Research article

Evolution of the sugar receptors in insects

Lauren B Kent and Hugh M Robertson*

Address: Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Email: Lauren B Kent - lkent@life.uiuc.edu; Hugh M Robertson* - hughrobe@life.uiuc.edu

* Corresponding author

Abstract

Background: Perception of sugars is an invaluable ability for insects which often derive quickly accessible energy from these molecules. A distinctive subfamily of eight proteins within the gustatory receptor (Gr) family has been identified as sugar receptors (SRs) in *Drosophila melanogaster* (Gr5a, Gr61a, and Gr64a-f). We examined the evolution of these SRs within the 12 available Drosophila genome sequences, as well as three mosquito, two moth, and beetle, bee, and wasp genome sequences.

Results: While most Drosophila species retain all eight genes, we find that the three Drosophila subgenus species have lost Gr64d, while *D. grimshawi* and the *D. pseudoobscura/persimilis* sibling species have also lost Gr5a function. The entire Gr64 gene complex was also duplicated in the *D. grimshawi* lineage, but only one potentially functional copy of each gene has been retained. The numbers of SRs range from two in the hymenopterans *Apis mellifera* and *Nasonia vitripennis* to 16 in the beetle *Tribolium castaneum*. An unusual aspect is the evolution of a novel exon from intronic sequence in an expanded set of four SRs in *Bombyx mori* (BmGr5-8), which appears to be the first example of such exonization in insects. Twelve intron gains and 63 losses are inferred within the SR family.

Conclusion: Examination of the SRs in these fly, mosquito, moth, beetle, and hymenopteran genome sequences reveals that they appear to have originated independently from single ancestral genes within the dipteran and coleopteran lineages, and two genes in the lepidopteran and hymenopteran lineages. The origin of the insect SRs will eventually be illuminated by additional basal insect and arthropod genome sequences.

Background

Sugars serve as some of the simplest, most easily metabolized forms of energy available to life. For example, despite an anautogenous female mosquito's need for a bloodmeal to nourish her developing eggs, it is the simple nectar of plants that fuels her flight muscles and daily energy needs. As sugar is a valuable resource, it seems fitting that most animals have the ability to taste sugars, and in many it forms a primary stimulatory signal for feeding.

The molecular basis for sugar detection in insects has been revealed in *Drosophila melanogaster* where it involves a series of at least eight genes in the gustatory receptor (Gr) family [1-3]. The first of these is Gr5a on the X chromosome, although identification of this gene as encoding a trehalose receptor was initially confused with the neighboring *Tre* locus [4]. In phylogenetic analyses, Gr5a clusters with seven other genes on the third chromosome, including the singleton Gr61a and Gr64a-f: six genes in a

tandem array [5], making all of these candidate sugar receptors (SRs). Recent work with these SRs has started to unravel their involvement in sugar detection, although much work remains to understand how these flies perceive sugars. Thorne et al. [6] and Wang et al. [7] showed that Gr5a is expressed widely in sensory neurons that detect sugars. Subsequently, Jiao et al. [8] showed that Gr5a-expressing cells also express undefined combinations of the other seven genes, and showed that Gr64a is required for sensing several sugars other than trehalose. Dahanukar et al. [9] showed that Gr61a and Gr64f are co-expressed with Gr5a in some but not all sugar-sensitive neurons, indicating that there is a complicated pattern of co-expression of these eight genes. Furthermore, they generated double-mutant flies for both Gr5a and Gr64a that cannot taste any sugars, suggesting that these two receptors co-function with the other six to achieve detection of sugars. Meanwhile, Slone et al. [10] generated a deletion mutant removing Gr64a-f and found that these flies could not detect most sugars, including trehalose, which is supposed to be detected by Gr5a. Together the evidence from these studies affirms that these eight proteins constitute the SRs in flies, and strongly suggests that they function as heterodimers, perhaps with Gr5a and Gr64a pairing with each other and/or the other less widely-expressed Gr61a and Gr64b-f. Many issues remain unresolved, including the exact ligand specificities of each heterodimeric pair of these eight SRs. Here we contribute to our understanding of these fly SRs by examining their evolution in the 11 newly available Drosophila species genomes [11], as well as more distant comparisons with the three available mosquito genomes, and the available moth, beetle, bee, and wasp genomes. This analysis reveals an unexpected history of expansion of these gene subfamilies from only one or two genes in each insect order, as well as such unusual features as evolution of a novel exon in a lineage of moth SRs.

Methods

Homologs of the eight SR genes in *D. melanogaster* were identified in the 11 newly available Drosophila species genome sequences using the assemblies available in Fly-Base as of October 2007, which are those employed in the genome paper [11], except that the *D. simulans* assembly is the "merged" assembly of six different strains. TBLASTN searches were employed to identify these genes, and gene models were constructed using the DmGrs as templates in the text editor of PAUP* v4 [12]. The *D. simulans* assembly available at FlyBase has numerous problems, including unexplained single base indels relative to the raw traces available in the Trace Archive at the National Center for Biotechnology Information (NCBI). Such errors were present in most of the genes and were corrected.

The mosquito *Anopheles gambiae* and *Aedes aegypti* gene models are from Hill et al. [13] and Kent et al. [14], but updated in light of gene models constructed for *Culex pipiens* using the CpipJ1 assembly available at VectorBase, the NCBI, the Broad Institute, and the J. Craig Venter Institute (JCVI). The *Bombyx mori* moth gene models are from Wanner and Robertson [15], while those for the red flour beetle *Tribolium castaneum* were constructed by HMR for the main genome publication [16]. The two honey bee *Apis mellifera* SRs are from Robertson and Wanner [17] and their homologs in the parasitoid wasp *Nasonia vitripennis* were built from the v1.0 assembly available from the Human Genome Sequencing Center at the Baylor College of Medicine and NCBI. The complete set of SRs is provided in a supplementary online FASTA file (Additional file 1).

All proteins were aligned using the multiple alignment program CLUSTALX with default settings [18]. The alignments were used to detect potential problems with the gene models, which were then refined. Phylogenetic analysis was performed using corrected distances, as well as supporting maximum parsimony and maximum likelihood analysis, as described in Robertson et al. [5], Robertson and Wanner [17] and Kent et al. [14]. Intron locations and phases were mapped to the protein alignment manually in the PAUP text editor and then mapped to branches in the phylogenetic tree using Dollo parsimony assuming that intron gains are unique but losses are independent events.

Results

The Drosophila SRs

For the most part the 12 available Drosophila genome sequences contain single intact orthologs for each of the eight Drosophila SR lineages (Figure 1). The only previously known exception is that Gr5a is missing from *D. pseudoobscura* [9,19] and, not surprisingly, its sibling species *D. persimilis*. There are, however, several other instances of gene subfamily evolution within this fly genus. Gr64e appears to be a pseudogene in both of these species because the intron donor splice site on the penultimate intron starts with GA instead of the canonical GT. Gr5a is a severely damaged pseudogene in the Hawaiian *D. grimshawi*, and is not included in the tree analysis. In addition, there was a duplication of the entire 3rd chromosome gene complex in *D. grimshawi*, roughly 2.6 Mbp apart, followed by the loss or pseudogenization of each gene in one or the other version of the complex, leaving a single intact copy of each gene (Figure 2). Thus the centromeric complex retains a functional copy of Gr61a, a pseudogenic copy of Gr64a, and functional copies of Gr64b, Gr64c, and Gr64e, followed by a fragment of Gr64f, while the telomeric complex has an intact copy of Gr64a and Gr64f, and fragments of Gr61a, Gr64b, Gr64c, and Gr64e. We designate genes in the centromeric complex by the number "1" after their name and the telomeric complex by the number "2". In addition, Gr64d is missing from the
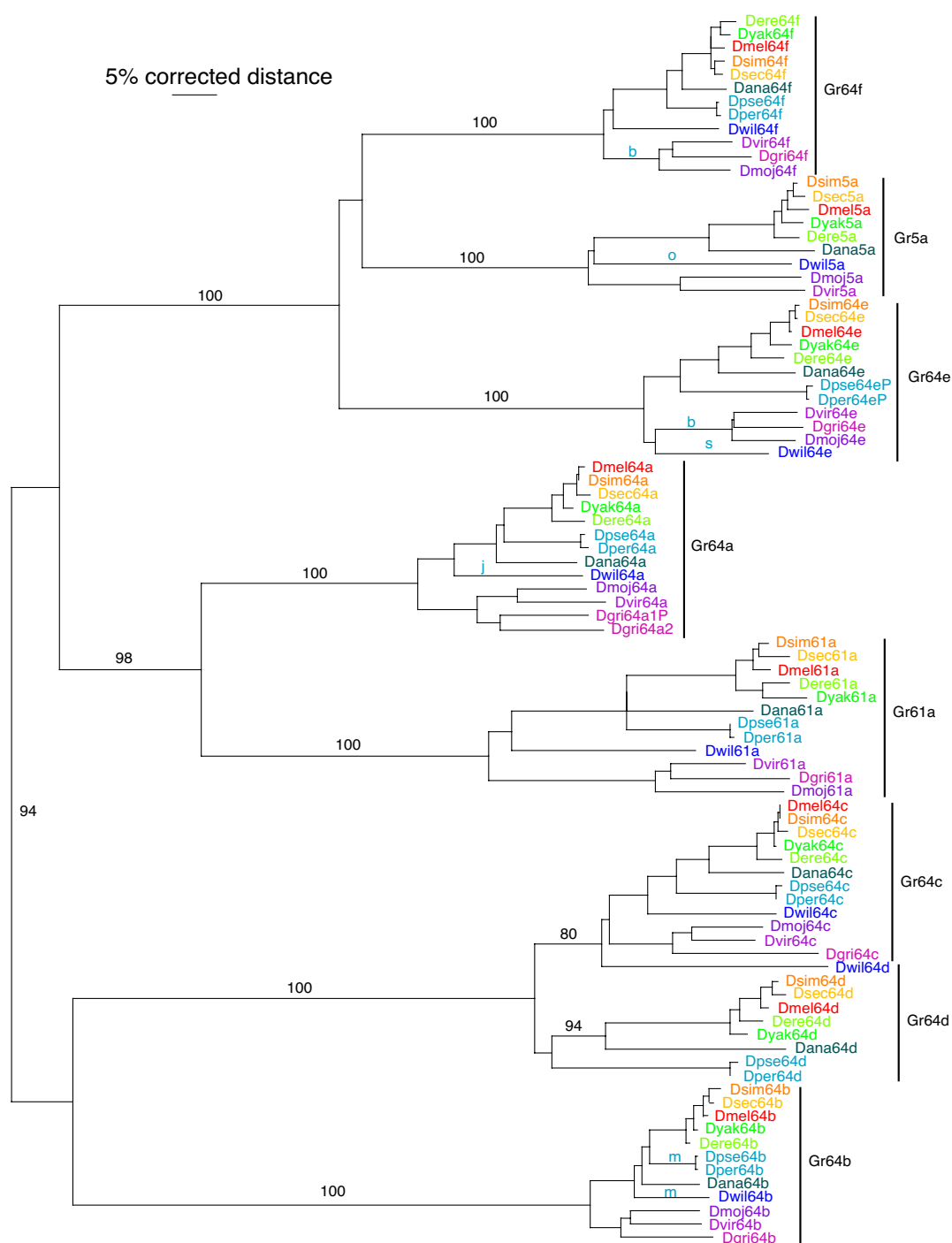
Figure 1
Phylogenetic relationships of eight SRs in 12 Drosophila species. This is a corrected distance tree based on aligned amino acids and rooted at the midpoint. The variable length and sequence N-termini upstream of the conserved "pre-peak" region (Figures 7, 8, 9, 10, 11, 12 and 13) were removed from the alignment. Species names are abbreviated to the genus plus the first three letters of the species name, and are color-coded. The eight proteins are indicated on the right. Bootstrap support from 1000 replications of uncorrected distance analysis is shown for the major branches. Most orthologous relationships are in accordance with the known species relationships, or near enough, and bootstrap support is not shown for them. Intron losses are shown in lower case light blue letters above the branches to which they map, but only if not shown in Figure 3.
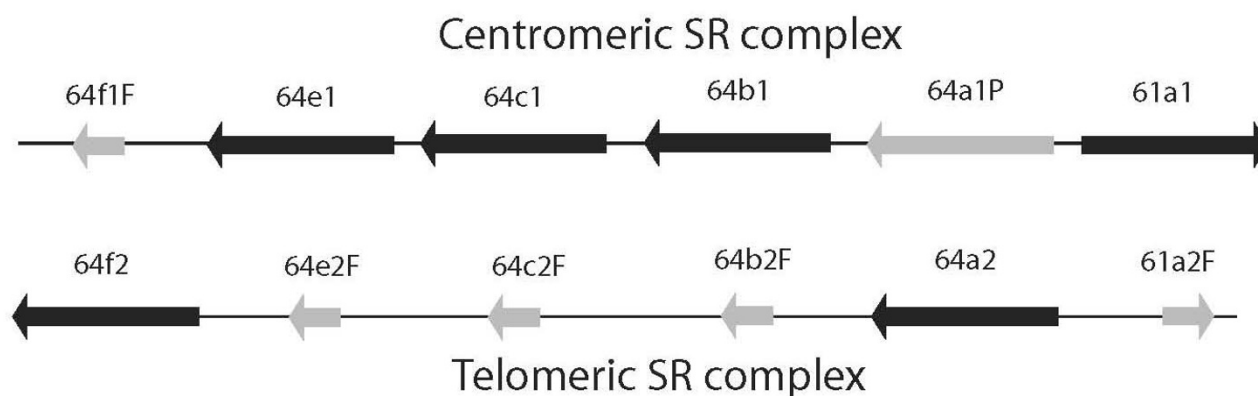
Figure 2
Schematic diagram of the two SR gene complexes in *D. grimshawi*. The two complexes and their genes are labeled 1 and 2 for the centromeric and telomeric complexes, with intact genes shown as solid boxes, pseudogenes as grey boxes, and pseudogenic gene fragments as short grey boxes. Direction of transcription is shown by arrow heads. Paralogy is indicated by dashed lines. Not to scale.

