Enormous expansion of the chemosensory gene repertoire in the omnivorous German cockroach *Blattella germanica*

Hugh M. Robertson[1] [iD] | Rachel L. Baits[1] | Kimberly K.O. Walden[1] |
Ayako Wada-Katsumata[2] | Coby Schal[2]

[1]Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[2]Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, North Carolina, USA

Correspondence
Hugh Robertson, Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL 61801.
Email: hughrobe@uiuc.edu

Abstract

The acquisition of genome sequences from a wide range of insects and other arthropods has revealed a broad positive correlation between the complexity of their chemical ecology and the size of their chemosensory gene repertoire. The German cockroach *Blattella germanica* is an extreme omnivore and has the largest chemosensory gene repertoire known for an arthropod, exceeding even the highly polyphagous spider mite *Tetranychus urticae*. While the Odorant Receptor family is not particularly large, with 123 genes potentially encoding 134 receptors (105 intact), the Gustatory Receptor family is greatly expanded to 431 genes potentially encoding 545 receptors (483 intact), the largest known for insects and second only to the spider mite. The Ionotropic Receptor family of olfactory and gustatory receptors is vastly expanded to at least 897 genes (604 intact), the largest size known in arthropods, far surpassing the 150 known from the dampwood termite *Zootermopsis nevadensis*. Commensurately, the Odorant Binding Protein family is expanded to the largest known for insects at 109 genes (all intact). Comparison with the far more specialized, but phylogenetically related termite, within the Dictyoptera, reveals considerable gene losses from the termite, and massive species-specific gene expansions in the cockroach. The cockroach has lost function of 11%–41% of these three chemoreceptor gene families to pseudogenization, and most of these are young events, implying rapid turnover of genes along with these major expansions, presumably in response to changes in its chemical ecology.

1 | INTRODUCTION

The past two decades have seen the availability of genome sequences for numerous insects and other arthropods. Among the many insights these have provided is recognition of the extremely variable sizes of the various gene families involved in allowing arthropods to sense external chemicals. These gene families include the Odorant Binding Protein (OBP), Odorant Receptor (OR), Gustatory Receptor (GR), and Ionotropic Receptor (IR) families, as well as some smaller families (Leal, 2013; Benton 2015; Joseph & Carlson, 2015; Rimal & Lee, 2018). Arthropods vary enormously in the complexity of their chemical ecology, and their chemosensory gene repertoires covary with it. Extreme examples of this relationship are the most prominent. For example, the smallest chemosensory gene families are found in obligate parasites or mutualists with limited need for diverse chemical

sensing ability, for example, the human body louse *Pediculus humanus* (Kirkness et al., 2010) and the fig wasp *Ceratosolen solmsi* (Xiao et al., 2013). In contrast, there are insects with complex chemically mediated social lives, like ants, with massive expansions of the OR family (Smith et al., 2011a, b; Zhou et al., 2012), many of which are now known to mediate perception of their highly diverse cuticular hydrocarbons (Pask et al., 2017), or highly polyphagous species like the moth *Spodoptera frugiperda* (Gouin et al., 2017) and the spider mite *Tetranychus urticae* (Ngoc et al., 2016), with massive expansions of the GR family. More subtle examples are also available, for example, the depauperate GR family in the honey bee *Apis mellifera* (Robertson & Wanner, 2006), thought to be commensurate with its mutualistic relationship with plants, and the considerable contraction of the OR and GR families of the tsetse fly *Glossina palpalis* (Obiero et al., 2014), commensurate with the reduced complexity of its chemical ecology

compared to most flies. Finer-grained examination of close relatives has revealed the on-going processes of gene gain and loss that mediate these grander patterns, for example, in the monophagous *Drosophila sechellia* on the Seychelles islands (McBride, 2007; McBride & Arguello, 2007) or the unusual pestiferous *Drosophila suzukii* (Hickner et al., 2016). Genome sequences are nearly essential for discovering the size and complexity of these processes as these gene families commonly encode highly divergent proteins making them difficult to identify by screening methods, and they are commonly expressed at such low levels that transcriptome studies, even of chemosensory tissues such as antennae, palps and legs, will only detect some of them unless carried out at extreme sequencing depths, for example, the neuro-transcriptome of the yellow fever mosquito *Aedes aegypti* (Matthews, McBride, DeGennaro, Despo, & Vosshall, 2016). Even then, there are gene family members not expressed in obvious chemosensory tissues, whereas pseudogenes are rarely transcribed and provide useful insights into gene family evolution (e.g., Smith et al., 2011a). The ongoing deluge of arthropod genome sequences promises to provide many more examples of the connection of chemosensory repertoire to chemical ecology, and one more extreme example is described here.

The German cockroach, *Blattella germanica* Linnaeus, is a widespread human commensal species (Schal, Gautier, & Bell, 1984; Schal, 2011), where it causes considerable problems beyond annoyance, including allergic responses leading to asthma (Gore & Schal, 2007; Rabito, Carlson, He, Werthmann, & Schal, 2017) and being a passive vector for potential pathogens. It is an extreme omnivore, feeding on almost any available foods, with a preference for "sweet" tastes (Schal et al., 1984). This biological preference was exploited in the development of insecticide baits that use sugars such as glucose (Schal & Hamilton, 1990). Resistance to these baits has evolved via the expected pathways of insecticide resistance (e.g., Gondhalekar & Scharf, 2012), but also remarkably by evolution of aversion to the sugar bait (Silverman & Bieman, 1993; Silverman & Ross, 1994). This aversiveness to glucose has been demonstrated to involve perception of glucose by the "bitter" neuron in each gustatory sensillum (Wada-Katsumata, Silverman, & Schal, 2011, 2013). This switch might involve misexpression of a glucose receptor in these "bitter-sensing" neurons or modified recognition of glucose by a receptor that normally senses a "bitter" compound. Thus, in addition to illuminating the chemosensory biology of this cockroach, documenting the major gene families encoding chemosensory proteins is a prerequisite to attempts to understand the molecular basis of this sugar aversiveness. Food preferences are also used in mate-recognition and acceptance, as the courting male offers the female a nuptial gift in his tergal gland rich in sugars (maltose and other oligosaccharides) and phospholipids (Wada-Katsumata, Ozaki, Yokohari, Nishikawa, & Nishida, 2009). This cockroach also uses chemoperception in other contexts, including long-range mate-finding with volatile sex pheromones (Nojima, Schal, Webster, Santangelo, & Roelofs, 2005), contact-based sexual recognition with derivatives of cuticular hydrocarbons (Eliyahu, Nojima, Mori, & Schal, 2008), and aggregation (Wada-Katsumata et al., 2015), so a complete documentation of its chemosensory genes and encoded proteins lays the ground for improved understanding of many aspects of the chemical ecology of *B. germanica*.

Three previous publications describe aspects of the chemosensory repertoire of this cockroach. Zhou et al. (2014) described partial sequences for 14 OBPs, two ORs, and four GRs from a whole body transcriptome conducted using pyrosequencing, whereas Niu, Liu, Dong, and Dong (2016) expanded the OBP total to 48 mostly complete sequences by performing an antennal transcriptome using ILLUMINA sequencing and also found five ORs and 5 IRs, albeit mostly partial sequences. The sequencing of the *B. germanica* genome as part of an i5k pilot project (Robinson et al., 2011; i5k Consortium 2013) allowed Harrison et al. (2018) to compare the repertoires of intact OR and IR genes of this cockroach with those of the dampwood termite *Zootermopsis nevadensis* (Terrapon et al., 2014) and two other termites they sequenced, revealing massive expansion of the IR family in this cockroach relative to the termites. Here we describe all four gene families in complete detail, including their many pseudogenes, and reveal that not only is the IR family massively expanded in this cockroach far beyond that known for any other arthropod, but the GR and OBP families are the largest known for insects, commensurate with the broad requirements of this cockroach to sense diverse chemicals in its environment.

2 | MATERIALS AND METHODS

Searches for *B. germanica* chemosensory genes were conducted on the genome assembly of Harrison et al. (2018) using TBLASTN at the i5k Workspace (Poelchau et al., 2014) with proteins from the *Z. nevadensis* families and other insects (Terrapon et al., 2014) and E values up to 1000. Iterative searches with newly discovered genes and their encoded proteins were undertaken in an effort to exhaustively discover all gene family members. Gene models were built in the Apollo browser at the i5k Workspace, with supporting evidence from RNAseq data from three sources, all generated with ILLUMINA sequencing: the antennal set from Niu et al. (2016), a head set from Drinnenberg, Henikoff, and Malik (2014) available from the Short Read Archive (SRA) at NCBI as SRX682022, and a head set generated by A. W-K. and C. S. (available in the SRA as SRX3189901/2). Partial gene models resulting from difficulties with the genome assembly were repaired when possible using a combination of the above RNAseq reads and raw genome reads from the SRA. Pseudogenes were translated as best possible, using Z for stop codons and X for frameshifts and other obvious pseudogenizing mutations like large insertions or deletions and splice junction mutants, but only named and included if they encoded at least half the length of a typical gene family member. Some pseudogenes were so badly degraded that despite being nearly full-length they were not reconstructed and are not included in the protein sets, but were included in the pseudogene statistics. The same length criterion was employed for gene fragments that could not be repaired. Proteins from *B. germanica* and *Z. nevadensis* were aligned within each gene family, along with representatives from other insects, using ClustalX v2.1 (Larkin et al., 2007), and gene models were refined in light of these alignments. All protein sequences are available as Supporting Information, and the transcripts and protein sequences for all intact and contiguous models are available from the i5k Workspace.

For phylogenetic analyses, aligned protein datasets were trimmed using TrimAl v1.4 (Capella-Gutierrez et al. 2009), using the "gappyout" option for the OR, GR, and OBP families, which are of reasonably uniform length, and the "strict" option for the IRs, which vary considerably in the length and sequence of their N-termini, most of which was effectively removed from the alignment. Maximum likelihood phylogenetic analysis was performed with PhyML v3.0 (Guindon et al., 2010) using default settings with support for nodes evaluated using their approximate Likelihood Ratio Tests (aLRT). Trees were arranged and colored with FigTree v1.4.2 (https://tree.bio.ed.ac.uk/software/figtree/), and figures prepared in Adobe Illustrator.

