), 498–503 (2000).
- Krieger, J. & Breer, H. Olfactory reception in invertebrates. *Science* **286**(5440), 720–723 (1999).
- Grabe, V. & Sachse, S. Fundamental principles of the olfactory code. *Biosystems* **164**, 94–101 (2018).
- Davis, R. L. Traces of *Drosophila* memory. *Neuron* **70**(1), 8–19 (2011).
- de Belle, J. S. & Heisenberg, M. Associative odor learning in *Drosophila* abolished by chemical ablation of mushroom bodies. *Science* **263**(5147), 692–695 (1994).
- Laurent, G. Dynamical representation of odors by oscillating and evolving neural assemblies. *Trends Neurosci.* **19**(11), 489–496 (1996).
- Wilson, R. I. Early olfactory processing in *Drosophila*: Mechanisms and principles. *Annu. Rev. Neurosci.* **36**, 217–241 (2013).
- Gao, Q. & Chess, A. Identification of candidate *Drosophila* olfactory receptors from genomic DNA sequence. *Genomics* **60**(1), 31–39 (1999).
- Vosshall, L. B. *et al.* A spatial map of olfactory receptor expression in the *Drosophila* antenna. *Cell* **96**(5), 725–736 (1999).
- Clyne, P. J. *et al.* A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron* **22**(2), 327–338 (1999).
- Smart, R. *et al.* *Drosophila* odorant receptors are novel seven transmembrane domain proteins that can signal independently of heterotrimeric G proteins. *Insect. Biochem. Mol. Biol.* **38**(8), 770–780 (2008).
- Larsson, M. C. *et al.* Or83b encodes a broadly expressed odorant receptor essential for *Drosophila* olfaction. *Neuron* **43**(5), 703–714 (2004).
- Laurent, G. & Davidowitz, H. Encoding of olfactory information with oscillating neural assemblies. *Science* **265**(5180), 1872–1875 (1994).
- Halle, E. A., Ho, M. G. & Carlson, J. R. The molecular basis of odor coding in the *Drosophila* antenna. *Cell* **117**(7), 965–979 (2004).
- Stocker, R. F. The organization of the chemosensory system in *Drosophila melanogaster*: A review. *Cell. Tissue Res.* **275**(1), 3–26 (1994).
- Ochieng, S., Hallberg, H. & Hansson, B. S. Fine structure and distribution of antennal sensilla of the desert locust, *Schistocerca gregaria* (Orthoptera: Acrididae). *Cell Tissue Res.* **291**, 525–536 (1998).
- Su, C. Y. *et al.* Non-synaptic inhibition between grouped neurons in an olfactory circuit. *Nature* **492**(7427), 66–71 (2012).
- Kay, L. M. & Stopfer, M. Information processing in the olfactory systems of insects and vertebrates. *Semin. Cell Dev. Biol.* **17**(4), 433–442 (2006).
- Leitch, B. & Laurent, G. GABAergic synapses in the antennal lobe and mushroom body of the locust olfactory system. *J. Comp. Neurol.* **372**(4), 487–514 (1996).
- Blaney, W. M. & Chapman, R. F. The fine structure of the terminal sensilla on the maxillary palps of *Schistocerca gregaria* (Forsk.) (Orthoptera, Acrididae). *Z. Zellforsch. Mikrosk. Anat.* **99**(1), 74–97 (1969).
- Blaney, W. M. The ultrastructure of an olfactory sensillum on the maxillary palps of *Locusta migratoria* (L.). *Cell Tissue Res.* **184**(3), 397–409 (1977).
- Jin, X., Zhang, S. G. & Zhang, L. Expression of odorant-binding and chemosensory proteins and spatial map of chemosensilla on labial palps of *Locusta migratoria* (Orthoptera: Acrididae). *Arthropod. Struct. Dev.* **35**(1), 47–56 (2006).