three Drosophila subgenus species, *D. virilis*, *D. mojavensis*, and *D. grimshawi*, so this loss predates the duplication of the complex in the *D. grimshawi* lineage. Judging from the branch lengths of the DgriGr64a1P/2 copies, this gene complex duplication is relatively old and may be present in all Hawaiian Drosophila. We have not determined how extensive the duplication is, but it presumably involves multiple flanking genes as well. Another slightly unusual problem is the phylogenetic placement of what we are calling Gr64d in *D. willistoni*. This gene is in the expected location for Gr64d, that is between Gr64c and Gr64e, however phylogenetically it is clearly closer to the Gr64c genes than the Gr64d genes. There is no simple explanation for this situation. A duplication of Gr64c in *D. willistoni* followed by loss of the original Gr64d gene should lead to our Gr64d clustering with the *D. willistoni* Gr64c, and there is no evidence of a partial gene conversion event. We are also able to date roughly the movement of Gr5a and Gr61a from the tandem complex of Gr64a-f. All Drosophila species appear to have Gr5a on their X chromosomes, so this gene relocation predates the genus. However, Gr61a is located in inverse orientation at the 5' end of the complex, in all species up to *D. ananassae*, so it must have relocated thereafter. Finally, Gr61a is relocated to the X chromosome in *D. yakuba*. The result is that the number of apparently intact SRs in these 12 Drosophila species is six in *D. pseudoobscura/persimilis* and *D. grimshawi*, seven in *D. virilis* and *D. mojavensis*, and eight in the remainder of the species.

Our analysis of the Drosophila Grs differs somewhat from that recently reported in Gardiner et al. [20], primarily in that they ignore the Gr5a pseudogene in *D. grimshawi*, and list multiple copies of Gr61a (5 genes and 3 pseudo-

genes), Gr64a (4/3), and Gr64b (2/1) in this species beyond what we include (Figure 2). This may be because they used early assemblies from January 2006, which for this species might have had multiple haplotypes alternatively assembled. Remnants of these remain in the October 2007 assemblies as short contigs and were not included in our analysis.

The mosquito SRs
We find that *An. gambiae* and *Ae. aegypti* have eight and seven functional SRs, respectively. In addition, *Ae. aegypti* also has three pseudogenes. [13,14]. We examined the newly available *Culex pipiens* genome sequence and find that this mosquito, which is ~50 Myr diverged from *Ae. aegypti* [21,22], has 14 sugar receptor genes, of which one is a pseudogene (Figure 3). Five of these extra receptor genes are the result of relatively recent duplications within the *C. pipiens* lineage. Phylogenetic analysis reveals, however, that what we had earlier considered to be an orthologous, albeit rather divergent, relationship of AgGr19 and AaGr13P [14], is in fact a paralogous comparison, because *C. pipiens* has clear orthologs of each of these receptors. Thus *An. gambiae* has lost the ortholog of AaGr13P/ CpGr16/17, while *Ae. aegypti* has lost the ortholog of AgGr19/CpGr6/7. We therefore infer that most mosquito lineages have nine SR gene lineages, although not all are present and intact in all species, and some have been duplicated in some species.

Evolution of the fly SRs
As noted in Kent et al. [14], the relationships of the eight Drosophila SR lineages and the then-eight and now-nine mosquito SRs is not one of simple orthology, but rather a complex pattern of gene duplications and losses, presum-
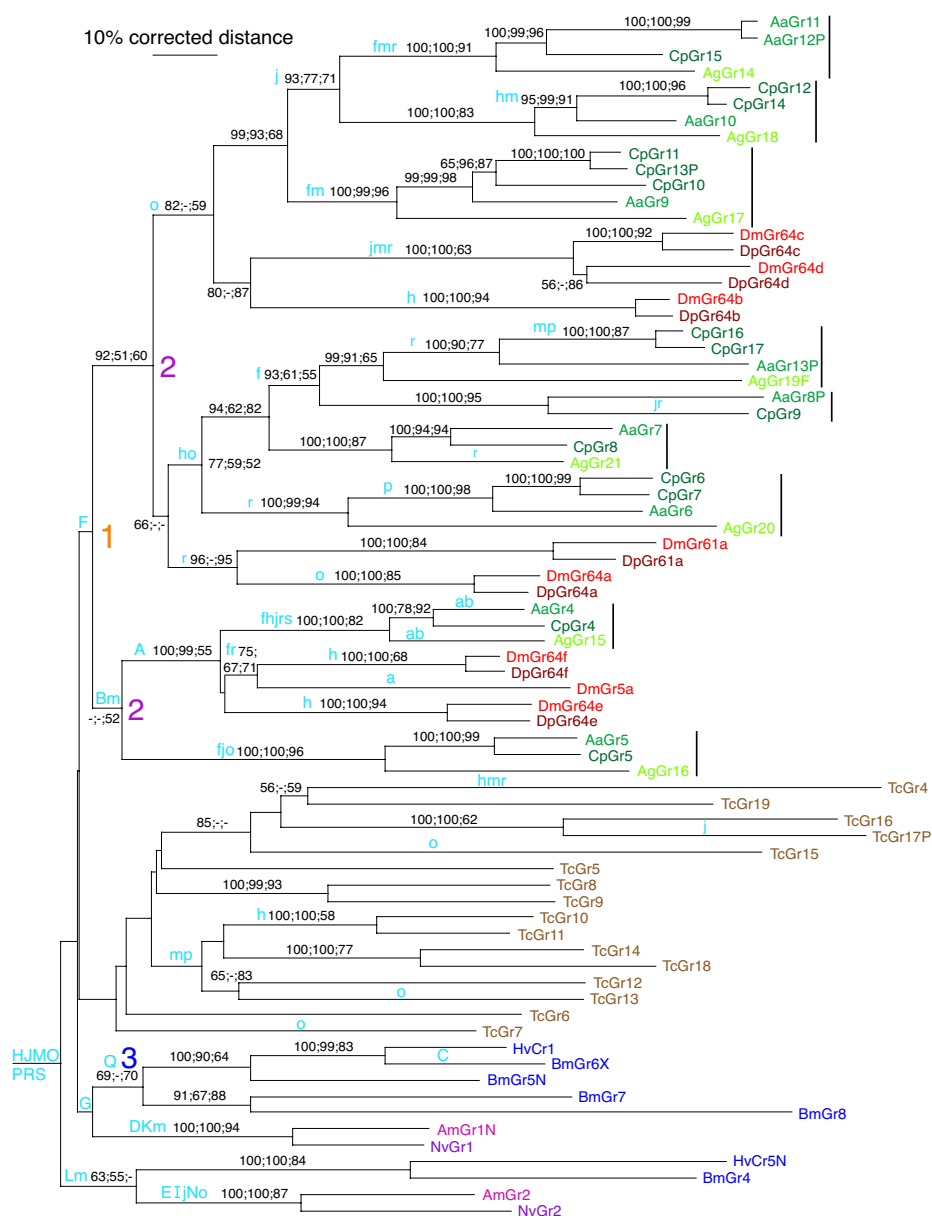
Figure 3
Phylogenetic relationships of SRs in insects. This corrected distance tree was rooted with the AmGr2/NvGr2 and HvCr5/BmGr4 lineage as the outgroup, based on their apparent basal clustering in larger analyses of the entire Gr family. Bootstrap support above 50% from 1000 replications of full heuristic uncorrected distance analysis, 1000 bootstrap replications of full heuristic maximum parsimony analysis, and 25000 maximum likelihood quartet puzzling steps are shown above relevant branches. Intron gains and losses are indicated by light blue upper case and lower case letters, respectively, above branches to which they are mapped by simple presence-absence Dollo parsimony. Drosophila SRs are shown in shades of red, mosquito green, beetle brown, moth blue, and hymenopteran purple. Only the *D. pseudoobscura* SRs are included along with those from *D. melanogaster* to help balance the dipteran SR analysis. Inferred orthologous relationships within the three mosquito species are indicated by bars on the right. The large orange 1 and purple 2's indicate the two major tandem duplications hypothesized at the base of the dipteran SR evolution (see Figure 4). The blue 3 indicates the origin of the novel exon and hence also intron "q" in the expanded moth SR lineage.

ably reflecting even older gene subfamily events like the more recent ones seen within Drosophila and mosquitoes. Although our phylogenetic analysis does not provide bootstrap support for the single basal root of this fly SR lineage in Figure 3, it is supported by the apparent acquisition of intron F which is unique to the flies (see details below), so we hypothesize that the lineage arose from a single ancestral SR gene in an early fly. This does not mean that an early fly had only one SR gene, because it could well have had several, with all others being lost subsequently. We hypothesize that this single gene underwent a simple tandem gene duplication (large orange "1" in Figures 3 and 4), leading to two lineages. Then before the split of the major dipteran suborders (Brachycera and Nematocera, represented by Drosophila and mosquitoes, respectively), ~260 Mya [23], each of these lineages underwent simple tandem duplications (large purple 2's in Figures 3 and 4). Subsequent to this major organismal lineage split, the four existing SR genes underwent independent duplications, losses, and transpositions, leading to the current SR phylogeny of Drosophila and mosquitoes. In particular, the drosophilid flies have lost the lineage related to AgGr16/AaGr5/CpGr5. As a consequence of this convoluted history, ligand specificity determined for the Drosophila SRs cannot be directly transferred to the mosquito SRs, although it might be suggestive.

The Tribolium castaneum beetle SRs

The *T. castaneum* Gr family was described by HMR in the main genome paper [16], however in the phylogenetic analysis performed therein, the 16 SRs did not cluster as a single lineage, perhaps because the analysis included the entire Gr family. With the current phylogenetic analysis restricted to the SR subfamily, we believe we obtain refined clustering of the Tribolium SRs into a single lineage, however again there is no bootstrap support for this monophyly. Furthermore, we again are not proposing that beetles once had a single SR, simply that the existing Tribolium SR complement of 16 genes is monophyletic. The relatively high divergence of the Tribolium SRs from the other SRs precludes any suggestion of ligand specificity and we propose that all ligand specificity of the Tribolium SRs evolved independently within the beetle lineage.