Expression levels of the OBPs were compared between the antennal RNAseq of Niu et al. (2016) and our RNAseq from heads with antennae (SRX3189901/2). Trimmed reads were aligned to transcripts from our complete gene models, with 5′ and 3′ untranslated regions included, using the Burrows-Wheeler Aligner (BWA) (Li & Durbin, 2009). Samtools (Li et al., 2009) was used to sort, index, and summarize the BWA. Read counts were standardized as counts per kb. These two libraries are of comparable size with 67,706,096 reads from Niu et al. (2016) and 77,726,077 reads from our heads-with-antennae RNAseq, so counts were not standardized by library depth as our comparisons are primarily within each dataset.

3 | RESULTS AND DISCUSSION

3.1 | The OR family

The OR family evolved from a lineage of the diverse GR family within insects (Robertson, Warr, & Carlson, 2003; Missbach et al., 2014; Ioannidis et al., 2017), with the wingless firebrat *Thermatobia domestica* expressing at least three members of the family in their antennae, but the slightly older lineage of the bristletail *Lepismachilis y-signata* appearing not to have any ORs (Missbach et al., 2014). In all insects with a genome sequence examined to date the family consists of a single highly conserved gene encoding a coreceptor, known as Orco, as well as 4–400 "specific" ORs that mediate the specificity and sensitivity of insect olfaction (Leal, 2013; Benton 2015; Joseph & Carlson, 2015). The family ranges enormously in size in insects, from a low of five genes in the damselfly *Calopteryx splendens* (Ioannidis et al., 2017), which was previously thought to be anosmic, and 13 in the obligate parasitic human body louse *P. humanus* (Kirkness et al., 2010), through 60 genes encoding 62 receptors in *D. melanogaster* (Robertson et al., 2003), up to ~400 in some ants (Smith et al., 2011a,b; Zhou et al., 2012), where up to half of them likely mediate perception of the enormously diverse cuticular hydrocarbons involved in nestmate recognition and other social cues (Smith et al., 2011a,b; Pask et al., 2017). The ligands of most of the *D. melanogaster* ORs are known (Hallem, Ho, & Carlson, 2004), as are those from *Anopheles gambiae* (Carey, Wang, Su, Zwiebel, & Carlson, 2010; Wang et al. 2010), and some others in other insects, for example, pheromone receptors in diverse species (Wanner et al., 2007; Leary et al., 2012; Andersson et al., 2016).

The OR family in *B. germanica* is of an intermediate size, with 134 potential transcripts from 123 genes in addition to the Orco gene.

Five genes exhibit an unusual form of alternative splicing previously described in this and the GR family in many insects including *D. melanogaster* (Clyne, Warr, & Carlson, 2000; Robertson et al., 2003), in which tandemly arrayed long first exons are alternatively spliced into a shared set of exons encoding the C-terminus of the protein, the most conserved region of the protein. Twenty-nine of these genes or transcripts are pseudogenic (21.6%), leaving 105 apparently intact specific ORs. Niu et al. (2016) described only five ORs from their antennal transcriptome, and all are partial sequences, so the family here was named afresh, attempting to give related receptors, especially in arrays within scaffolds, consecutive names (their OR1 is Orco, Or2 is Or49, Or3 is Or33, Or4 is Or100, and Or5 is Or48).

Phylogenetic analysis of these OR proteins with their 70 relatives from *Z. nevadensis* (Terrapon et al., 2014), reveals the kinds of gene family evolution well known for this family (e.g., Benton, 2015) (Figure 1). Beyond the highly conserved Orco lineage, there are many instances of apparent 1:1 orthologs between the cockroach and termite, as expected from their close phylogenetic relationship, as well as many instances of gene losses and duplications in one or both lineages. The most extreme examples of the latter are expansions of 18 proteins in the termite and 15 and 20 proteins in the cockroach. Most pseudogenes are found in these and other smaller expansions, suggesting loss of gene function as these receptors sometimes became irrelevant to the species chemical ecology. The overall larger size of the cockroach OR repertoire is a combination of fewer gene losses and more duplications, many of which have short branches in the tree indicating that they are the result of recent duplications in the cockroach.

3.2 | The GR family

The GR family consists of multiple divergent subfamilies (Clyne et al., 2000; Robertson et al., 2003) and dates back at least to the earliest animals where they are called GR-Like genes (Benton 2015; Robertson, 2015; Saina et al., 2015; Eyun et al., 2017). It varies in size from a low of five genes in the obligate mutualist fig wasp *C. solmsi* (Xiao et al., 2013) and 12 in the honey bee *A. mellifera* (Robertson & Wanner, 2006), through 60 genes encoding 68 receptors in *D. melanogaster* (Clyne et al., 2000; Scott et al., 2001; Robertson et al., 2003), to 222 and 286 in the red flour beetle *Tribolium castaneum* (Richards et al., 2008) and the Asian longhorned beetle *Anoplophora glabripennis* (McKenna et al., 2016), respectively. The sugar receptors are the oldest distinctive recognizable subfamily of known function, dating back to crustaceans (Peñalva-Arana, Lynch, & Robertson, 2009), whereas the distinctive subfamily that contains the carbon dioxide receptors of endopterygotes dates at least to the odonates (Iaonnidis et al. 2017). In *D. melanogaster* most of the remaining GRs are involved in perception of bitter tastants (Weiss, Dahanukar, Kwon, Banerjee, & Carlson, 2011), however some are implicated in perception of pheromones (Bray & Amrein, 2003; Watanabe, Toba, Koganezawa, & Yamamoto, 2011), Gr43a (a fructose receptor) and other GRs are involved in both peripheral detection and central monitoring of tastants (Miyamoto, Slone, Song, & Amrein, 2012; Chen & Dahanukar, 2017), and others are expressed in non-chemosensory tissues (Thorne & Amrein, 2008) and Gr28b senses light and heat (Ni et al., 2013).
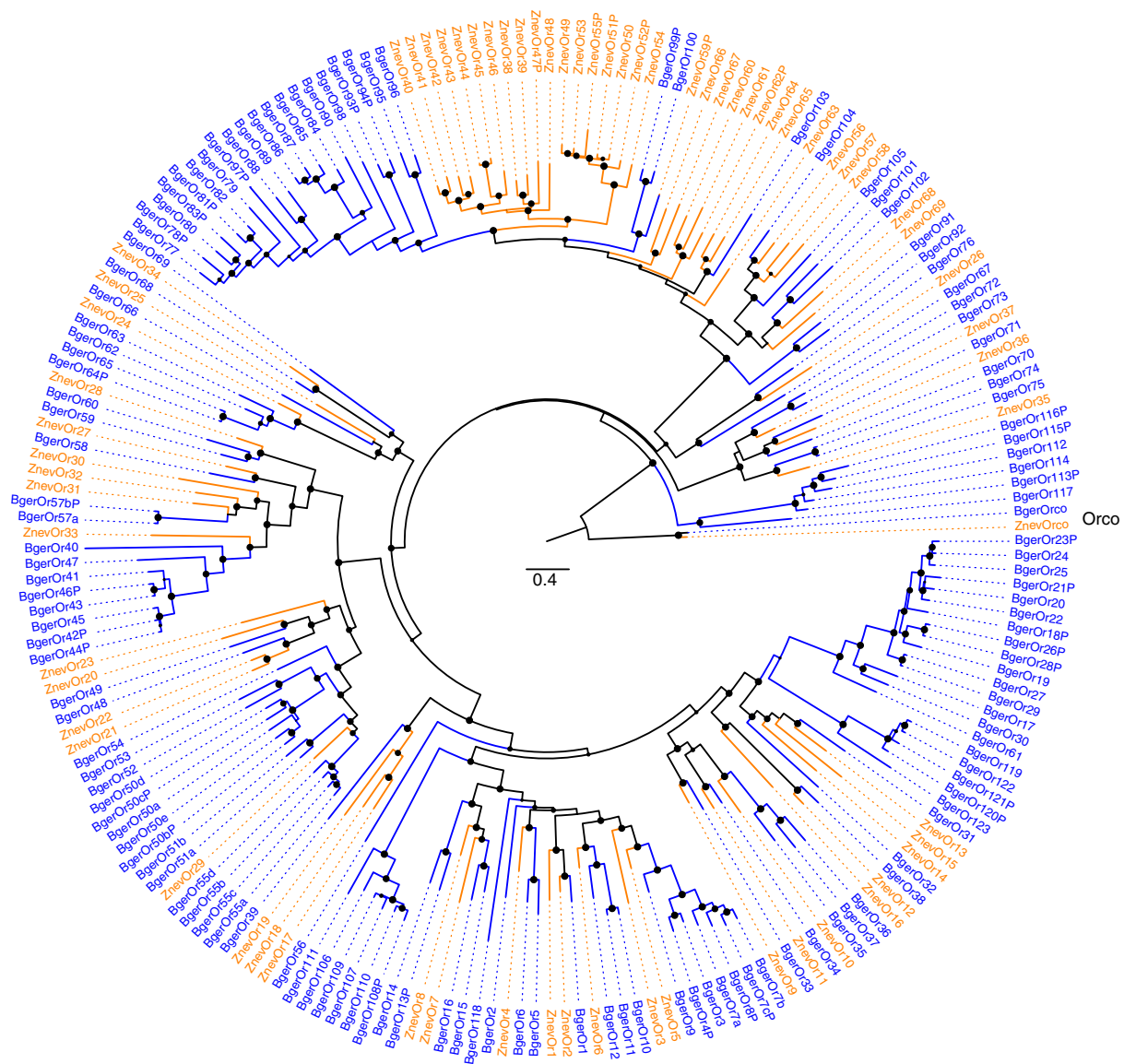
FIGURE 1 Phylogenetic relationships of the OR family. The tree was rooted with the Orco proteins, based on the conserved sequence and basal position of this protein within the OR family in analyses of the entire insect chemoreceptor superfamily (Robertson et al., 2003). *B. germanica* proteins are in blue and *Z. nevadensis* in orange, as are the branches leading to them. The suffix P after a name indicates that it is pseudogenic. The scale bar is substitutions per site, and filled circles on nodes indicate support levels as aLRT (approximate Likelihood Ratio Test) values from PHYML v3.0 ranging from 0 to 1 [Color figure can be viewed at wileyonlinelibrary.com]

B. germanica has the largest known GR family in insects, with 431 genes potentially encoding 545 proteins through alternative splicing of 26 genes, although 62 (11.4%) of these genes or transcripts are pseudogenic, leaving a potentially intact set of 483 GRs. Only the extremely polyphagous spider mite *T. urticae* has more GRs with 689 (Ngoc et al., 2016). The sugar receptor subfamily consists of 14 genes (BgerGr1–14), compared with six in the termite, through both loss of genes from the termite and duplications in the cockroach (Figure 2). The subfamily that contains the carbon dioxide receptors of endoterygotes is also larger in the cockroach, with 36 genes (Bger15–50) including a cockroach-specific expansion of 16 genes, compared with 25 in the termite. A complex mixture of orthologs, gene losses, and species-specific duplications has been involved in this subfamily. Although these proteins cluster confidently with the carbon diox-

ide receptors of endopterygote insects, their ligands remain to be determined and are unlikely all to be related to sensing of carbon dioxide.