- Pregitzer, P. *et al.* Expression of odorant-binding proteins in mouthpart palps of the desert locust *Schistocerca gregaria*. *Insect. Mol. Biol.* **28**(2), 264–276 (2019).
- Sun, L. *et al.* Odor-induced vomiting is combinatorially triggered by palp olfactory receptor neurons that project to the lobus glomerulatus in locust brain. *Front. Physiol.* **13**, 855522 (2022).
- Dweck, H. K. *et al.* Olfactory channels associated with the *Drosophila* maxillary palp mediate short- and long-range attraction. *Elife* **5**, 14965 (2016).
- Blaney, W. M. Electrophysiological responses of the terminal sensilla on the maxillary palps of *Locusta migratoria* (L.) to some electrolytes and non-electrolytes. *J. Exp. Biol.* **60**(2), 275–293 (1974).
- Jia, L. & Long, Z. Indication of bioactive candidates among body volatiles of gregarious adult locusts *Locusta migratoria* manilensis by electroantennography (EAG) test. *Afr. J. Biotech.* **10**(45), 9170–9176 (2011).
- Zhang, L. *et al.* An odorant-binding protein mediates sexually dimorphic behaviors via binding male-specific 2-heptanone in migratory locust. *J. Insect. Physiol.* **118**, 103933 (2019).
- Smith-Unna, R. *et al.* TransRate: Reference-free quality assessment of *de novo* transcriptome assemblies. *Genome Res.* **26**(8), 1134–1144 (2016).
- Simao, F. A. *et al.* BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**(19), 3210–3212 (2015).
- Raghavan, V. *et al.* A simple guide to *de novo* transcriptome assembly and annotation. *Brief Bioinform.* **23**(2), 563 (2022).
- Yang, Y. *et al.* The olfactory co-receptor Orco from the migratory locust (*Locusta migratoria*) and the desert locust (*Schistocerca gregaria*): Identification and expression pattern. *Int. J. Biol. Sci.* **8**(2), 159–170 (2012).
- Jiang, X., Breer, H. & Pregitzer, P. Sensilla-specific expression of odorant receptors in the desert locust *Schistocerca gregaria*. *Front. Physiol.* **10**, 1052 (2019).
- Wang, Z. *et al.* Identification and functional analysis of olfactory receptor family reveal unusual characteristics of the olfactory system in the migratory locust. *Cell Mol. Life Sci.* **72**(22), 4429–4443 (2015).
- Pregitzer, P. *et al.* In search for pheromone receptors: Certain members of the odorant receptor family in the desert locust *Schistocerca gregaria* (Orthoptera: Acrididae) are co-expressed with SNMP1. *Int. J. Biol. Sci.* **13**(7), 911–922 (2017).
- Jones, W. D. *et al.* Functional conservation of an insect odorant receptor gene across 250 million years of evolution. *Curr. Biol.* **15**(4), R119–R121 (2005).
- Tauxe, G. M. *et al.* Targeting a dual detector of skin and CO₂ to modify mosquito host seeking. *Cell* **155**(6), 1365–1379 (2013).
- Eldem, V. *et al.* Transcriptome Analysis for Non-Model Organism: Current Status and Best-Practices. in *Applications of RNA-Seq and Omics Strategies: From Microorganisms to Human Health*. (2017).
- Krieger, J. *et al.* A candidate olfactory receptor subtype highly conserved across different insect orders. *J. Comp. Physiol. A* **189**(7), 519–526 (2003).
- Prieto-Godino, L. L. *et al.* Olfactory receptor pseudo-pseudogenes. *Nature* **539**(7627), 93–97 (2016).
- Li, H. *et al.* Expressions of olfactory proteins in locust olfactory organs and a palp odorant receptor involved in plant aldehydes detection. *Front. Physiol.* **9**, 663 (2018).