The moth SRs

Krieger et al. [24] included cDNA sequences for two SRs in their initial description of a set of candidate chemoreceptors from the noctuid moth *Heliothis virescens* based on private partial genome sequences generated by Bayer and Exelixis Corporations. They subsequently obtained cDNAs, called Cr1 and Cr5 in their publication, which turn out to represent two divergent SR lineages within moths, as revealed by examination of the Gr family in the genome of the silkmoth *Bombyx mori* [25,26]. Details of our analysis of the Gr family in *B. mori* are published elsewhere [15], but we include the five SRs here for complete-

ness. *B. mori* has a single ortholog of HvCr5, which we have named BmGr4 (BmGr1-3 are the carbon dioxide receptors [27]). *B. mori* also has a simple ortholog of HvCr1, named BmGr6, as well as an expansion of three other genes, BmGr5, BmGr7, and BmGr8, in a monophyletic lineage with HvCr1/BmGr6. The monophyly of this lineage is supported not only by bootstrapping, but by each member's (BmGr5-8 and we predict HvCr1 as well) remarkable possession of a newly evolved short exon from within the ancestral phase 2 intron p (see intron details below). This novel exon is supported by a partial cDNA sequence for BmGr6, which was submitted to GenBank as "candidate olfactory receptor BmOR20" [GenBank:BAF31192.1 by T. Sakusai, Y. Hashimoto, and T. Nishioka in 2006], but which has a deletion of an upstream exon preventing complete translation. It has also been confirmed by sequencing of an RT/PCR product between the flanking exons for BmGr8 [15]; [GenBank:EU769119], and is inferred bioinformatically for BmGr5 and BmGr7. It encodes 15–20 amino acids in BmGr5-8 (Figure 5), and at least that number of amino acids in HvCr1 (where the genome sequence is not available, but the cDNA encodes these extra amino acids). These amino acids show no sequence conservation and are part of the extracellular loop 3 (ECL3; see below). This loop is usually very short in all insect Grs, so this evolution of a novel exon encoding a short stretch of variable amino acids that lengthen ECL3 is unusual.

The hymenopteran SRs

We examined the newly available parasitoid jewel wasp *Nasonia vitripennis* genome sequence for SRs, and find only simple orthologs of the *Apis mellifera* Gr1 and Gr2 genes [17], sharing 64 and 55 percent sequence identity, respectively (Figure 3). Bees and wasps form a sister group within the order Hymenoptera, so it appears likely that they all have only these two SRs. These two genes are next to and facing each other in a contig of the *N. vitripennis* genome assembly, but 2 Mbp apart on chromosome 5 in *A. mellifera*, suggesting that they have remained neighbors in the wasp lineage since their duplication in a common ancestor, but became separated in the bee lineage. These hymenopteran SRs form two sister lineages with the two SR lineages in moths, suggesting that these are rather old gene lineages. Although bootstrap support for this notion is not robust, much like the fly SR lineage sharing the unique intron f, each of these SR lineages in moths and hymenopterans has a unique intron position (g and l, respectively, see below).

Intron evolution

Robertson et al. [5] inferred considerable intron evolution within the insect chemoreceptor superfamily, and specifically the Gr family, from comparisons of the genes within *Drosophila melanogaster* alone, something that was also evident from the initial gene descriptions (e.g. [1,28]).
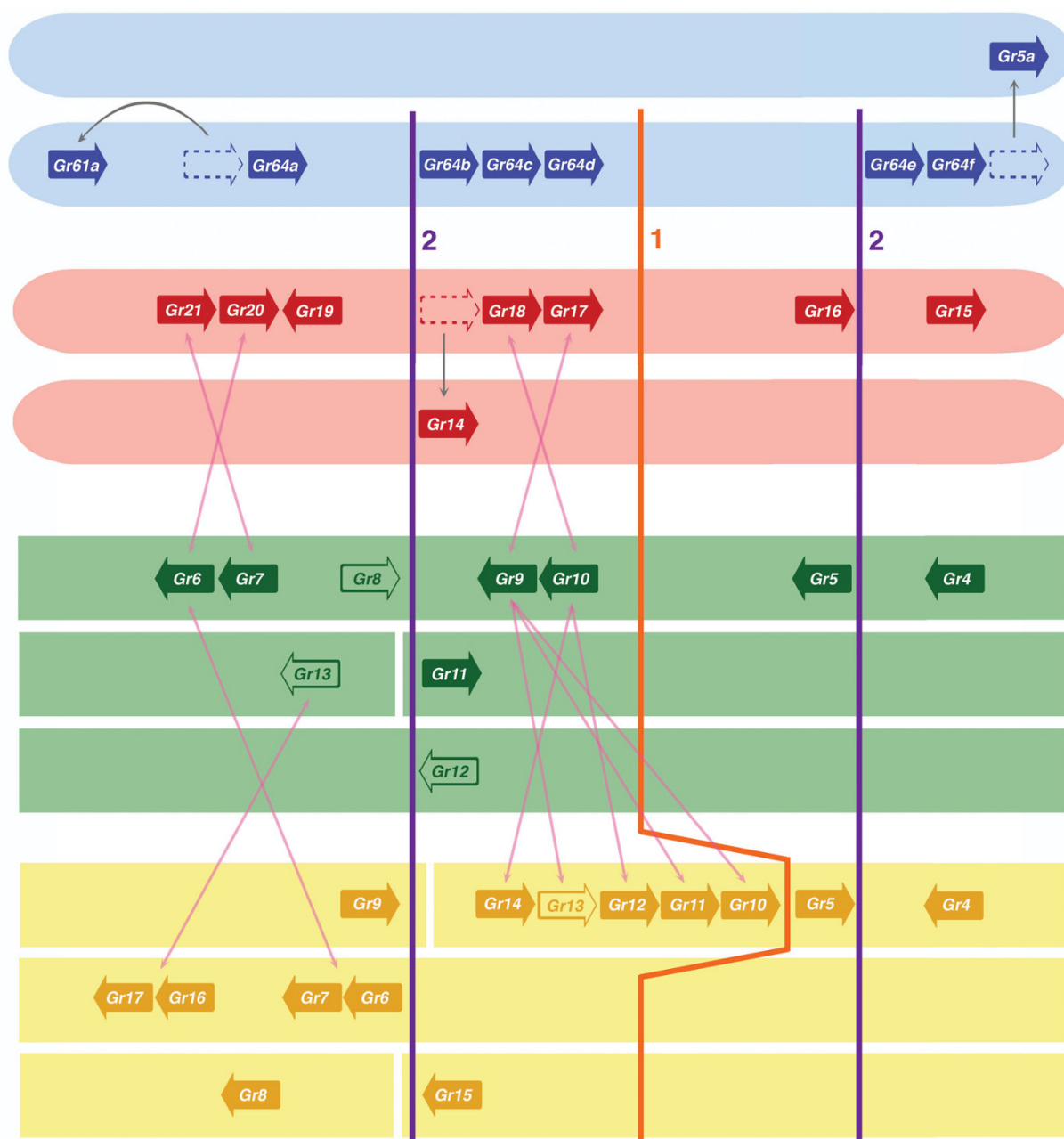
Figure 4
Model for the evolution of the SR genes in flies. Schematic diagram showing conservation of microsynteny of sugar Grs in *Drosophila melanogaster* (blue), *Anopheles gambiae* (red), *Aedes aegypti* (green) and *Culex pipiens* (yellow). Most of the genes share both phylogenetic relatedness and similar placement in chromosomes (*Ae. gambiae* and *D. melanogaster*) and supercontigs (*Ae. aegypti* and *C. pipiens*) across the four species, even after duplication events (shown with line 1 and lines 2, as in Fig. 3). Although it appears that some genes have moved within their respective genomes in *D. melanogaster* and *An. gambiae* (shown with grey arrows), their phylogenetic relationships (Fig. 3) reveal where they most likely had previously resided on the chromosomes. Likewise, it appears that there has been some level of gene shuffling in *Ae. aegypti* and *C. pipiens* (shown with pink arrows) with respect to the arrangements seen in *D. melanogaster* and *An. gambiae*. Although the genes have not yet been mapped to *Aedes* and *Culex* chromosomes, the ordered supercontigs of the genome sequences are likely accurate representations of microsynteny in each species.

```
Name    ----------TM5--------                   Novel exon                          --------TM6--------
HvCr1 LGGIVILSCSCNFYFICLQMFLGIT(?)---------------------------QGLSSDLLSLIYYVISLAWLCTRVISVVLA
BmGr4 LGALVLLSNVNNLYFICLQLFLGIN(2)---------------------------SKDRGSFINRLYYFISLGWLMFRACGVVLA
HvCr5 ISGVVFISFANNLFFICLQLFNTLE(?)DGLKGTGECTQLNSQSKLKK(?)IVVSKSGPLGGHEAAAYFLFSLVYLLSRSVAVSLI
BmGr6 INGILFISFANNLFFVCVQLFNTFD(2)DSVDMVGL-----CYNYSER(2)-----RTKPVGREPVIYLLFSLGFLISRSITVSLI
BmGr5 ISSVIFLSFASNLFFICLQLYNILS(2)NGVTSKYNLLKEMCPNYPSG(2)-------PLGGYEQIMYLLFSLSFLLGRSLVVSLV
BmGr7 LDYFVLISYFTNLFFICFQLYNSLN(2)RIYDANDV----CNENMDII(2)-------ATASVTYLTYYVFSFLFLVTRALLLSIM
BmGr8 LNPFVFISFTANLSYICAQLFYILN(2)KLTSSRT--VKITSFLEDKR(2)---------CDWETVLYISISFALVVLKVLLVSII
Intron                          p                   q
```

Figure 5
A novel exon in BmGr5-8. The amino acids of the TM5 and TM6 regions encoded by the ends of the exons flanking the novel exon are shown aligned with each other for the seven moth SRs. Amino acids shared by at least six of the sequences are highlighted in bold font. Like other SRs, BmGr4 does not have the novel exon, and the same is predicted for HvCr1. The location of the phase 2 intron in BmGr4 (intron p in Figure 6) is shown. BmGr5-8 have a short novel exon encoding 15–20 unconserved amino acids, and an additional flanking phase 2 intron (intron q in Figure 6). HvCr5 is predicted to have an intron q, however the precise location cannot be predicted.

Examination of this issue within the three-gene carbon dioxide lineage revealed considerable intron gain and loss even just within this gene lineage [27]. Intron evolution within these SR genes is again evident. We rename the introns from those in Robertson et al. [5], because otherwise the alphabetical naming system becomes cumbersome. The intron locations and phases are shown in Figure 6, and their gains and losses are mapped on the tree in Figure 3. Inclusion of the other insect SRs leads to some revision of when some introns were gained, compared with Figure 3 in Robertson et al. [5]. For example, intron f (d in [5]) can now be inferred to have been gained in the single ancestor of the fly SR genes, rather than being diagnostic of the SRs in general, because it is absent in the non-Dipteran genes. Indeed, it now provides useful support for our proposal that the fly SRs are monophyletic, all originating from a single SR gene in a basal fly lineage. Intron f was then lost independently six times within the

dipteran gene expansion. Conversely, intron r (m in [5]), which was initially inferred to have been gained within the Drosophila SRs in Robertson et al. [5], is clearly much older because it is shared by the Bombyx and Tribolium SRs. Indeed, we now infer that it is ancestral for the entire SR subfamily.

We infer 12 intron gains within the SRs (excluding hjmo-prs), including the two N-terminal introns (a and b) in the Drosophila Gr5a/Gr64e/Gr64f lineage and mosquito relatives, the splitting of the ancestral phase 2 intron p in the expanded moth SR lineage yielding novel intron q, and eight other novel introns in the non-dipteran genes. Most of these intron locations are not only unique within this SR lineage, but also within the entire Gr family, which represents most of the diversity of the insect chemoreceptor superfamily [5]. Indeed, only the two C-terminal introns r and s (2 and 3 in [5]) are shared across the Gr family.
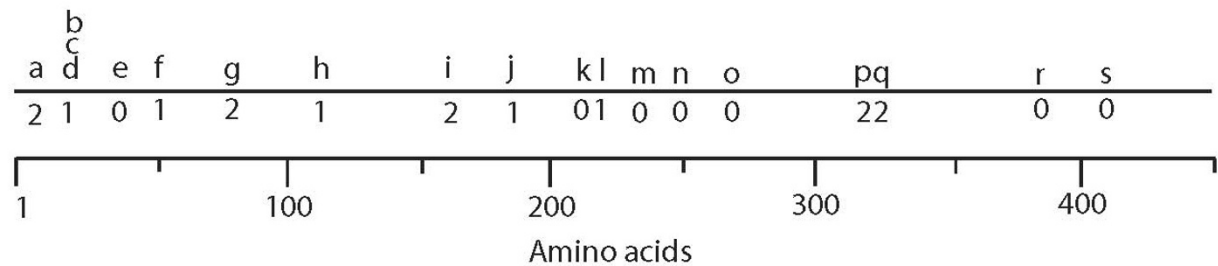

Figure 6
Intron locations and phases in the insect SR genes. Introns are named in an alphabetical series from 5' to 3' ends of the genes, positioned above a line representing a 450aa SR, with the longer N-terminus typical of the DmGr5a and 64e/f lineage. Hence the coordinates of the introns are ~20 aa later than those shown in Figure 4 of Robertson et al. [5]. Correspondence between intron names in Robertson et al. [5] and those herein is: a/a, b/b, d/f, j/h, o/j, r/m, y/o, a'/p, 2/r, 3/s. Intron phase is shown below the line (0 is between codons, 1 is after the first codon position, and 2 after the second codon position).