The remaining GRs fall broadly into two large groupings, as do the termite ones. The first is a set of genes that contain three phase-0 introns near the 3′ end, in positions compatible with being homologous with the three introns inferred to be ancestral to the insect chemoreceptor superfamily (Robertson et al., 2003) and present throughout the GRLs in other animals (Robertson 2015; Saina et al., 2015). These genes are Gr51–94 and 418–430. Gr420–430 form basal clusters nearer the root of the tree (indicated in Figure 3), and Gr429 and 430 have acquired three additional introns interrupting the usually long first exon, so they have introns in phases 1-0-1-0-0-0. Four relatives of the latter set were not previously recognized in
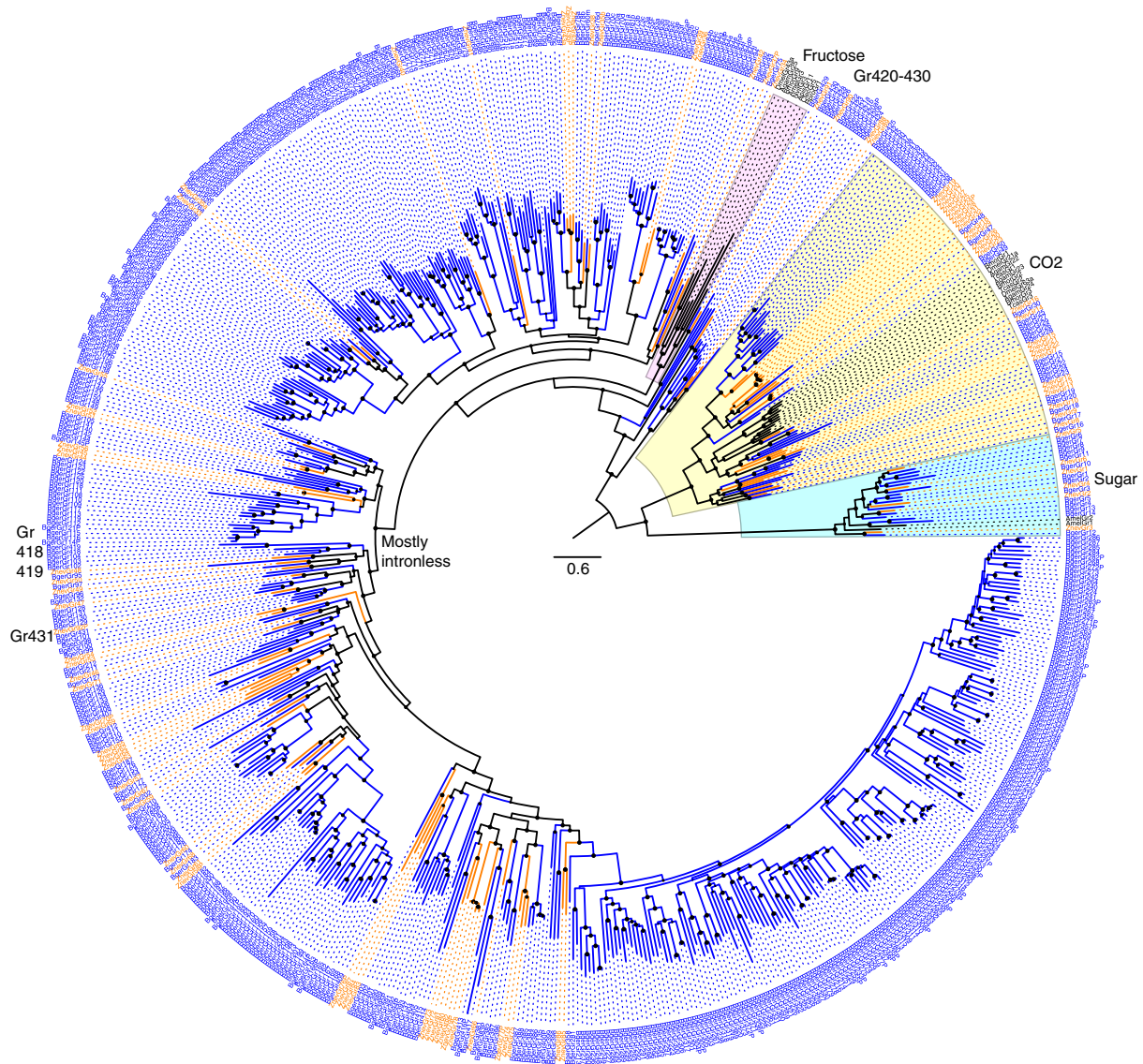
FIGURE 2 Phylogenetic relationships of the GR family. The sugar and carbon dioxide receptor subfamilies together rooted the tree, based on their basal location together in analyses with GRLs of other animals (Robertson, 2015). Major subfamilies are highlighted in colors, and the branch leading to the mostly intronless clade is indicated inside the circle. Non-blattodean insect GRs are in black. Other details are as in the Figure 1 legend [Color figure can be viewed at wileyonlinelibrary.com]

the termite, so are newly described here, specifically ZnevGr88/89 are related to BgerGr424/425, whereas ZnevGr90 and 91 are related to BgerGr426 and 429, respectively. Enigmatically, Gr418/419 cluster in the tree with the intronless genes below. These intron-containing genes are commonly alternatively spliced in a fashion similar to some of the OR genes, with multiple long first exons alternatively spliced into the three shared short final exons, commonly with RNAseq support for at least some of these alternative splices. The largest of these is Gr60a-m encoding 13 distinct receptors, all apparently intact, whereas some isoforms encoded by other genes in this set are pseudogenic. Comparison of the Gr51–94 and 418–430 set with their termite relatives reveals major expansions of several gene lineages, for a total of 171 potential proteins versus just 18 in the termite. Again, although there are a few apparent orthologous relationships, the remaining relationships involve minor expansions in each species, apparent loss from

the termite, and large expansions through alternative splicing in the cockroach.

The second and larger GR set are intronless genes, at least as far as the coding sequences are concerned (for some genes there is RNAseq evidence of an intron in the 5′ untranslated region), comprising a single lineage in the tree (indicated in Figure 2). This intronless clade consists of 44 termite and 322 cockroach genes (Gr95–417). Again, although there are nine apparent orthologous relationships, many other relationships involve duplications in one or both species, there are many apparent gene losses from the termite and very few from the cockroach, and most impressively, there is a massive cockroach-specific expansion of 166 genes (Gr213–379). By comparison with other insects such as *D. melanogaster*, these two large clades of GRs are likely to be involved in perception of "bitter" compounds.
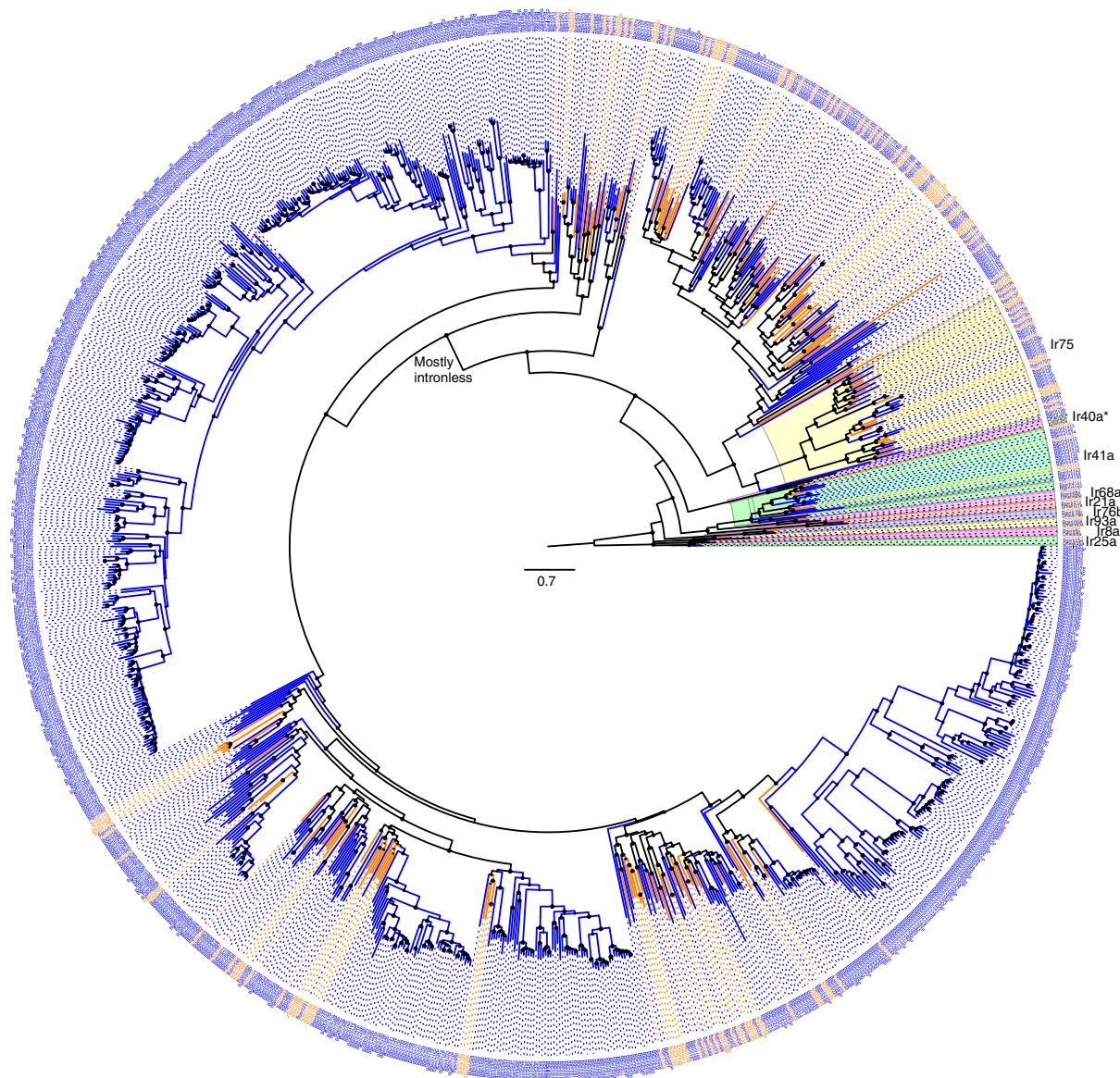
FIGURE 3 Phylogenetic relationships of the IR family. The Ir8a and 25a lineages were declared the outgroup as these two proteins are most similar to the ionotropic glutamate receptors (Croset et al., 2010; Terrapon et al., 2014). Major conserved lineages are highlighted in colors, and their names indicated on the outside of the circle (the asterisk on Ir40a indicates it is missing from both species). The branch leading to the mostly intronless clade is indicated inside the circle. Other details are as in Figures 1 and 2 legends [Color figure can be viewed at wileyonlinelibrary.com]