47. Iatrou, K. & Biessmann, H. Sex-biased expression of odorant receptors in antennae and palps of the African malaria vector *Anopheles gambiae*. *Insect. Biochem. Mol. Biol.* **38**(2), 268–274 (2008).
48. Couto, A., Alenius, M. & Dickson, B. J. Molecular, anatomical, and functional organization of the *Drosophila* olfactory system. *Curr. Biol.* **15**(17), 1535–1547 (2005).
49. Tom, M. T. *et al.* Sex- and tissue-specific expression of chemosensory receptor genes in a hawkmoth. *Front. Ecol. Evol.* **10**, 521 (2022).
50. Lemke, R. S. *et al.* SNMP1 and odorant receptors are co-expressed in olfactory neurons of the labial and maxillary palps from the desert locust *Schistocerca gregaria* (Orthoptera: Acrididae). *Cell Tissue Res.* **379**(2), 275–289 (2020).
51. Levai, O. *et al.* Cells in the vomeronasal organ express odorant receptors but project to the accessory olfactory bulb. *J. Comp. Neurol.* **498**(4), 476–490 (2006).
52. Dobritsa, A. A. *et al.* Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. *Neuron* **37**(5), 827–841 (2003).
53. Krieger, J. *et al.* Genes encoding candidate pheromone receptors in a moth (*Heliothis virescens*). *Proc. Natl. Acad. Sci. USA* **101**(32), 11845–11850 (2004).
54. Sakurai, T. *et al.* Identification and functional characterization of a sex pheromone receptor in the silkworm *Bombyx mori*. *Proc. Natl. Acad. Sci.* **101**(47), 16653–16658 (2004).
55. Grosse-Wilde, E. *et al.* Sex-specific odorant receptors of the tobacco hornworm *Manduca sexta*. *Front. Cell Neurosci.* **4**, 22 (2010).
56. Greenwood, M. & Chapman, R. F. Differences in numbers of sensilla on the antennae of solitary and gregarious *Locusta migratoria* L. (Orthoptera: Acrididae). *Int. J. Insect Morphol. Embryol.* **13**(4), 295–301 (1984).
57. Chapman, R. F. Development of phenotypic differences in sensillum populations on the antennae of a grasshopper, *Schistocerca americana*. *J. Morphol.* **254**(2), 186–194 (2002).
58. Guo, X. *et al.* 4-Vinylanisole is an aggregation pheromone in locusts. *Nature* **584**(7822), 584–588 (2020).
59. You, Y. *et al.* A broadly tuned odorant receptor in neurons of trichoid sensilla in locust *Locusta migratoria*. *Insect. Biochem. Mol. Biol.* **79**, 66–72 (2016).
60. Cui, X., Wu, C. & Zhang, L. Electrophysiological response patterns of 16 olfactory neurons from the trichoid sensilla to odorant from fecal volatiles in the locust, *Locusta migratoria manilensis*. *Arch. Insect. Biochem. Physiol.* **77**(2), 45–57 (2011).
61. Moreno-Santillan, D. D. *et al.* De Novo Transcriptome Assembly and Functional Annotation in Five Species of Bats. *Sci Rep* **9**(1), 6222 (2019).
62. Hallgren, J., *et al.* DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. *BioRxiv* (2022).
63. Trifinopoulos, J. *et al.* W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**(W1), W232–W235 (2016).
64. Hoang, D. T. *et al.* UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**(2), 518–522 (2018).
65. Letunic, I. & Bork, P. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **49**(W1), W293–W296 (2021).
66. Bray, N. L. *et al.* Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**(5), 525–527 (2016).
67. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**(12), 550 (2014).

Acknowledgements

We thank Michael Rosario for help with initial experiments, Vincent Schram from the Microscopy and Imaging Core NICHD for help with confocal scanning, and the Molecular and Genomics Core NICHD for help with the RNA sequencing. We also thank members of Stopfer lab for helpful discussion and comments, and Kui Sun for her excellent animal care. This work utilized the computational resources of the NIH High Performance Computing Biowulf cluster (<https://hpc.nih.gov>). E.D.M. was funded by Pew Latin American postdoctoral fellowship. This work was supported by an intramural Grant from NIH-NICHD to M.S.

Author contributions

Conceptualization: A.B.G. and M.S.; Methodology: A.B.G., M.S., J.I., E.D.M.; Investigation: A.B.G. and M.S.; Writing—original draft: A.B.G., and M.S.; Writing—Review and Editing: A.B.G., M.S., J.I., E.D.M.; Supervision: A.B.G., and M.S.

Funding

Open Access funding provided by the National Institutes of Health (NIH).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-27199-3>.

Correspondence and requests for materials should be addressed to M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

This is a U.S. Government work and not under copyright protection in the US; foreign copyright protection may apply 2022