Intron losses are rather more frequent, totaling 58 in Figure 3, plus seven in the Drosophila species in Figure 1 that do not overlap with Figure 3, for a total of 65. The two C-terminal and ancient introns, r and s, reveal the extremes of intron loss, with intron r being lost 10 times while intron s was only lost twice, in a mosquito lineage and in DwilGr64e. It remains unclear why the final intron is so seldom lost. This pattern is found not only in the SRs, but is consistent throughout the entire superfamily, as this C-terminal intron is almost always present. The only major ambiguity in this analysis of intron evolution is the series of phase 1 introns near the N-terminus, named b, c, and d. These three introns are in roughly the same location, but are found in subsets of the dipteran, moth, and hymenopteran SRs. Unfortunately, the N-terminal sequences are so divergent across these three insect orders that they cannot confidently be considered to be homologous intron placements, and given their disparate locations in the tree, are considered here to be independent gains. As was true for the carbon dioxide receptors [27], and appears generally true across entire genomes (e.g. [29]), intron losses are more frequent in the Diptera, with all dipteran gene lineages having lost at least one intron, while some moth, beetle, and hymenopteran lineages have lost none and some only gained introns (although it is also formally possible that some of these non-dipteran introns are ancestral to the SR family and were lost from the dipteran gene lineages).

Distinctive features of SRs
The SRs form a distinctive subfamily, with a long branch connecting them to the rest of the Gr family [5]. They appear, therefore, to have been evolving independently from the rest of the family for some time, which in part explains their distinctive set of introns. SRs are also slightly longer than most Grs, most often attributed to an N-terminal extension. This extension is particularly long in the DmGr5a/DmGr64e/DmGr64f and related mosquito SR lineage, which includes one or two additional N-terminal exons (separated by introns a and/or b in Figure 6, except for AaGr4 and AgGr15 which lost both introns). All SRs have a "pre-peak" of hydrophobic amino acids in the N-terminus that is sometimes predicted to be an eighth TM domain by various TM-domain prediction programs (see below). The amino acids within this "pre-peak" region are remarkably well conserved, with a motif that can be described as hHxAh(G/A/S)Phhhh(G/A/S)Qhh(G/A/S)hhPh, where h stands for any hydrophobic amino acid (F, I, L, M, or V; alignment positions 84–103 in Figures 7, 8, 9, 10, 11, 12 and 13). The final proline is the most conserved position. This motif is shared by most other Grs, including the carbon dioxide receptors, which in larger phylogenetic analyses usually form the sister group to the SRs. However, the somewhat conserved histidine at the start of the motif is distinctive to the SRs. The otherwise

unconserved TM1 domain starts with a well-conserved serine (S alignment position 121) that is only shared with the carbon dioxide receptors, while the TM2 domain ends with a completely conserved tryptophan (W 189) that is idiosyncratically present in other Grs. TM3 has a highly conserved glutamic acid (E 238), while TM4 has a completely conserved aspartic acid (D 309), both seemingly unique to the SRs, although these regions are poorly aligned across all Grs. The intracellular loop 2 (ICL2) contains a highly conserved tryptophan (W 357), shared only with the carbon dioxide receptors, followed after four amino acids by a distinctively conserved arginine (R 361). The C-terminal TM6/ICL3/TM7 region is the most highly conserved region of the Grs, so it is not surprising that the SRs contain several conserved residues here that are shared with the rest of the Gr family, including the TY (521/2) and QF (528/9) pairs in TM7, although the QF pair was replaced in about half the fly SRs. The most distinctive residue in the SRs is a completely conserved glutamic acid (E 523) immediately after the TY pair in TM7. In other Grs this residue is usually a hydrophobic amino acid. This glutamic acid is seen in no other available insect Grs, hence is apparently diagnostic for the SR subfamily. Its conservation within the SR subfamily suggests that it somehow plays a crucial role in the perception of sugars.