Finally, a single divergent gene named Gr431 was discovered while searching for relatives of the DmelGr43a fructose receptor lineage. This receptor lineage evolved from within the "bitter" clade of GRs (as opposed to being a divergent member of the sugar receptor sub-family), and is present in all neopteran insects examined to date as at least one gene and sometimes an expanded lineage, except for the termite *Z. nevadensis* (Terrapon et al., 2014). The damselfly *C. splendens* has a set of five proteins that might be related to this lineage, however the association is not robust (Ioannidis et al., 2017). Gr431 has a completely different gene structure than all the others, with introns in phases 0-2-0-1, none of which correspond to any other GR introns in this species or the fructose receptor homologs in other insects, splitting the CDS into five roughly equal-size exons. The termite has a previously unrecognized ortholog of this gene (ZnevGr92P) with the same gene structure, but it is a pseudogene with a stop codon

in the middle of the fourth exon, whose existence is supported by both raw genome reads and some expressed sequence tags, although it is always possible this is a pseudogenic allele specific to the sequenced strain. These two proteins cluster phylogenetically well within the intronless clade (indicated in Figure 2), suggesting that all four of these introns are novel gains in this gene before the cockroach/termite split. This GR is nevertheless unlikely to be a fructose receptor as that clade is quite distinct in the phylogenetic tree.

3.3 | The IR family

The IR family is a variant lineage of the ionotropic glutamate receptor superfamily of ligand-gated ion channels (Benton et al. 2009; Rytz, Croset, & Benton, 2013; Rimal & Lee, 2018), present throughout the protostomes (Croset et al., 2010; Eyun et al., 2017). Like the other

families, IRs are best understood in *D. melanogaster* where they have been shown to mediate a distinct set of olfactory capabilities, especially involving acids and amines (Ai et al., 2010; Silbering et al., 2011; Rytz et al., 2013; Ahn, Chen, & Amrein, 2017; Chen & Amrein, 2017), as well as perception of temperature and humidity (Knecht et al., 2016, 2017; Ni et al., 2016), whereas the large Ir20a clade are involved in gustation (Koh et al., 2014; Stewart, Koh, Ghosh, & Carlson, 2015). The family has three coreceptors, Ir8a, 25a, and 76b, involved in different sensory aspects, as well as various conserved lineages dating back to early insects (Ioannidis et al., 2017). The family ranges from 14 and 19 genes in the pea aphid *Acythosiphon pisum* and *P. humanus*, respectively (Croset et al., 2010; Terrapon et al., 2014) and similarly tens of genes in most Hymenoptera (Croset et al., 2010), through 62 genes in *D. melanogaster* (Benton et al. 2009), to 150 genes in the termite *Z. nevadensis* (Terrapon et al., 2014).

B. germanica has far and away the largest IR family known in any arthropod, with 897 genes, 393 (43.8%) of which are pseudogenes, leaving 604 intact genes. In addition there are many gene fragments encoding less than 50% of a related protein, as well as pseudogenes so badly damaged they could not easily be reconstructed and hence were not included in the naming or analysis. The naming of IRs is complicated. In their survey of IRs across various insects, Croset et al. (2010) gave clear relatives of the *D. melanogaster* proteins the same names and for the rest assigned sequential numbers to each IR independent of species, but this approach is not tenable in the long run. Terrapon et al. (2014) started an alternative approach of naming the remaining IRs in a series beginning with Ir101, which avoids confusion with the *D. melanogaster* names as they were named for their cytological locations and hence only go up to 100a. This approach has now been employed for several species, including the predatory mite *Metaseiulus occidentalis* (Hoy et al., 2016), the damselfly *C. splendens* (Ioannidis et al., 2017), the milkweed bug *Oncopeltus fasciatus* (Panfilio et al., 2017), the Asian longhorned beetle *A. glabripennis* (McKenna et al., 2016), the Colorado potato beetle *Leptinotarsa decemlineata* (Schoville et al., 2017), the tick *Ixodes scapularis* (Josek et al. 2018), and updated *Ae. aegypti* and *An. gambiae* genomes (Matthews et al. 2018). Thus the single orthologs of Ir8a, 21a, 25a, 68a, 76b, and 93a are given those names, and as in the termite, the expanded set of genes related to DmelIr41a are named Ir41a1–16, whereas the expanded set related to DmelIr75a–d are named Ir75a–z. Both the termite and cockroach do not have an obvious ortholog for the Ir40a gene, which is present in the odonate (Iaonnidis et al. 2017). Niu et al. (2016) identified five IRs in their antennal transcriptome, and like the ORs most are partial sequences, however their Ir4 is full length. Their IRs correspond with those described here as follows: their Ir1 is Ir41a11, Ir2 is Ir41a15, Ir3 is Ir41a12, Ir4 is Ir25a, and Ir5 is Ir76b.

Phylogenetic analysis of the IR family reveals the simple orthologous relationships of the conserved IRs, the absence of an Ir40a ortholog, and the expansions of the Ir41a and Ir75 lineages in the cockroach, once again involving some orthologs, gene losses primarily from the termite lineage, and duplications in the cockroach (Figure 3). The remaining IRs, much like the GRs, form two large groupings. The first set of genes generally contains eight introns comparable to those seen in these IRs in many other insects. In the cock-

roach, these are BgerIr101–196 and in the termite ZnevIr101–155. This set of genes again contains some potentially orthologous relationships, as well as the usual losses primarily from the termite, and duplications in one or both species, the largest of which are eight genes, BgerIr152–159, related to ZnevIr132 and 14 genes, BgerIr163–176, related to ZnevIr137/138.

The truly massive IR expansions in the cockroach, however, have occurred within a clade of largely intronless genes. Some of these genes have acquired idiosyncratic single, and rarely two, introns, commonly near their 5′ end and barely interrupting the N-terminal coding region. It is possible that some of these were introns in 5′ untranslated regions that have now become coding. This massively expanded clade is BgerIr196–950 (755 genes) and ZnevIr156–222 (66 genes). As usual it contains a few apparent orthologous relationships, many gene losses from the termite and a few from the cockroach, and most impressively massive cockroach-specific expansions. In stark contrast to the termite, where the largest expansion is just five genes, in the cockroach expansions include 17 genes (Ir226–239) related to ZnevIr163, 46 genes (Ir702–747) with no clear termite relative, 61 genes (Ir592–653) related to ZnevIr190–193, 147 genes (Ir751–870 and 924–950) related to ZnevIr213, and 347 genes (Ir262–567, 748–750, and 884–923) without a clear termite relative.

3.4 | The OBP family

OBPs are small globular secreted proteins, which in the context of insect chemoreception are secreted into the lumen of sensilla from support cells at their base (Leal, 2013; Pelosi, Iovinella, Felicioli, & Dani, 2014). Not all OBPs are involved in chemoperception, however, with some expressed in other cells and tissues (e.g. Foret and Maleszka 2006; Pelosi et al., 2014). The Classic structure for an OBP is to have six highly-conserved cysteines that maintain their tertiary shape via three disulfide bonds, however there are variants with four cysteines called Minus-C, and ones with two additional cysteines, called Plus-C, as well as apparent OBP dimers (Pelosi et al., 2014). The family ranges considerably in size up to 90 genes in the parasitoid wasp *Nasonia vitripennis* (Vieira et al., 2012), and appears to be largely an insect-specific invention (Missbach, Vogel, Hansson, & Groβe-Wilde, 2015), extending back to basal hexapods like Collembola (Pelosi et al., 2014), although similar if not homologous proteins are known from other arthropods (e.g., Renthal et al., 2017; Vizueta et al. 2017; Josek et al. 2018).

Niu et al. (2016) described 48 mostly full-length OBPs from their antennal transcriptome of *B. germanica*, along with extensive data on their expression patterns. Those 48 names were therefore retained, whereas gene models were completed for their partial ones, and 61 models were added for a total of 109 genes, all of which are intact (OBP103 is a pseudogene in the genome assembly with a single-base-insertion frameshift in a homopolymer region of exon7, but examination of raw genome and RNAseq reads indicates it is polymorphic with an intact version, which was employed instead). Eleven additional models required repair of the genome assembly and all models are now full length. This is the largest known OBP family. Because Niu et al. (2016) described and named their proteins from a transcriptome, the genes
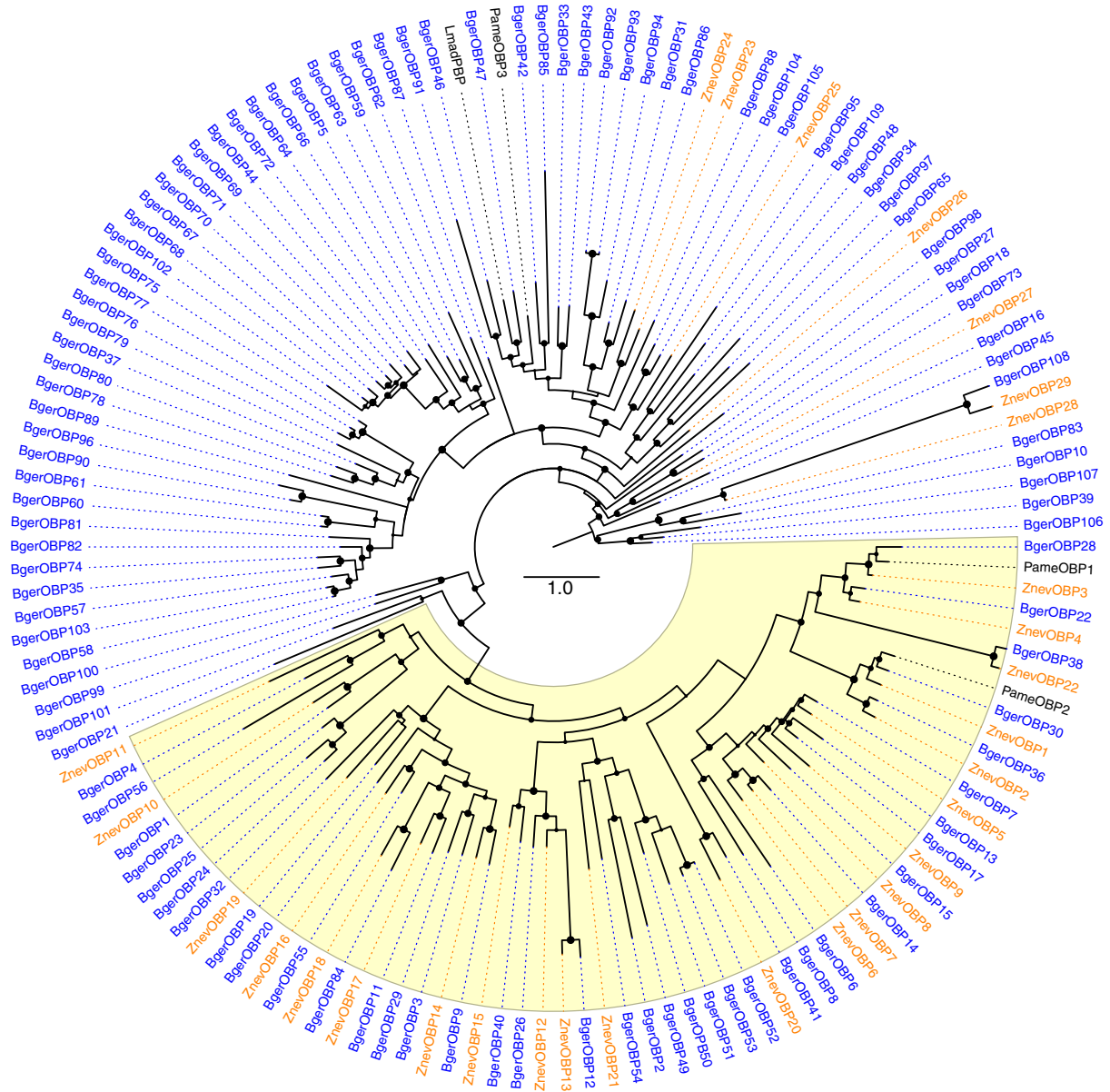
FIGURE 4 Phylogenetic relationships of the OBP family. In the absence of an established outgroup, the tree was rooted at the midpoint. The Classic and Minus-C OBPs are highlighted in yellow, the remaining OBPs being Plus-C and a distinctive clade. Other details are as in Figures 1 and 2 legends, except that here four additional cockroach OBPs from other species are in black [Color figure can be viewed at wileyonlinelibrary.com]

encoding them have no logical order in the genome. The new ones were also added in no particular order, so unlike the chemoreceptor families, consecutive gene names have no meaning for this OBP family, including in a large array of 61 genes spanning approximately 1.6 Mb across four scaffolds (scaffolds1250, 997, 2545, and 116). The newly added OBPs mostly have RNAseq support from head RNAseq and less from the antennal RNAseq of Niu et al. (2016) (Supporting Information Figure 1), largely explaining why they were not included in their compilation. These OBPs might be expressed in gustatory rather than olfactory sensilla, whereas others might not be involved in chemoperception at all. Like OBP genes in other insects, most of these have a phase-0 intron separating the first exon encoding the signal sequence from the rest of the gene, which typically consists of 7–9 short coding exons commonly spanning several kilobase pairs.

Phylogenetic analysis, along with the 29 OBPs from *Z. nevadensis* (Terrapon et al., 2014), three from *P. americana* (Li, He, Zhang, & Dong, 2017), and a pheromone-binding OBP from *Leucophaea* (= *Rhyparobia*) *maderae* (Riviere et al. 2003) reveals that there is a major clade of 71 genes, most of which are in the large 61-gene array in the cockroach, including a 33-gene cockroach-specific expansion. These are all Plus-C OBPs, while the Classic and Minus-C OBPs cluster together (highlighted in yellow in Figure 4). Once again, several gene losses from the termite and almost no termite-specific expansions are evident from the tree. The cockroach genes that cluster with this 71-gene clade, but are not in the 61-gene array, are OBP27 which is a singleton in another large scaffold, OBP85 in another large scaffold with other OBPs, OBP33 and 43 in yet another large separate scaffold along with other OBPs, and OBP39 and 104–107 in yet another large scaffold

(OBP108 is discussed below). These genes have presumably relocated from the large array, as they clearly belong evolutionarily with it. Five of the six ZnevOBPs that cluster with this cockroach OBP expansion in the tree are also in an array spanning 100 kb across two scaffolds (ZnevOBP23–28 in scaffolds1046 and 631), so this expansion is an old one, and there are relatives in other insects not shown in Figure 4. ZnevOBP29, and its cockroach ortholog BgerOBP108, are not only elsewhere in their respective genomes but are far longer than the average OBP at 331 and 324 amino acids, respectively, with a long section of simple sequence between the N-terminal secretion signal sequence and the OBP-homologous C-terminus. Finally, OBP109 is an even more unusual gene encoding a 596 amino acid protein, the C-terminus of which has homology to OBPs, but the N-terminus has no similarity to other proteins in the nonredundant protein database at NCBI, and the termite ortholog appears to have been lost. Like ZnevOBP29 and Bger-OBP108, this gene model is deeply supported by RNAseq from antennae and heads, so its unusual length is real. OBP109 is at one end of the 61-gene array and appropriately clusters phylogenetically with the others from the array.

3.5 | Chemoreceptor pseudogenes

The proportion of pseudogenes in the three chemoreceptor families is fairly high (21.6%, 11.4%, and 43.8% for the OR, GR, and IR families). These large numbers present an opportunity to examine them in more detail, specifically by counting the numbers of obvious pseudogenizing mutations in each gene. This approach previously revealed an excess of "middle-aged" pseudogenes in the OR and GR families of the red harvester ant *Pogonomyrmex barbatus*, something not seen in the honey bee *A. mellifera*, suggesting that it had undergone a major shift in its chemical ecology in the distant past (Smith et al., 2011a). As noted in the methods, in addition to the many pseudogenes named and translated for analysis herein, there were some, especially in the IR family, that despite being near full length were not formally included in the families as they were too damaged to be easily reconstructed. For the purpose of examining the pseudogenes, these were counted as having more than seven pseudogenizing mutations, thus the numbers in the histograms are somewhat higher than the numbers given in the family descriptions above. The distribution of pseudogenes is clearly dominated by those with single mutations, with a clear reduction in numbers of pseudogenes with more mutations (Figure 5). This is the pattern to be expected if older pseudogenes are removed from the genome by deletions, although at 2.2 Gbp, this genome is not that small and hence pseudogenes are probably not removed as quickly as in smaller genomes like that of *D. melanogaster*, but would be removed more quickly than from a large orthopteran genome (Petrov, Sangster, Johnston, Hartl, & Shaw, 2000).

4 | GENERAL DISCUSSION

The connection between complexity of chemosensory repertoire and chemical ecology of an arthropod species is now clear, as it is for vertebrates (Niimura, 2012), and this example of the German cockroach,
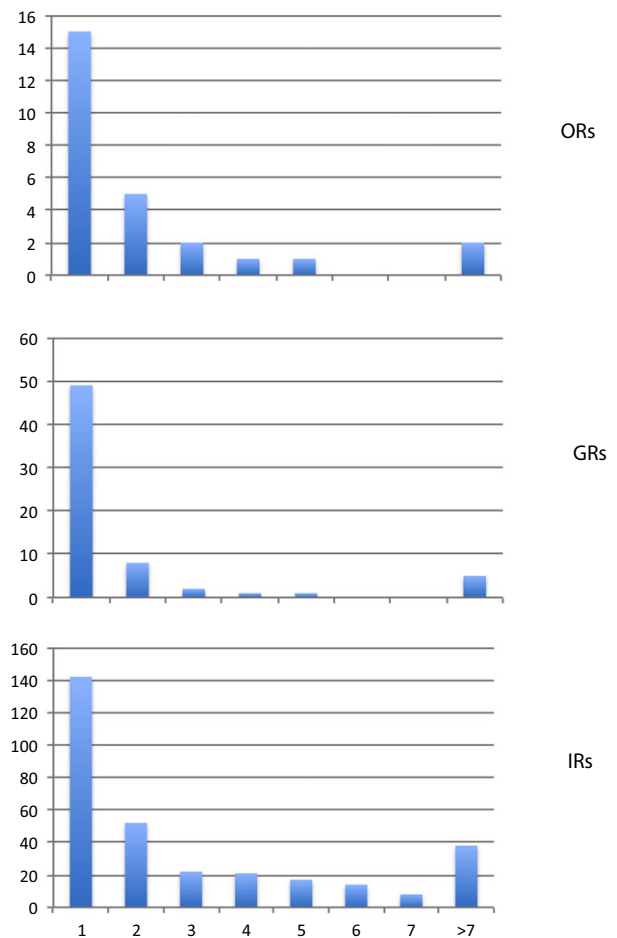

FIGURE 5 Histograms of numbers of pseudogenizing mutations in the three chemoreceptor families. The *x*-axis is the number of pseudogenizing mutations and *y*-axes are the number of pseudogenes in each category [Color figure can be viewed at wileyonlinelibrary.com]

and its comparison with the dampwood termite, demonstrates it abundantly. Termites evolved from cockroaches, and comparison of their genomes has revealed many genomic processes that parallel those of the independent evolution of sociality in Hymenoptera (Harrison et al., 2018). Termites have lost many chemosensory genes that their cockroach ancestors had, presumably because they were no longer necessary for their more specialized ecological niche. In addition, they have only expanded a few chemoreceptor lineages, in particular one of 18 genes in the OR family, which might be involved in their social behavior. In radical contrast, *B. germanica* has lost only a few gene lineages that once were present in the common ancestor with the termite, and has expanded many lineages in dramatic fashion, most spectacularly in the GR and IR families. These major expansions, many of them fairly young as indicated by the short terminal branches to many proteins in the phylogenetic analyses, along with reasonably high percentages of pseudogenes, especially in the IR family, suggests that these chemoreceptor gene families continue to undergo rampant evolution in this cockroach lineage.