Membrane topology
Recent studies have shown that the insect Ors, while most likely containing seven TM domains, have the opposite membrane topology to that of the G-protein coupled receptors that constitute chemoreceptors in vertebrates and nematodes [30-33]. That is, their N-termini are internal while their C-termini are external. We believe this topology is present throughout the Gr family as well and hence the entire superfamily [27]. These SRs provide a particularly clear illustration of this topology. Examination of the CLUSTALX alignment in Figure 7 shows clearly the relatively long intracellular (IC) loop 2 between TM4 and TM5, and IC3 between TM6 and TM7. Furthermore, these intracellular loops each have several conserved positively charged arginine (R) and lysine (K) residues, in agreement with the "positive inside" rule of von Heijne [34,35]. In contrast, the extracellular loop 3 between TM5 and TM6 is particularly short (except in the lepidopteran lineage with a novel exon), and devoid of conserved positively charged residues. Similar trends apply to the more N-terminal loops. A remaining uncertainty with respect to the membrane topology of these SRs, and Grs in general, is the presence of a "pre-peak" of hydrophobic amino acids that is less evident in the Ors, the only family studied experimentally to date. This pre-peak is around 21 amino acids long, the minimum required for a TM domain. For most of the SRs, it is predicted to be a TM domain by most hydropathy and TM domain prediction programs, as summarized in the ConPredII website [36]. Most of these pro-
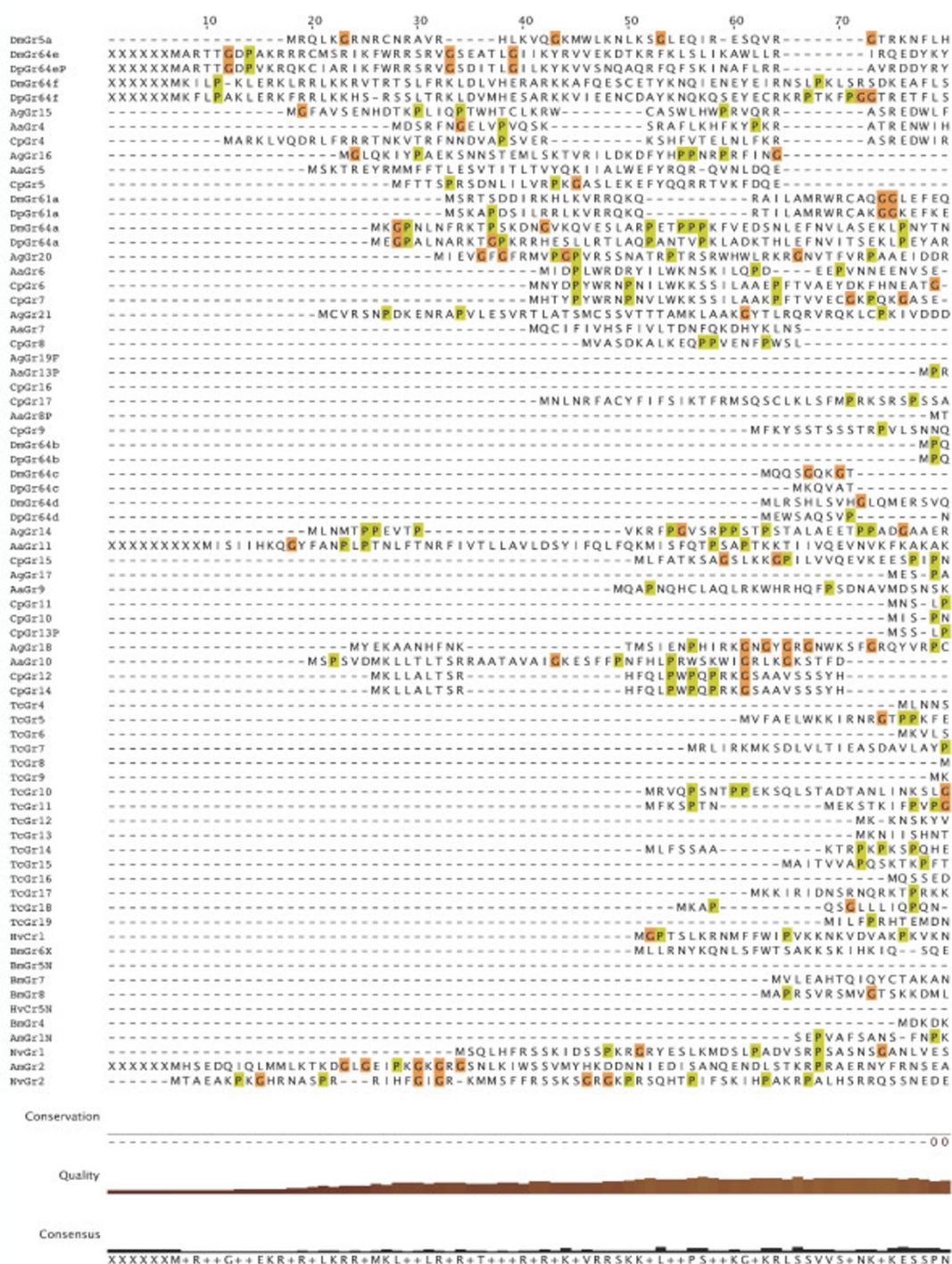
Figure 7
Multiple alignment of the insect SRs. This is the CLUSTALX alignment employed for the phylogenetic analysis in Figure 3, except that positions of uncertain alignment and large gaps, specifically alignment positions 1–82, 150–160, 209–219, 340–355, 405–435, and 531–559, were excluded from the phylogenetic analysis. The Xs at the start of some sequences were added to facilitate splitting the figure into six convenient parts. Alignment positions are shown at the bottom, along with the "conservation" histogram from CLUSTALX. The predicted TM domains are evident as vertical bands of hydrophobic amino acids shaded blue, and the "pre-peak" is approximately alignment positions 84–103, TM1 is 121–142, TM2 is 161–189, TM3 is 224–234, TM4 is 291–314, TM5 is 378–398, TM6 is 438–457, and TM7 is 511–529.
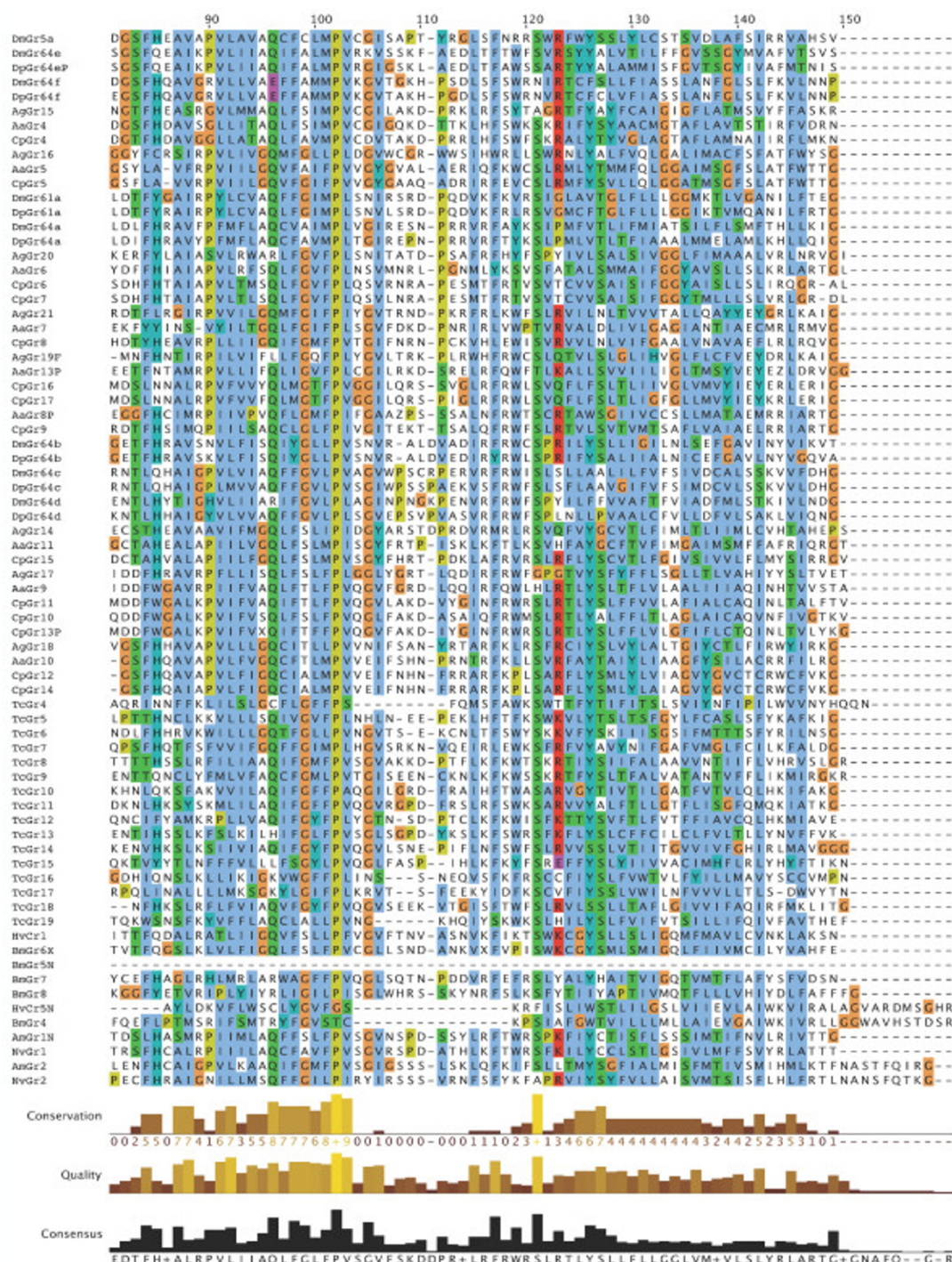
Figure 8
Multiple alignment of the insect SRs. This is the CLUSTALX alignment employed for the phylogenetic analysis in Figure 3, except that positions of uncertain alignment and large gaps, specifically alignment positions 1–82, 150–160, 209–219, 340–355, 405–435, and 531–559, were excluded from the phylogenetic analysis. The Xs at the start of some sequences were added to facilitate splitting the figure into six convenient parts. Alignment positions are shown at the bottom, along with the "conservation" histogram from CLUSTALX. The predicted TM domains are evident as vertical bands of hydrophobic amino acids shaded blue, and the "pre-peak" is approximately alignment positions 84–103, TM1 is 121–142, TM2 is 161–189, TM3 is 224–234, TM4 is 291–314, TM5 is 378–398, TM6 is 438–457, and TM7 is 511–529.
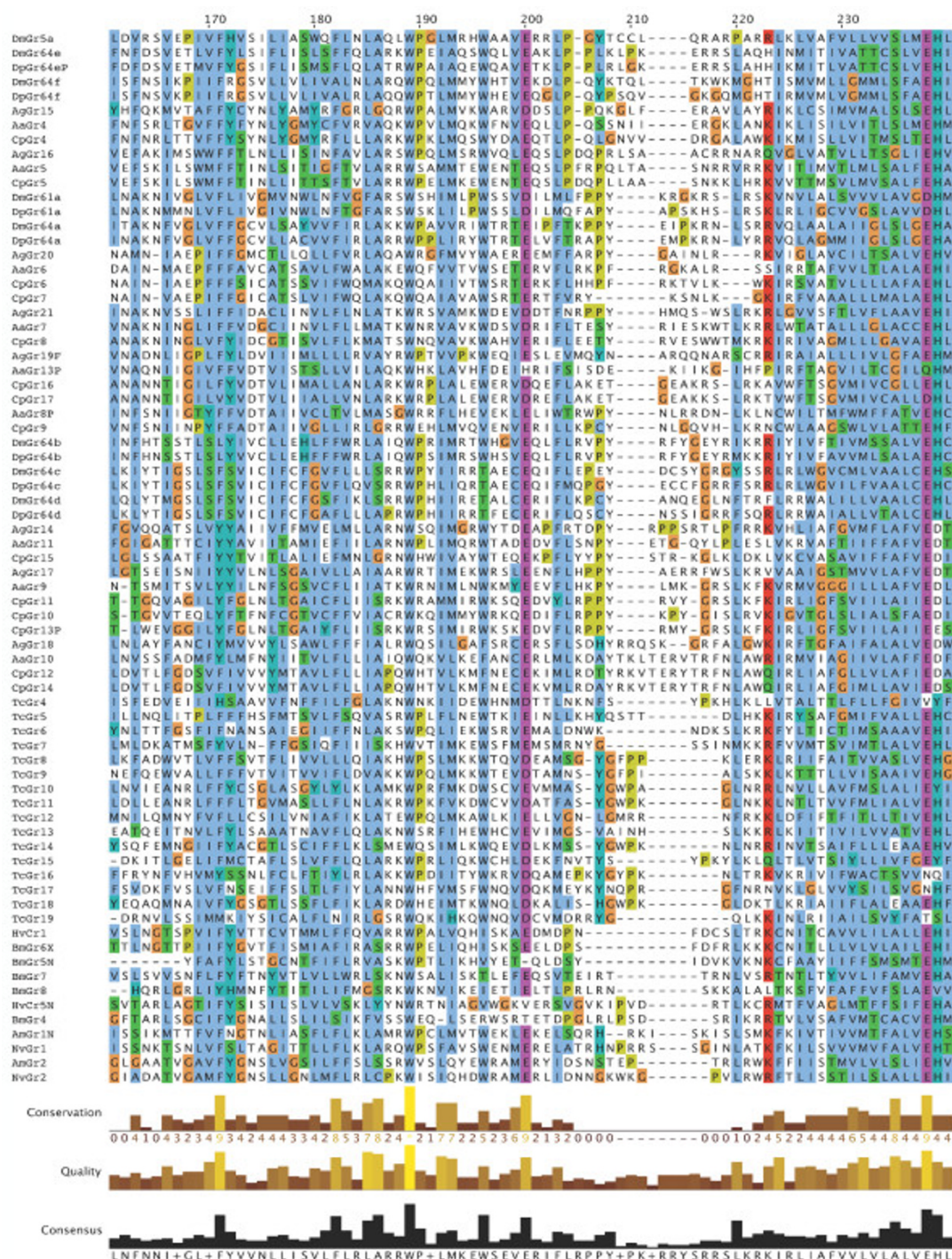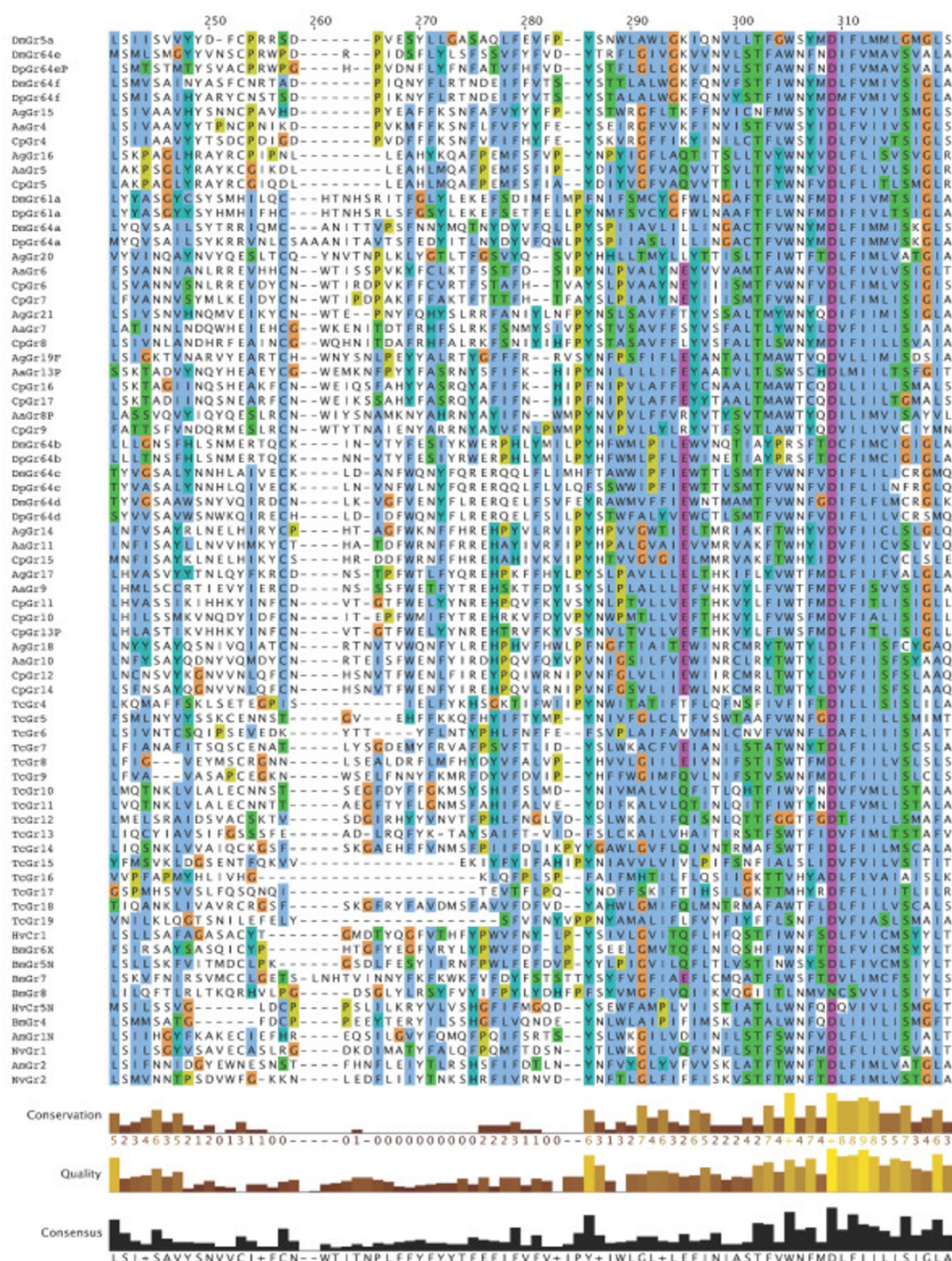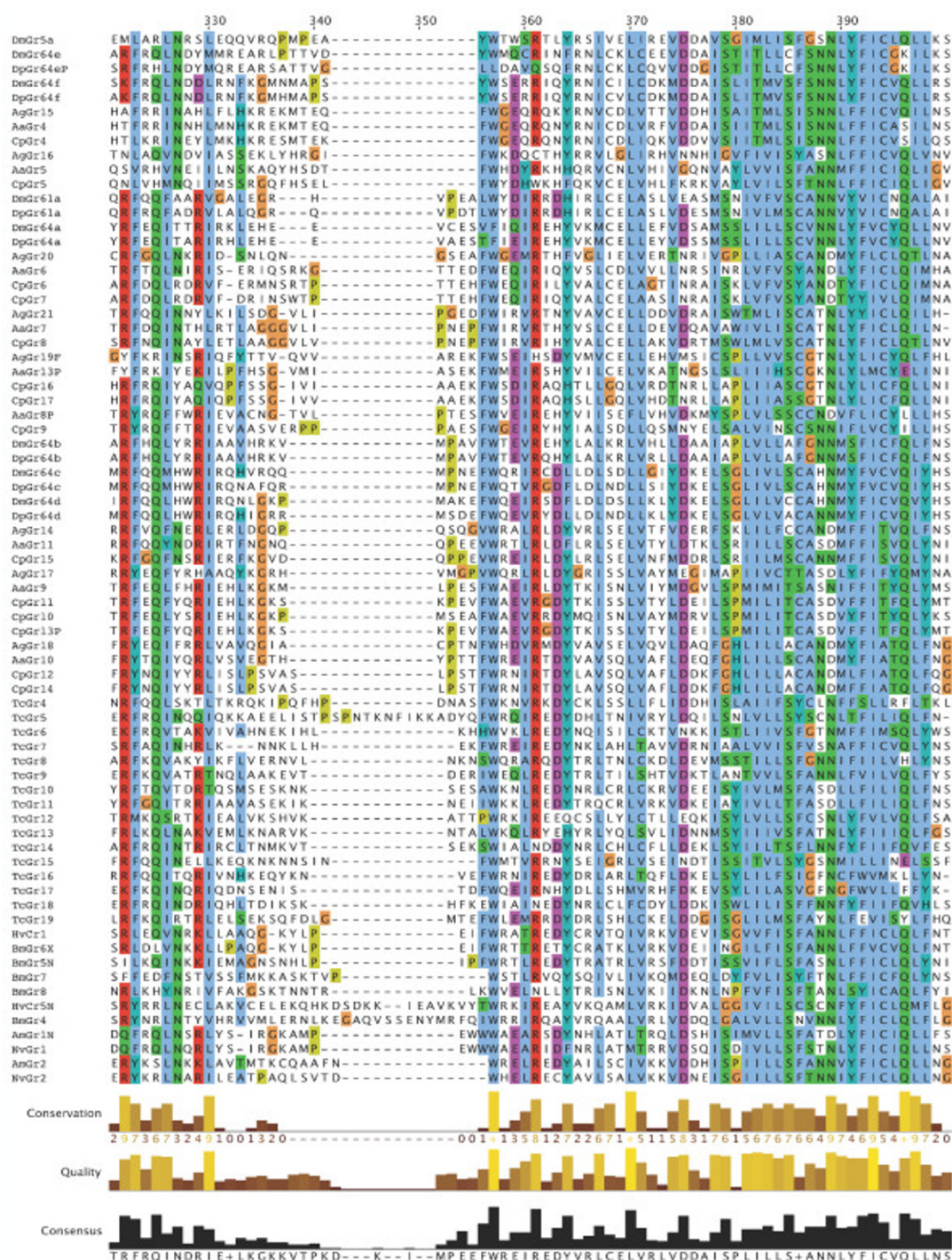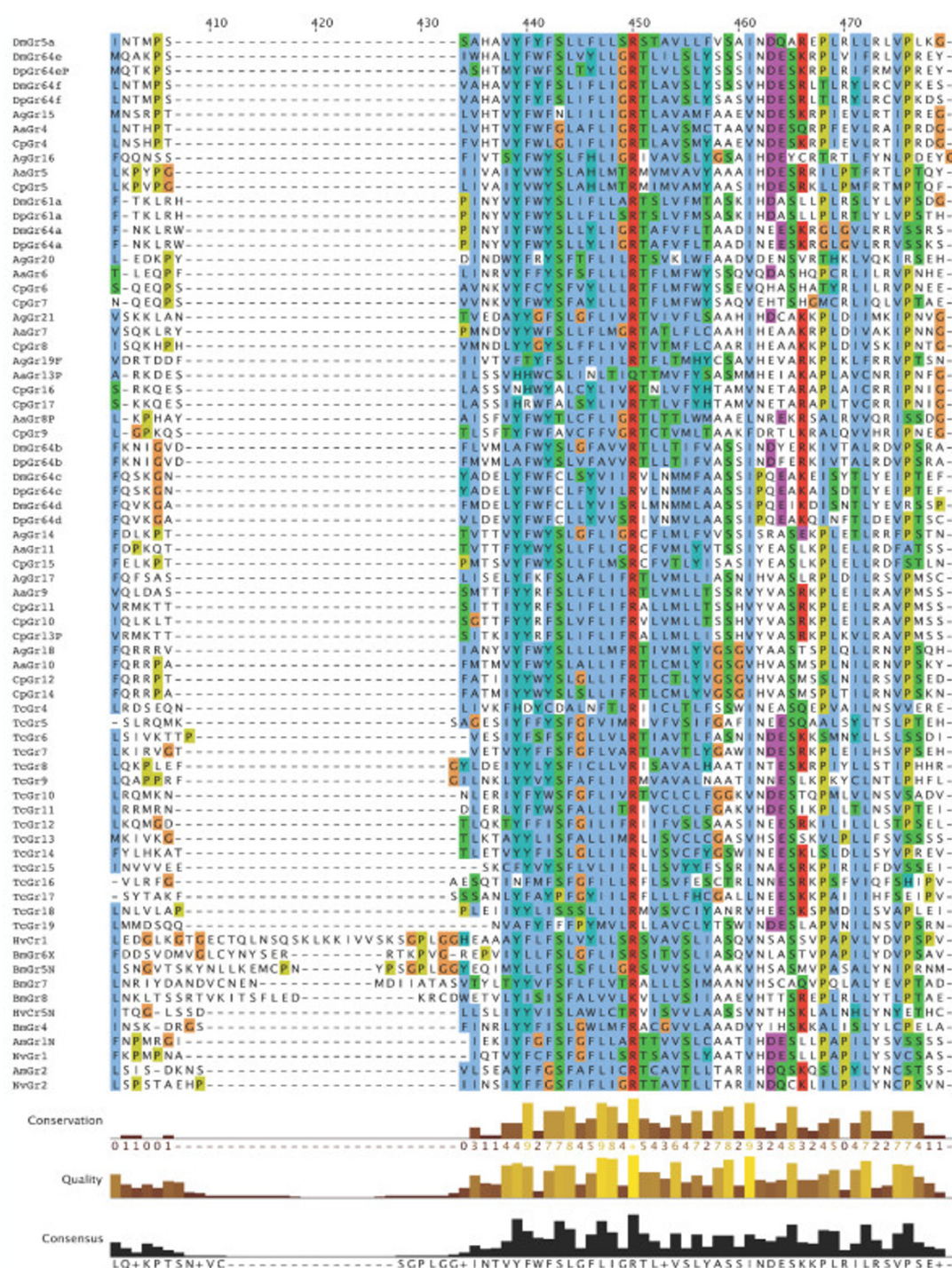
Figure 9
Multiple alignment of the insect SRs. This is the CLUSTALX alignment employed for the phylogenetic analysis in Figure 3, except that positions of uncertain alignment and large gaps, specifically alignment positions 1–82, 150–160, 209–219, 340–355, 405–435, and 531–559, were excluded from the phylogenetic analysis. The Xs at the start of some sequences were added to facilitate splitting the figure into six convenient parts. Alignment positions are shown at the bottom, along with the "conservation" histogram from CLUSTALX. The predicted TM domains are evident as vertical bands of hydrophobic amino acids shaded blue, and the "pre-peak" is approximately alignment positions 84–103, TM1 is 121–142, TM2 is 161–189, TM3 is 224–234, TM4 is 291–314, TM5 is 378–398, TM6 is 438–457, and TM7 is 511–529.
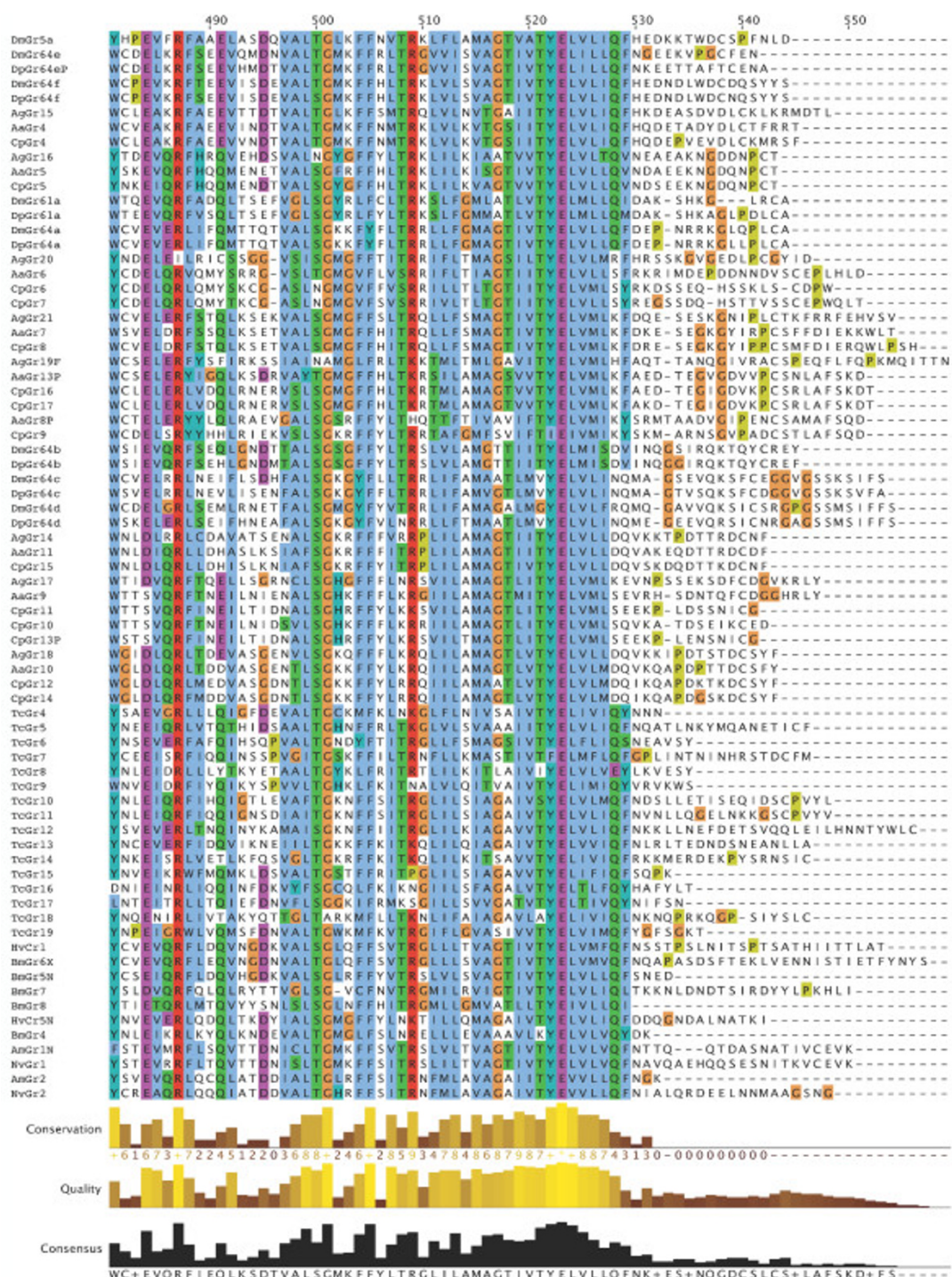
Figure 10
Multiple alignment of the insect SRs. This is the CLUSTALX alignment employed for the phylogenetic analysis in Figure 3, except that positions of uncertain alignment and large gaps, specifically alignment positions 1–82, 150–160, 209–219, 340–355, 405–435, and 531–559, were excluded from the phylogenetic analysis. The Xs at the start of some sequences were added to facilitate splitting the figure into six convenient parts. Alignment positions are shown at the bottom, along with the "conservation" histogram from CLUSTALX. The predicted TM domains are evident as vertical bands of hydrophobic amino acids shaded blue, and the "pre-peak" is approximately alignment positions 84–103, TM1 is 121–142, TM2 is 161–189, TM3 is 224–234, TM4 is 291–314, TM5 is 378–398, TM6 is 438–457, and TM7 is 511–529.