Blattodeans are evolutionarily distant from the dipteran and other endopterygotan species for which evidence of ligand specificity is available for at least some chemoreceptors. Hence, it is only

possible to make general inferences about the roles of most of these proteins. In the OR family, the well-conserved Orco protein is present as a single ortholog, as is the case in almost all insects examined to date (the three named Orco proteins in the basal zygentoman *T. domestica* described by Missbach et al. (2014) are likely actually a single Orco and two "specific" ORs (Iaonnidis et al. 2017)). The remaining 123 cockroach ORs are therefore likely "specific" receptors that function as dimers with Orco, but their divergence from all endopterygotan ORs of known ligand specificity precludes any speculation about their functions.

The GR family consists of several subfamilies, the most distinctive and ancient of which is the sugar receptors, here with 14 genes. Although the ligands of these proteins are almost certain to be sugars, the precise ligand specificities of sugar receptors remain unclear even in *D. melanogaster* (e.g., Fujii et al., 2015). In any case, the sugar GRs in this cockroach do not have simple orthologous relationships with those of the flies, having mostly expanded within this lineage in two gene clusters of five and six genes in different large scaffolds, as well as a singleton and pair of genes in other scaffolds. Basal insects have an expanded GR subfamily to which the carbon dioxide receptors of endopterygotes belong, here consisting of 36 genes, however it seems unlikely that all of these are involved in detecting this gas, but rather represent the subfamily from which the carbon dioxide receptors evolved. The remaining GRs in most insects appear to function in perception of "bitter" chemicals, although some in *D. melanogaster* are pheromone and light sensors. These comprise the vast majority of the expanded GR family in this cockroach, including an intronless clade of 322 genes, and are inferred to primarily sense a wide diversity of chemicals in their diverse foods. Somewhat surprisingly, the otherwise well-conserved fructose receptor lineage related to Gr43a in *D. melanogaster* is absent from both these blattodean genomes. This lineage was also not confidently identified in the odonate *C. splendens* (Iaonnidis et al. 2017) so it is possible that it originated from a "bitter" receptor at some point later in insect evolution, although its role in brain nutrient sensing would seem to be an ancient and essential role (Miyamoto et al., 2012).

The IR family has several conserved members whose functions can be assigned with some confidence by comparison with *D. melanogaster*, and are named for their fly orthologs. The Ir8a, 25a, and 76b proteins are coreceptors with other IRs, and involved in diverse aspects of IR function. The Ir21a, 68a, and 93a proteins have recently been demonstrated to mediate perception of temperature and humidity in *D. melanogaster* (Knecht et al., 2016, 2017; Ni et al., 2016), although these blattodeans have lost the fourth gene involved, Ir40a, as it is present in *C. splendens* (Iaonnidis et al. 2017). The cockroach has considerably expanded the Ir41a lineage to 16 intact genes. DmIr41a and 76a are close relatives and cluster with Ir92a, and Ir92a and Ir41a are involved in olfactory perception of amines, in cooperation with Ir76b, at least in the case of Ir41a (Min, Ai, Shin, & Suh, 2013; Hussain et al., 2016). Interestingly Niu et al. (2016), in addition to detecting the coreceptors Ir25a and 76b in their antennal transcriptome, identified three of these Ir41a lineage genes, suggesting that like in *Drosophila* they are olfactory receptors that partner with Ir76b to detect amines. The Ir75 subfamily in *Drosophila* consists of the Ir75a–d, 31a, 64a, and 84a pro-

teins, which along with Ir8a are involved in perception of various acids (Ai et al., 2010, 2013; Grosjean et al., 2011; Gorter et al., 2016; Prieto-Godino et al., 2016; 2017). This subfamily is considerably expanded in this cockroach to 26 genes, all but two intact, which can reasonably be inferred to also partner with Ir8a to sense a diversity of acids. The remaining 850 BgIrs, especially the 755 intronless Ir genes, are likely to encode gustatory receptors, by analogy with the large Ir20a clade of *D. melanogaster*, which are expressed in gustatory tissues (Koh et al., 2014; Stewart et al., 2015), and like the GR expansions presumably sense various chemicals in foods.

Finally, the OBP family is the largest known at 109 genes. This large number, many of which are expressed at low levels in antennae, is appropriate given the huge expansion of GRs and IRs, and because many of these nonantennal OBPs are expressed in head RNAseq where they might serve in gustatory sensilla in support of the numerous gustatory receptors. Li et al. (2017) described three OBPs from *P. americana*, and the first two are close relatives of BgerOBP28 and 30, respectively, and are also expressed in antennae and bind a variety of relevant odorants, so similar roles can be ascribed to these and probably many other of the 48 OBPs Niu et al. (2016) described. Their third OBP is a Plus-C subfamily member, and Li et al. (2017) did not describe a close *B. germanica* relative, but it is related to the 71-gene expansion in *Blattella*, and it is widely expressed in *P. americana* and did not bind any of 87 chemicals tested. Thus, it and some of the *B. germanica* relatives might not be involved in chemosensation. It is, however, also related to the Pheromone Binding Protein of another cockroach, *Leucophaea maderae* (Riviere et al. 2003). The biological roles of OBPs remain unclear as their assumed role of transporting commonly hydrophobic odorants and tastants through the sensillar lymph to the membrane-bound receptors (Leal, 2013; Pelosi et al., 2014) has been challenged by the first direct mutational evidence for the physiological role of an OBP in *D. melanogaster*, where DmOBP28a is not involved in odorant transport, but rather in buffering the responses of the relevant olfactory sensory neurons (Larter, Sun, & Carlson, 2016).

In summary, *B. germanica* has the largest chemosensory repertoire known for any arthropod, including the second-largest GR family, the largest OBP family, and by far the largest IR family. It therefore supports the contention that chemosensory repertoires evolve in concert with the complexity, as well as presumably the details, of the chemical ecology of the species. Complete documentation of these four large gene families will facilitate future work towards identifying the roles of some of these proteins in various aspects of cockroach biology, from their sex and aggregation pheromones to the remarkable evolution of sugar-aversive strains.

CONFLICT OF INTEREST

None.

ORCID

Hugh M. Robertson http://orcid.org/0000-0001-8093-0950

REFERENCES

Ahn, J. E., Chen, Y., & Amrein, H. O. (2017). Molecular basis of fatty acid taste in *Drosophila*. *eLife*, *6*, e30115.

Ai, M., Min, S., Grosjean, Y., Leblanc, C., Bell, R., Benton, R., & Suh, G. S. (2010). Acid sensing by the *Drosophila* olfactory system. *Nature*, *468*(7324), 691–695.

Ai, M., Blais, S., Park, J. Y., Min, S., Neubert, T. A., & Suh, G. S. (2013). Ionotropic glutamate receptors IR64a and IR8a form a functional odorant receptor complex in vivo in *Drosophila*. *Journal of Neuroscience*, *33*(26), 10741–10749.

Andersson, M. N., Corcoran, J. A., Zhang, D. D., Hillbur, Y., Newcomb, R. D., & Löfstedt, C. (2016). A sex pheromone receptor in the Hessian fly *Mayetiola destructor* (Diptera, Cecidomyiidae). *Frontiers in Cellular Neuroscience*, *10*, 212.

Benton, R. (2015). Multigene family evolution: Perspectives from insect chemoreceptors. *Trends in Ecology & Evolution*, *30*(10), 590–600.

Bray, S., & Amrein, H. (2003). A putative *Drosophila* pheromone receptor expressed in male-specific taste neurons is required for efficient courtship. *Neuron*, *39*(6), 1019–1029.

Capella-Gutiérrez, S., Silla-Martínez, J. M., & Gabaldón, T. (2009). trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, *25*(15), 1972–1973.

Carey, A. F., Wang, G., Su, C. Y., Zwiebel, L. J., & Carlson, J. R. (2010). Odorant reception in the malaria mosquito *Anopheles gambiae*. *Nature*, *464*(7285), 66–71.

Chen, Y., & Amrein, H. (2017). Ionotropic receptors mediate *Drosophila* oviposition preference through sour gustatory receptor neurons. *Current Biology*, *27*(18), 2741–2750.

Chen, Y-C. D., & Dahanukar, A. (2017). Molecular and cellular organization of taste neurons in adult *Drosophila* pharynx. *Cell Reports*, *21*, 2978–2991.

Clyne, P. J., Warr, C. G., & Carlson, J. R. (2000). Candidate taste receptors in *Drosophila*. *Science*, *287*(5459), 1830–1834.

Croset, V., Rytz, R., Cummins, S. F., Budd, A., Brawand, D., Kaessmann, H., ... Benton, R. (2010). Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genetics*, *6*(8), e1001064.

Drinnenberg, I. A., Henikoff, S., & Malik, H. S. (2014). Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *eLife*, *3*, e03676.

Eliyahu, D., Nojima, S., Mori, K., & Schal, C. (2008). New contact sex pheromone components of the German cockroach, *Blattella germanica*, predicted from the proposed biosynthetic pathway. *Journal of Chemical Ecology*, *34*, 229–237.

Eyun, S. I., Young Soh, H., Posavi, M., Munro, J. B., Hughes, D. S., Murali, S. C., … Lee, C. E. (2017). Evolutionary history of chemosensory-related gene families across the Arthropoda. *Molecular Biology and Evolution*, *34*(8), 1838–1862.

Forêt, S., & Maleszka, R. (2006). Function and evolution of a gene family encoding odorant binding-like proteins in a social insect, the honey bee (*Apis mellifera*). *Genome Research*, *16*(11), 1404–1413.

Fujii, S., Yavuz, A., Slone, J., Jagge, C., Song, X., & Amrein, H. (2015). *Drosophila* sugar receptors in sweet taste perception, olfaction, and internal nutrient sensing. *Current Biology*, *25*(5), 621–627.

Gondhalekar, A. D., & Scharf, M. E. (2012). Mechanisms underlying fipronil resistance in a multiresistant field strain of the German cockroach (Blattodea: Blattellidae). *Journal of Medical Entomology*, *49*(1), 122–131.

Gore, J. C., & Schal, C. (2007). Cockroach allergen biology and mitigation in the indoor environment. *Annual Review of Entomology*, *52*, 439–463.

Gorter, J. A., Jagadeesh, S., Gahr, C., Boonekamp, J. J., Levine, J. D., & Billeter, J. C. (2016). The nutritional and hedonic value of food modulate sexual receptivity in *Drosophila* melanogaster females. *Scientific Reports*, *6*, 19441.

Gouin, A., Bretaudeau, A., Nam, K., Gimenez, S., Aury, J. M., Duvic, B., … Fournier, P. (2017). Two genomes of highly polyphagous lepidopteran pests (*Spodoptera frugiperda*, Noctuidae) with different host-plant ranges. *Scientific Reports*, *7*(1), 11816.

Grosjean, Y., Rytz, R., Farine, J. P., Abuin, L., Cortot, J., Jefferis, G. S., & Benton, R. (2011). An olfactory receptor for food-derived odours promotes male courtship in *Drosophila*. *Nature*, *478*(7368), 236–240.

Guindon, S., Dufayard, J. F., 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. *Systematic Biology*, *59*(3), 307–321.

Hallem, E. A., Ho, M. G., & Carlson, J. R. (2004). The molecular basis of odor coding in the *Drosophila* antenna. *Cell*, *117*(7), 965–979.

Harrison, M. C., Jongepier, E., Robertson, H. M., Arning, N., Bitard-Feildel, T., Chao, H., … Bornberg-Bauer, E. (2018). Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nature Ecology and Evolution*, *2*, 557–566.

Hickner, P. V., Rivaldi, C. L., Johnson, C. M., Siddappaji, M., Raster, G. J., & Syed, Z. (2016). The making of a pest: Insights from the evolution of chemosensory receptor families in a pestiferous and invasive fly, *Drosophila suzukii*. *BMC Genomics*, *17*(1), 648.

Hoy, M. A., Waterhouse, R. M., Wu, K., Estep, A. S., Ioannidis, P., Palmer, W. J., … Murphy, T. D. (2016). Genome sequencing of the phytoseiid predatory mite Metaseiulus occidentalis reveals completely atomized Hox genes and superdynamic intron evolution. *Genome Biology and Evolution*, *8*(6), 1762–1775.

Hussain, A., Zhang, M., Üçpunar, H. K., Svensson, T., Quillery, E., Gompel, N., … Kadow, I. C. G. (2016). Ionotropic chemosensory receptors mediate the taste and smell of polyamines. *PLoS Biology*, *14*(5), e1002454.

i5K Consortium (2013). The i5K Initiative: Advancing arthropod genomics for knowledge, human health, agriculture, and the environment. *Journal of Heredity*, *104*(5), 595–600.

Ioannidis, P., Simao, F. A., Waterhouse, R. M., Manni, M., Seppey, M., Robertson, H. M., … Zdobnov, E. M. (2017). Genomic features of the damselfly *Calopteryx splendens* representing a sister clade to most insect orders. *Genome Biology and Evolution*, *9*(2), 415–430.

Joseph, R. M., & Carlson, J. R. (2015). *Drosophila* chemoreceptors: A molecular interface between the chemical world and the brain. *Trends in Genetics*, *31*(12), 683–695.

Josek, J., Walden, K. K. O., Allan, B. F., Alleyne, M., & Robertson, H. M. (2018). A foreleg transcriptome for Ixodes scapularis ticks: candidates for chemoreceptors and binding proteins that might be expressed in the sensory Haller's organ. Ticks and Tick-borne Diseases, in press.

Kirkness, E. F., Haas, B. J., Sun, W., Braig, H. R., Perotti, M. A., Clark, J. M., … Pittendrigh, B. R. (2010). Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(27), 12168–12173.

Knecht, Z. A., Silbering, A. F., Ni, L., Klein, M., Budelli, G., Bell, R., ... Garrity, P. A. (2016). Distinct combinations of variant ionotropic glutamate receptors mediate thermosensation and hygrosensation in *Drosophila*. *eLife*, *5*, e17879.

Knecht, Z. A., Silbering, A. F., Cruz, J., Yang, L., Croset, V., Benton, R., & Garrity, P. A. (2017). Ionotropic Receptor-dependent moist and dry cells control hygrosensation in *Drosophila*. *eLife*, *6*, e26654.

Koh, T. W., He, Z., Gorur-Shandilya, S., Menuz, K., Larter, N. K., Stewart, S., & Carlson, J. R. (2014). The *Drosophila* IR20a clade of ionotropic receptors are candidate taste and pheromone receptors. *Neuron*, *83*(4), 850–865.

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., … Higgins, D. J. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947–2948.

Larter, N. K., Sun, J. S., & Carlson, J. R. (2016). Organization and function of *Drosophila* odorant binding proteins. *eLife*, *5*, e20242.

Leal, W. S. (2013). Odorant reception in insects: Roles of receptors, binding proteins, and degrading enzymes. *Annual Review of Entomology*, *58*, 373–391.

Leary, G. P., Allen, J. E., Bunger, P. L., Luginbill, J. B., Linn, C. E., Macallister, I. E., … Wanner, K. W. (2012). Single mutation to a sex pheromone receptor provides adaptive specificity between closely related moth species. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(35), 14081–14086.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., … Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Li, Z. Q., He, P., Zhang, Y. N., & Dong, S. L. (2017). Molecular and functional characterization of three odorant-binding protein from *Periplaneta americana*. *PloS One*, *12*(1), e0170072.

Matthews, B. J., McBride, C. S., DeGennaro, M., Despo, O., & Vosshall, L. B. (2016). The neurotranscriptome of the *Aedes aegypti* mosquito. *BMC Genomics*, *17*(1). 32.

Matthews, B. J., Dudchenko, O., Kingan, S., Koren, S., Antoshechkin, I., Crawford, J. E., … Vosshall, L. B. (2017). Improved *Aedes aegypti* mosquito reference genome assembly enables biological discovery and vector control. *Nature*, in press.

McBride, C. S. (2007). Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(12), 4996–5001.

McBride, C. S., & Arguello, J. R. (2007). Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics*, *177*(3), 1395–1416.

McKenna, D. D., Scully, E. D., Pauchet, Y., Hoover, K., Kirsch, R., Geib, S. M., … Richards, S. (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 Biology*, *17*(1), 227.

Min, S., Ai, M., Shin, S. A., & Suh, G. S. (2013). Dedicated olfactory neurons mediating attraction behavior to ammonia and amines in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(14), E1321–E1329.

Missbach, C., Dweck, H. K., Vogel, H., Vilcinskas, A., Stensmyr, M. C., Hansson, B. S., & Grosse-Wilde, E. (2014). Evolution of insect olfactory receptors. *eLife*, *3*, e02115.

Missbach, C., Vogel, H., Hansson, B. S., & Groβe-Wilde, E. (2015). Identification of odorant binding proteins and chemosensory proteins in antennal transcriptomes of the jumping bristletail *Lepismachilis y-signata* and the firebrat *Thermobia domestica*: Evidence for an independent OBP–OR origin. *Chemical Senses*, *40*(9), 615–626.

Miyamoto, T., Slone, J., Song, X., & Amrein, H. (2012). A fructose receptor functions as a nutrient sensor in the *Drosophila* brain. *Cell*, *151*(5), 1113–1125.

Ngoc, P. C. T., Greenhalgh, R., Dermauw, W., Rombauts, S., Bajda, S., Zhurov, V., … Clark, R. M. (2016). Complex evolutionary dynamics of massively expanded chemosensory receptor families in an extreme generalist chelicerate herbivore. *Genome Biology and Evolution*, *8*(11), 3323–3339.

Ni, L., Bronk, P., Chang, E. C., Lowell, A. M., Flam, J. O., Panzano, V. C., … Garrity, P. A. (2013). A gustatory receptor paralogue controls rapid warmth avoidance in *Drosophila*. *Nature*, *500*(7464), 580–584.

Ni, L., Klein, M., Svec, K. V., Budelli, G., Chang, E. C., Ferrer, A. J., … Garrity, P. A. (2016). The ionotropic receptors IR21a and IR25a mediate cool sensing in *Drosophila*. *eLife*, *5*, e13254.

Niimura, Y. (2012). Olfactory receptor multigene family in vertebrates: From the viewpoint of evolutionary genomics. *Current Genomics*, *13*(2), 103–114.

Niu, D. J., Liu, Y., Dong, X. T., & Dong, S. L. (2016). Transcriptome based identification and tissue expression profiles of chemosensory genes in *Blattella germanica* (Blattaria: Blattidae). *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics*, *18*, 30–43.

Nojima, S., Schal, C., Webster, F. X., Santangelo, R. G., & Roelofs, W. L. (2005). Identification of the sex pheromone of the German cockroach, *Blattella germanica*. *Science*, *307*(5712), 1104–1106.

Obiero, G. F., Mireji, P. O., Nyanjom, S. R., Christoffels, A., Robertson, H. M., & Masiga, D. K. (2014). Odorant and gustatory receptors in the tsetse fly *Glossina morsitans morsitans*. *PLoS Neglected Tropical Diseases*, *8*(4), e2663.

Panfilio, K. A., Jentzsch, I. M. V., Benoit, J. B., Erezyilmaz, D., Suzuki, Y., Colella, S., … Richards, S. (2017). Molecular evolutionary trends and feeding ecology diversification in the Hemiptera, anchored by the milkweed bug genome. *bioRxiv*, 201731.

Pask, G. M., Slone, J. D., Millar, J. G., Das, P., Moreira, J. A., Zhou, X., … Ray, A. (2017). Specialized odorant receptors in social insects that detect cuticular hydrocarbon cues and candidate pheromones. *Nature Communications*, *8*(1), 297.