Figure 11
Multiple alignment of the insect SRs. This is the CLUSTALX alignment employed for the phylogenetic analysis in Figure 3, except that positions of uncertain alignment and large gaps, specifically alignment positions 1–82, 150–160, 209–219, 340–355, 405–435, and 531–559, were excluded from the phylogenetic analysis. The Xs at the start of some sequences were added to facilitate splitting the figure into six convenient parts. Alignment positions are shown at the bottom, along with the "conservation" histogram from CLUSTALX. The predicted TM domains are evident as vertical bands of hydrophobic amino acids shaded blue, and the "pre-peak" is approximately alignment positions 84–103, TM1 is 121–142, TM2 is 161–189, TM3 is 224–234, TM4 is 291–314, TM5 is 378–398, TM6 is 438–457, and TM7 is 511–529.

Figure 12
Multiple alignment of the insect SRs. This is the CLUSTALX alignment employed for the phylogenetic analysis in Figure 3, except that positions of uncertain alignment and large gaps, specifically alignment positions 1–82, 150–160, 209–219, 340–355, 405–435, and 531–559, were excluded from the phylogenetic analysis. The Xs at the start of some sequences were added to facilitate splitting the figure into six convenient parts. Alignment positions are shown at the bottom, along with the "conservation" histogram from CLUSTALX. The predicted TM domains are evident as vertical bands of hydrophobic amino acids shaded blue, and the "pre-peak" is approximately alignment positions 84–103, TM1 is 121–142, TM2 is 161–189, TM3 is 224–234, TM4 is 291–314, TM5 is 378–398, TM6 is 438–457, and TM7 is 511–529.

Figure 13
Multiple alignment of the insect SRs. This is the CLUSTALX alignment employed for the phylogenetic analysis in Figure 3, except that positions of uncertain alignment and large gaps, specifically alignment positions 1–82, 150–160, 209–219, 340–355, 405–435, and 531–559, were excluded from the phylogenetic analysis. The Xs at the start of some sequences were added to facilitate splitting the figure into six convenient parts. Alignment positions are shown at the bottom, along with the "conservation" histogram from CLUSTALX. The predicted TM domains are evident as vertical bands of hydrophobic amino acids shaded blue, and the "pre-peak" is approximately alignment positions 84–103, TM1 is 121–142, TM2 is 161–189, TM3 is 224–234, TM4 is 291–314, TM5 is 378–398, TM6 is 438–457, and TM7 is 511–529.

teins are predicted to have eight TM domains with the N-terminus external, although the range is from six to nine TM domains (insect chemoreceptor TM domains are seldom as well-defined as those of most other TM proteins, while TM4 is sometimes split into two). An example of these hydropathy plots, predicted TM domains, and predicted topology is shown for AmGr2 in Figures 14 and 15. Further experimental work will be required to resolve this issue, however, it has now been shown that at least the Ors are in fact ligand-gated ion channels [33,37,38], and the same is surely true of these Grs.

Discussion

Our analysis of the evolution of the SRs in insects reveals a remarkable pattern (Figure 3). Each major lineage of SRs within an insect order appears to have originated from just one or two genes. Thus we hypothesize that all the fly SRs, all the Tribolium SRs, and most of the moth SRs originated from a single basal gene within each organismal lineage. In contrast, both moths and the hymenopteran wasp/bee lineage appear to have shared two SRs lineages for a long time. Recent work on the SRs in *D. melanogaster* strongly suggests that, like the Ors and the carbon dioxide receptors, they function as heterodimers [9,10,39]. If this is the case, then we can predict that the two existing SRs in the wasp/bee lineage function as a single receptor capable of recognizing all sugars that these hymenopterans can sense. This implies that, much like mammals which have a single heterodimeric SR pair [40], these species should not be able to differentiate different sugars. We infer then that moths, through duplication of one of their two ancestral SRs into four genes, probably do have the ability to discriminate different sugars, most likely by combining one of these four proteins with the single BmGr4/HvCr5 protein in different gustatory sensory neurons. Finally, although all existing fly and Tribolium SRs each appear to
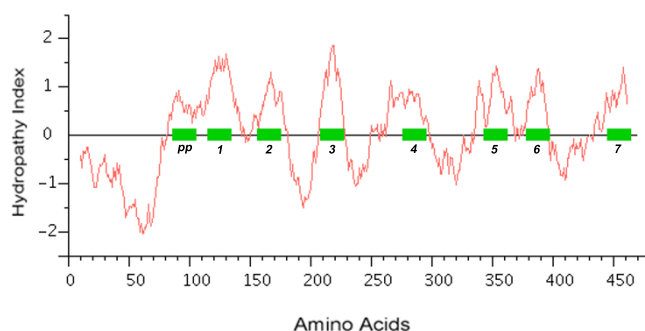

Figure 14
Hydropathy plot and predicted TM domains for AmGr2. Regions of hydrophobic amino acids yield stretches of positive hydropathy, and these are predicted to be TM domains by the ConPredII prediction program (indicated by thick green bars).
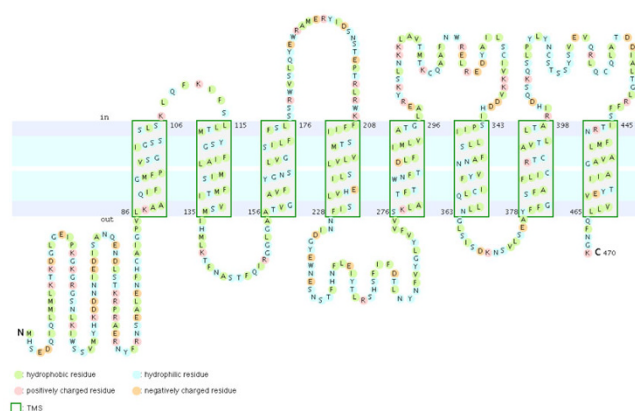

Figure 15
Predicted membrane topology for AmGr2. ConPredII predicts eight TM domains for this protein and many other SRs, usually as here with the N-terminus internal and the C-terminus external.

have evolved from a single SR gene, as noted in the results this does not imply that ancestral flies and beetles had a single SR, because additional genes could have been lost. Today, however, flies apparently employ combinations of their SRs allowing recognition and discrimination of diverse sugars. Dahanukar et al. [9] infer that DmGr5a and DmGr64a are crucial to sugar perception because a double mutant removing both of them is incapable of recognizing any sugars. Since Gr5a and Gr64a are the most widely expressed of the SRs, with Gr61a and Gr64b-f apparently being expressed in limited sets of neurons overlapping with Gr5a and Gr64a [9,10,39], a simple model is that functional heterodimers require either Gr5a or Gr64a. An obvious problem with this simple model is that Gr5a has been lost independently from both the *D. pseudoobscura/persimilis* and *D. grimshawi* lineages, and it seems unlikely that these species would have lost such a major portion of their sugar-sensing abilities. Gr5a and Gr64a nevertheless do represent the two major SR fly lineages after an initial duplication (Figure 3), so it appears that one daughter gene from each of these two lineages has specialized in being the more widely expressed partner, while the others, Gr61a and Gr64b-f, might be involved in recognition of particular suites of sugars. It is not obvious from the Tribolium SRs which protein(s) might be the widely expressed heterodimeric partner(s) of the others.

An unusual aspect of these SRs is the origin of a novel exon from within an intron in the expanded lineage of moth SRs. Novel exons are known to have evolved from intronic sequences in various vertebrates, in a process called "exonization". Most such instances have resulted

from the evolution of splice sites involving a short retro-transposon or SINE, such as Alu elements in humans (reviewed by [41]), however no such examples appear to have been published from an insect. Exonization is thought to occur with such a retroelement inserted in the opposite orientation to transcription with the inverse "poly-A" tail of the retroelement forming a pseudo 3' splice acceptor site, along with de novo formation of a 5' splice donor site within the retroelement. SINEs are wide-spread in *B. mori* [25,26] and likely other moth genomes, so perhaps such exonization events will be relatively common in moths. This particular event is too old for any vestiges of the potentially originating retroelement to remain. The novel exon in the four BmGr5-8 genes is short, encoding just 15–20 amino acids. The exon exhibits no sequence conservation among the four genes. These extra amino acids nevertheless more than double the length of the third extracellular loop in these four moth SRs relative to all the other SRs, and most other Grs. The origin of the one or two N-terminal exons in the Drosophila Gr5a/64e/f lineage and mosquito relatives, and hence the existence of introns a and b, is also a novelty in the SR subfamily and Gr family, but whether these evolved by insertion of introns into an extended 5' exon, extension of the start of translation into a 5' UTR exon, or true exonization is unclear.

Conclusion

Our investigation reveals that the repertoire of extant insect sugar receptors can be traced to one or two ancestral genes in each major insect order. We are unable to say much about the even older evolutionary history of the insect SRs because the body louse *Pediculus humanus*, representing a more basal insect lineage in the Exopterygota as compared with the endopterygote insects herein, does not have SRs (HMR unpublished results). The long branch leading to the SRs from the rest of the Gr family [5], suggests that the louse should have SRs but may have lost them during evolution of its obligate ectoparasitic lifestyle. The imminent availability of genome sequences for two other exopterygote insect lineages, the pea aphid *Acyrthosiphon pisum* and the kissing bug *Rhodnius prolixus*, as well as other arthropod genomes, will hopefully further illuminate the origin of the insect sugar receptors from within the Gr family. We predict, however, that those with SRs will always have at least two proteins forming a heterodimer capable of detecting diverse sugars, as represented today by the two SRs in bees and wasps.

Authors' contributions

LBK annotated mosquito Grs (*Culex*; *Aedes* already published), analyzed data, designed and edited figures, helped draft and revise manuscript. HMR annotated Grs, analyzed data, drafted manuscript and designed figures.

Additional material

Additional file 1

Amino acid sequences of SRs. We have provided the amino acid sequences of the complete set of sugar receptors used in our analyses. Sequences are presented in FASTA format, but have been converted to a PDF for simplified access.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2148-9-41-S1.pdf]

References

1. Clyne PJ, Warr CG, Carlson JR: **Candidate taste receptors in *Drosophila*.** *Science* 2000, **287(5459):**1830-1834.
2. Scott K, Brady R Jr, Cravchik A, Morozov P, Rzhetsky A, Zuker C, Axel R: **A Chemosensory Gene Family Encoding Candidate Gustatory and Olfactory Receptors in *Drosophila*.** *Cell* 2001, **104:**661-673.
3. Dunipace L, Meister S, McNealy C, Amrein H: **Spatially restricted expression of candidate taste receptors in the *Drosophila* gustatory system.** *Current Biology* 2001, **11:**822-835.
4. Chyb S, Dahanukar A, Wickens A, Carlson JR: ***Drosophila* Gr5a encodes a taste receptor tuned to trehalose.** *PNAS* 2003, **100(suppl 2):**14526-14530.
5. Robertson H, Warr CG, Carlson JR: **Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*.** *PNAS* 2003, **100(Suppl 2):**14537-14542.
6. Thorne N, Chromey C, Bray S, Amrein H: **Taste perception and coding in *Drosophila*.** *Current Biology* 2004, **14(12):**1065-1079.
7. Wang Z, Singhvi A, Kong P, Scott K: **Taste representations in the *Drosophila* brain.** *Cell* 2004, **117(7):**981-991.
8. Jiao Y, Moon SJ, Montell C: **A *Drosophila* gustatory receptor required for the responses to sucrose, glucose, and maltose identified by mRNA tagging.** *PNAS* 2007, **104(35):**14110-14115.
9. Dahanukar A, Lei Y-T, Kwon JY, Carlson JR: **Two *Gr* Genes Underlie Sugar Reception in *Drosophila*.** *Neuron* 2007, **56(3):**503-516.
10. Slone J, Daniels J, Amrein H: **Sugar receptors in *Drosophila*.** *Current Biology* 2007, **17(20):**1809-1816.
11. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, Pollard DA, Sackton TB, Larracuente AM, Singh ND, Abad JP, Abt DN, Adryan B, Aguade M, Akashi H, Anderson WW, Aquadro CF, Ardell DH, Arguello R, Artieri CG, Barbash DA, Barker D, Barsanti P, Batterham P, Batzoglou S, Begun D, Bhutkar A, Blanco E, Bosak SA, Bradley RK, Brand AD, Brent MR, Brooks AN, Brown RH, Butlin RK, Caggese C, Calvi BR, Bernardo de Carvalho A, Caspi A, Castrezana S, Celniker SE, Chang JL, Chapple C, Chatterji S, Chinwalla A, Civetta A, Clifton SW, Comeron JM, Costello JC, Coyne JA, Daub J, David RG, Delcher AL, Delehaunty K, Do CB, Ebling H, Edwards K, Eickbush T, Evans JD, Filipski A, Findeiss S, Freyhult E, Fulton L, Fulton R, Garcia AC, Gardiner A, Garfield DA, Garvin BE, Gibson G, Gilbert D, Gnerre S, Godfrey J, Good R, Gotea V, Gravely B, Greenberg AJ, Griffiths-Jones S, Gross S, Guigo R, Gustafson EA, Haerty W, Hahn MW, Halligan DL, Halpern AL, Halter GM, Han MV, Heger A, Hillier L, Hinrichs AS, Holmes I, Hoskins RA, Hubisz MJ, Hultmark D, Huntley MA, Jaffe DB, Jagadeeshan S, Jeck WR, Johnson J, Jones CD, Jordan WC, Karpen GH, Kataoka E, Keightley PD, Kheradpour P, Kirkness EF, Koerich LB, Kristiansen K, Kudrna D, Kulathinal RJ, Kumar S, Kwok R, Lander E, Langley CH, Lapoint R, Lazzaro BP, Lee SJ, Levesque L, Li R, Lin CF,