Pelosi, P., Iovinella, I., Felicioli, A., & Dani, F. R. (2014). Soluble proteins of chemical communication: An overview across arthropods. *Frontiers in Physiology*, *5*, e320.

Peñalva-Arana, D. C., Lynch, M., & Robertson, H. M. (2009). The chemoreceptor genes of the waterflea *Daphnia pulex*: Many Grs but no Ors. *BMC Evolutionary Biology*, *9*(1), 79.

Petrov, D. A., Sangster, T. A., Johnston, J. S., Hartl, D. L., & Shaw, K. L. (2000). Evidence for DNA loss as a determinant of genome size. *Science*, *287*(5455), 1060–1062.

Poelchau, M., Childers, C., Moore, G., Tsavatapalli, V., Evans, J., Lee, C. Y., … Hackett, K. (2014). The i5k Workspace@ NAL—Enabling genomic data access, visualization and curation of arthropod genomes. *Nucleic Acids Research*, *43*(D1), D714–D719.

Prieto-Godino, L. L., Rytz, R., Bargeton, B., Abuin, L., Arguello, J. R., Dal Peraro, M., & Benton, R. (2016). Olfactory receptor pseudo-pseudogenes. *Nature*, *539*(7627), 93–97.

Prieto-Godino, L. L., Rytz, R., Cruchet, S., Bargeton, B., Abuin, L., Silbering, A. F., … Benton, R. (2017). Evolution of acid-sensing olfactory circuits in drosophilids. *Neuron*, *93*(3), 661–676.

Rabito, F. A., Carlson, J. C., He, H., Werthmann, D., & Schal, C. (2017). A single intervention for cockroach control reduces cockroach exposure and asthma morbidity in children. *Journal of Allergy and Clinical Immunology*, *140*, 565–570.

Renthal, R., Manghnani, L., Bernal, S., Qu, Y., Griffith, W. P., Lohmeyer, K., … León, A. (2017). The chemosensory appendage proteome of *Amblyomma americanum* (Acari: Ixodidae) reveals putative odorant-binding and other chemoreception-related proteins. *Insect Science*, *24*(5), 730–742.

Richards, S., Gibbs, R. A., Weinstock, G. M., Brown, S. J., Denell, R., Beeman, R. W., … Bucher, G. (2008). The genome of the model beetle and pest *Tribolium castaneum*. *Nature*, *452*(7190), 949–955.

Rimal, S., & Lee, Y. (2018). The multidimensional ionotropic receptors of *Drosophila melanogaster*. *Insect Molecular Biology*, *27*(1), 1–7.

Rivière, S., Lartigue, A., Quennedey, B., Campanacci, V., Farine, J. P., Tegoni, M., … Brossut, R. (2003). A pheromone-binding protein from the cockroach *Leucophaea maderae*: Cloning, expression and pheromone binding. *Biochemical Journal*, *371*(2), 573–579.

Robertson, H. M. (2015). The insect chemoreceptor superfamily is ancient in animals. *Chemical Senses*, *40*(9), 609–614.

Robertson, H. M., & Wanner, K. W. (2006). The chemoreceptor superfamily in the honey bee, *Apis mellifera*: Expansion of the odorant, but not gustatory, receptor family. *Genome Research*, *16*(11), 1395–1403.

Robertson, H. M., Warr, C. G., & Carlson, J. R. (2003). Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(suppl 2), 14537–14542.

Robinson, G. E., Hackett, K. J., Purcell-Miramontes, M., Brown, S. J., Evans, J. D., Goldsmith, M. R., … Schneider, D. J. (2011). Creating a buzz about insect genomes. *Science*, *331*(6023), 1386–1386.

Rytz, R., Croset, V., & Benton, R. (2013). Ionotropic receptors (IRs): Chemosensory ionotropic glutamate receptors in *Drosophila* and beyond. *Insect Biochemistry and Molecular Biology*, *43*(9), 888–897.

Saina, M., Busengdal, H., Sinigaglia, C., Petrone, L., Oliveri, P., Rentzsch, F., & Benton, R. (2015). A cnidarian homologue of an insect gustatory receptor functions in developmental body patterning. *Nature Communications*, *6*, 6243. https://doi.org/10.1038/ncomms7243

Schal, C. (2011). Cockroaches. In S. Hedges & D. Moreland (Eds.), *Handbook of Pest Control* (pp. 150–291). Richfield, OH: GIE Media.

Schal, C., Gautier, J. Y., & Bell, W. J. (1984). Behavioural ecology of cockroaches. *Biological Reviews*, *59*(2), 209–254.

Schal, C., & Hamilton, R. L. (1990). Integrated suppression of synanthropic cockroaches. *Annual Review of Entomology*, *35*(1), 521–551.

Schoville, S. D., Chen, Y. H., Andersson, M. N., Benoit, J. B., Bhandari, A., Bowsher, J. H., … Richards, S. (2017). A model species for agricultural pest genomics: The genome of the Colorado potato beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae). *bioRxiv*, 192641.

Scott, K., Brady, R., Cravchik, A., Morozov, P., Rzhetsky, A., Zuker, C., & Axel, R. (2001). A chemosensory gene family encoding candidate gustatory and olfactory receptors in *Drosophila*. *Cell*, *104*(5), 661–673.

Silbering, A. F., Rytz, R., Grosjean, Y., Abuin, L., Ramdya, P., Jefferis, G. S., & Benton, R. (2011). Complementary function and integrated wiring of the evolutionarily distinct *Drosophila* olfactory subsystems. *Journal of Neuroscience*, *31*(38), 13357–13375.

Silverman, J., & Bieman, D. N. (1993). Glucose aversion in the German cockroach, *Blattella germanica*. *Journal of Insect Physiology*, *39*(11), 925–933.

Silverman, J., & Ross, M. H. (1994). Behavioral resistance of field-collected German cockroaches (Blattodea: Blattellidae) to baits containing glucose. *Environmental Entomology*, *23*(2), 425–430.

Smith, C. R., Smith, C. D., Robertson, H. M., Helmkampf, M., Zimin, A., Yandell, M., … Gadau, J. (2011a). Draft genome of the red harvester ant *Pogonomyrmex barbatus*. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5667–5672.

Smith, C. D., Zimin, A., Holt, C., Abouheif, E., Benton, R., Cash, E., … Tsutsui, N. D. (2011b). Draft genome of the globally widespread and invasive Argentine ant (*Linepithema humile*). *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5673–5678.

Stewart, S., Koh, T. W., Ghosh, A. C., & Carlson, J. R. (2015). Candidate ionotropic taste receptors in the *Drosophila* larva. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(14), 4195–4201.

Terrapon, N., Li, C., Robertson, H. M., Ji, L., Meng, X., Booth, W., … Liebig, J. (2014). Molecular traces of alternative social organization in a termite genome. *Nature Communications*, *5*, 3636.

Thorne, N., & Amrein, H. (2008). Atypical expression of *Drosophila* gustatory receptor genes in sensory and central neurons. *Journal of Comparative Neurology*, *506*(4), 548–568.

Vieira, F. G., Forêt, S., He, X., Rozas, J., Field, L. M., & Zhou, J. J. (2012). Unique features of odorant-binding proteins of the parasitoid wasp *Nasonia vitripennis* revealed by genome annotation and comparative analyses. *PLoS One*, *7*(8), e43034.

Vizueta, J., Frías-López, C., Macías-Hernández, N., Arnedo, M. A., Sánchez-Gracia, A., & Rozas, J. (2017). Evolution of chemosensory gene families in arthropods: Insight from the first inclusive comparative transcriptome analysis across spider appendages. *Genome Biology and Evolution*, *9*(1), 178–196.

Wada-Katsumata, A., Ozaki, M., Yokohari, F., Nishikawa, M., & Nishida, R. (2009). Behavioral and electrophysiological studies on the sexually biased synergism between oligosaccharides and phospholipids in gustatory perception of nuptial secretion by the German cockroach. *Journal of Insect Physiology*, *55*, 742–750.

Wada-Katsumata, A., Silverman, J., & Schal, C. (2011). Differential inputs from chemosensory appendages mediate feeding responses to glucose in wild-type and glucose-averse German cockroaches, *Blattella germanica*. *Chemical Senses*, *36*(7), 589–600.

Wada-Katsumata, A., Silverman, J., & Schal, C. (2013). Changes in taste neurons support the emergence of an adaptive behavior in cockroaches. *Science*, *340*(6135), 972–975.

Wada-Katsumata, A., Zurek, L., Nalyanya, G., Roelofs, W. L., Zhang, A., & Schal, C. (2015). Gut bacteria mediate aggregation in the German cockroach. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(51), 15678–15683.

Wang, G., Carey, A. F., Carlson, J. R., & Zwiebel, L. J. (2010). Molecular basis of odor coding in the malaria vector mosquito *Anopheles gambiae*. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(9), 4418–4423.

Wanner, K. W., Nichols, A. S., Walden, K. K., Brockmann, A., Luetje, C. W., & Robertson, H. M. (2007). A honey bee odorant receptor for the queen substance 9-oxo-2-decenoic acid. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(36), 14383–14388.

Watanabe, K., Toba, G., Koganezawa, M., & Yamamoto, D. (2011). Gr39a, a highly diversified gustatory receptor in *Drosophila*, has a role in sexual behavior. *Behavior Genetics*, *41*(5), 746–753.

Weiss, L. A., Dahanukar, A., Kwon, J. Y., Banerjee, D., & Carlson, J. R. (2011). The molecular and cellular basis of bitter taste in *Drosophila*. *Neuron*, *69*(2), 258–272.

Xiao, J. H., Yue, Z., Jia, L. Y., Yang, X. H., Niu, L. H., Wang, Z., ... Huang, D. W. (2013). Obligate mutualism within a host drives the extreme specialization of a fig wasp genome. *Genome Biology, 14*(12), R141.

Zhou, X., Slone, J. D., Rokas, A., Berger, S. L., Liebig, J., Ray, A., ... Zwiebel, L. J. (2012). Phylogenetic and transcriptomic analysis of chemosensory receptors in a pair of divergent ant species reveals sex-specific signatures of odor coding. *PLoS Genetics, 8*(8), e1002930.

Zhou, X., Qian, K., Tong, Y., Zhu, J. J., Qiu, X., & Zeng, X. (2014). De novo transcriptome of the hemimetabolous German cockroach (*Blattella germanica*). *PLoS One, 9*(9), e106932.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.