Lin MF, Lindblad-Toh K, Llopart A, Long M, Low L, Lozovsky E, Lu J, Luo M, Machado CA, Makalowski W, Marzo M, Matsuda M, Matzkin L, McAllister B, McBride CS, McKernan K, McKernan K, Mendez-Lago M, Minx P, Mollenhauer MU, Montooth K, Mount SM, Mu X, Myers E, Negre B, Newfeld S, Nielsen R, Noor MA, O'Grady P, Pachter L, Papaceit M, Parisi MJ, Parisi M, Parts L, Pedersen JS, Pesole G, Phillippy AM, Ponting CP, Pop M, Porcelli D, Powell JR, Prohaska S, Pruitt K, Puig M, Quesneville H, Ram KR, Rand D, Rasmussen MD, Reed LK, Reenan R, Reily A, Remington KA, Rieger TT, Ritchie MG, Robin C, Rogers YH, Rohde C, Rozas J, Rubenfield MJ, Ruiz A, Russo S, Salzberg SL, Sanchez-Gracia A, Saranga DJ, Sato H, Schaeffer SW, Schatz MC, Schlenke T, Schwartz R, Segarra C, Singh RS, Sirot L, Sirota M, Sisneros NB, Smith CD, Smith TF, Spieth J, Stage DE, Stark A, Stephan W, Strausberg RL, Strempel S, Sturgill D, Sutton G, Sutton GG, Tao W, Teichmann S, Tobari YN, Tomimura Y, Tsolas JM, Valente VL, Venter E, Venter JC, Vicario S, Vieira FG, Vilella AJ, Villasante A, Walenz B, Wang J, Wasserman M, Watts T, Wilson D, Wilson RK, Wing RA, Wolfner MF, Wong A, Wong GK, Wu CI, Wu G, Yamamoto D, Yang HP, Yang SP, Yorke JA, Yoshida K, Zdobnov E, Zhang P, Zhang Y, Zimin AV, Baldwin J, Abdouelleil A, Abdulkadir J, Abebe A, Abera B, Abreu J, Acer SC, Aftuck L, Alexander A, An P, Anderson E, Anderson S, Arachi H, Azer M, Bachantsang P, Barry A, Bayul T, Berlin A, Bessette D, Bloom T, Blye J, Boguslavskiy L, Bonnet C, Boukhgalter B, Bourzgui I, Brown A, Cahill P, Channer S, Cheshatsang Y, Chuda L, Citroen M, Collymore A, Cooke P, Costello M, D'Aco K, Daza R, De Haan G, DeGray S, DeMaso C, Dhargay N, Dooley K, Dooley E, Doricent M, Dorje P, Dorjee K, Dupes A, Elong R, Falk J, Farina A, Faro S, Ferguson D, Fisher S, Foley CD, Franke A, Friedrich D, Gadbois L, Gearin G, Gearin CR, Giannoukos G, Goode T, Graham J, Grandbois E, Grewal S, Gyaltsen K, Hafez N, Hagos B, Hall J, Henson C, Hollinger A, Honan T, Huard MD, Hughes L, Hurhula B, Husby ME, Kamat A, Kanga B, Kashin S, Khazanovich D, Kisner P, Lance K, Lara M, Lee W, Lennon N, Letendre F, LeVine R, Lipovsky A, Liu X, Liu J, Liu S, Lokyitsang T, Lokyitsang Y, Lubonja R, Lui A, MacDonald P, Magnisalis V, Maru K, Matthews C, McCusker W, McDonough S, Mehta T, Meldrim J, Meneus L, Mihai O, Mihalev A, Mihova T, Mittelman R, Mlenga V, Montmayeur A, Mulrain L, Navidi A, Naylor J, Negash T, Nguyen T, Nguyen N, Nicol R, Norbu C, Norbu N, Novod N, O'Neill B, Osman S, Markiewicz E, Oyono OL, Patti C, Phunkhang P, Pierre F, Priest M, Raghuraman S, Rege F, Reyes R, Rise C, Rogov P, Ross K, Ryan E, Settipalli S, Shea T, Sherpa N, Shi L, Shih D, Sparrow T, Spaulding J, Stalker J, Stange-Thomann N, Stavropoulos S, Stone C, Strader C, Tesfaye S, Thomson T, Thoulutsang Y, Thoulutsang D, Topham K, Topping I, Tsamla T, Vassiliev H, Vo A, Wangchuk T, Wangdi T, Weiand M, Wilkinson J, Wilson A, Yadav S, Young G, Yu Q, Zembek L, Zhong D, Zimmer A, Zwirko Z, Alvarez P, Brockman W, Butler J, Chin C, Grabherr M, Kleber M, Mauceli E, MacCallum I: **Evolution of genes and genomes on the Drosophila phylogeny.** *Nature* 2007, **450(7167):**203-218.

12. Swofford DL: **Phylogenetic Analysis Using Parsimony (*and other Methods).** 4th edition. Sunderland, Massachusetts: Sinauer Associates; 2002.

13. Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, Cravchik A, Collins FH, Robertson HM, Zwiebel LJ: **G protein-coupled receptors in Anopheles gambiae.** *Science* 2002, **298(5591):**176-178.

14. Kent L, Walden KK, Robertson HM: **The Gr family of candidate gustatory and olfactory receptors in the yellow-fever mosquito Aedes aegypti.** *Chemical Senses* 2008, **33(1):**79-93.

15. Wanner KW, Robertson HM: **The gustatory receptor (Gr) family in the silkworm moth Bombyx mori is characterized by a large expansion of a single lineage of putative bitter receptors.** *Insect Molecular Biology* 2008, **17:**621-629.

16. Richards S, Gibbs RA, Weinstock GM, Brown SJ, Denell R, Beeman RW, Gibbs R, Bucher G, Friedrich M, Grimmelikhuijzen CJ, Klingler M, Lorenzen M, Roth S, Schroder R, Tautz D, Zdobnov EM, Muzny D, Attaway T, Bell S, Buhay CJ, Chandrabose MN, Chavez D, Clerk-Blankenburg KP, Cree A, Dao M, Davis C, Chacko J, Dinh H, Dugan-Rocha S, Fowler G, Garner TT, Garnes J, Gnirke A, Hawes A, Hernandez J, Hines S, Holder M, Hume J, Jhangiani SN, Joshi V, Khan ZM, Jackson L, Kovar C, Kowis A, Lee S, Lewis LR, Margolis J, Morgan M, Nazareth LV, Nguyen N, Okwuonu G, Parker D, Ruiz SJ, Santibanez J, Savard J, Scherer SE, Schneider B, Sodergren E, Vattahil S, Villasana D, White CS, Wright R, Park Y, Lord J, Oppert B, Brown S, Wang L, Weinstock G, Liu Y, Worley K, Elsik CG, Reese JT, Elhaik E, Landan G, Graur D, Arensburger P, Atkinson P, Beidler J, Demuth JP, Drury DW, Du YZ, Fujiwara H, Maselli V, Osanai M, Robertson HM, Tu Z,

Wang JJ, Wang S, Song H, Zhang L, Werner D, Stanke M, Morgenstern B, Solovyev V, Kosarev P, Brown G, Chen HC, Ermolaeva O, Hlavina W, Kapustin Y, Kiryutin B, Kitts P, Maglott D, Pruitt K, Sapojnikov V, Souvorov A, Mackey AJ, Waterhouse RM, Wyder S, Kriventseva EV, Kadowaki T, Bork P, Aranda M, Bao R, Beermann A, Berns N, Bolognesi R, Bonneton F, Bopp D, Butts T, Chaumot A, Denell RE, Ferrier DE, Gordon CM, Jindra M, Lan Q, Lattorff HM, Laudet V, von Levetsow C, Liu Z, Lutz R, Lynch JA, da Fonseca RN, Posnien N, Reuter R, Schinko JB, Schmitt C, Schoppmeier M, Shippy TD, Simonnet F, Marques-Souza H, Tomoyasu Y, Trauner J, Zee M Van der, Vervoort M, Wittkopp N, Wimmer EA, Yang X, Jones AK, Sattelle DB, Ebert PR, Nelson D, Scott JG, Muthukrishnan S, Kramer KJ, Arakane Y, Zhu Q, Hogenkamp D, Dixit R, Jiang H, Zou Z, Marshall J, Elpidina E, Vinokurov K, Oppert C, Evans J, Lu Z, Zhao P, Sumathipala N, Altincicek B, Vilcinskas A, Williams M, Hultmark D, Hetru C, Hauser F, Cazzamali G, Williamson M, Li B, Tanaka Y, Predel R, Neupert S, Schachtner J, Verleyen P, Raible F, Walden KK, Angeli S, Foret S, Schuetz S, Maleszka R, Miller SC, Grossmann D: **The genome of the model beetle and pest Tribolium castaneum.** *Nature* 2008, **452(7190):**949-955.

17. Robertson HM, Wanner KW: **The chemoreceptor superfamily in the honey bee, Apis mellifera: expansion of the odorant, but not gustatory, receptor family.** *Genome Res* 2006, **16(11):**1395-1403.

18. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG: **The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Research* 1997, **25:**4876-4882.

19. Robertson HM: **The insect chemoreceptor superfamily in *Drosophila pseudoobscura*: molecular evolution of ecologically-relevant genes over 25 million years.** *Journal of Insect Science* 2008 in press.

20. Gardiner A, Barker D, Butlin RK, Jordan WC, Ritchie MG: **Drosophila chemoreceptor gene evolution: selection, specialization and genome size.** *Mol Ecol* 2008, **17(7):**1648-1657.

21. Besansky NJ, Fahey GT: **Utility of the white gene in estimating phylogenetic relationships among mosquitoes (Diptera: Culicidae).** *Mol Biol Evol* 1997, **14(4):**442-454.

22. Foley DH, Bryan JH, Yeates D, Saul A: **Evolution and systematics of *Anopheles*: Insights from a molecular phylogeny of Australasian mosquitoes.** *Molecular Phylogenetics and Evolution* 1998, **9(2):**262-275.

23. Gaunt MW, Miles MA: **An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks.** *Mol Biol Evol* 2002, **19(5):**748-761.

24. Krieger J, Raming K, Dewer YM, Bette S, Conzelmann S, Breer H: **A divergent gene family encoding candidate olfactory receptors of the moth Heliothis virescens.** *Eur J Neurosci* 2002, **16(4):**619-628.

25. Mita K, Kasahara M, Sasaki S, Nagayasu Y, Yamada T, Kanamori H, Namiki N, Kitagawa M, Yamashita H, Yasukochi Y, Kadono-Okuda K, Yamamoto K, Ajimura M, Ravikumar G, Shimomura M, Nagamura Y, Shin IT, Abe H, Shimada T, Morishita S, Sasaki T: **The genome sequence of silkworm, Bombyx mori.** *DNA Res* 2004, **11(1):**27-35.

26. Xia Q, Zhou Z, Lu C, Cheng D, Dai F, Li B, Zhao P, Zha X, Cheng T, Chai C, Pan G, Xu J, Liu C, Lin Y, Qian J, Hou Y, Wu Z, Li G, Pan M, Li C, Shen Y, Lan X, Yuan L, Li T, Xu H, Yang G, Wan Y, Zhu Y, Yu M, Shen W, Wu D, Xiang Z, Yu J, Wang J, Li R, Shi J, Li H, Su J, Wang X, Zhang Z, Wu Q, Li J, Zhang Q, Wei N, Sun H, Dong L, Liu D, Zhao S, Zhao X, Meng Q, Lan F, Huang X, Li Y, Fang L, Li D, Sun Y, Yang Z, Huang Y, Xi Y, Qi Q, He D, Huang H, Zhang X, Wang Z, Li W, Cao Y, Yu Y, Yu H, Ye J, Chen H, Zhou Y, Liu B, Ji H, Li S, Ni P, Zhang J, Zhang Y, Zheng H, Mao B, Wang W, Ye C, Wong GK, Yang H: **A draft sequence for the genome of the domesticated silkworm (Bombyx mori).** *Science* 2004, **306(5703):**1937-1940.

27. Robertson HM, Kent LB: **Evolution of the gene lineage encoding the carbon dioxide receptor in insects.** *Journal of Insect Science* 2008 in press.

28. Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR: **A Novel Family of Divergent Seven-Transmembrane Proteins: Candidate Odorant Receptors in *Drosophila*.** *Neuron* 1999, **22:**327-338.

29. Roy SW, Penny D: **On the incidence of intron loss and gain in paralogous gene families.** *Mol Biol Evol* 2007, **24(8):**1579-1581.

30. Benton R, Sachse S, Michnick SW, Vosshall LB: **Atypical membrane topology and heteromeric function of Drosophila odorant receptors in vivo.** *PLoS Biol* 2006, **4(2):**e20.
31. Wistrand M, Kall L, Sonnhammer EL: **A general model of G protein-coupled receptor sequences and its application to detect remote homologs.** *Protein Sci* 2006, **15(3):**509-521.
32. Lundin C, Kall L, Kreher SA, Kapp K, Sonnhammer EL, Carlson JR, Heijne G, Nilsson I: **Membrane topology of the Drosophila OR83b odorant receptor.** *FEBS Lett* 2007, **581(29):**5601-5604.
33. Smart R, Kiely A, Beale M, Vargas E, Carraher C, Kralicek AV, Christie DL, Chen C, Newcomb RD, Warr CG: **Drosophila odorant receptors are novel seven transmembrane domain proteins that can signal independently of heterotrimeric G proteins.** *Insect Biochem Mol Biol* 2008, **38(8):**770-780.
34. von Heijne G: **The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology.** *EMBO Journal* 1986, **5(11):**3021-3027.
35. von Heijne G, Gavel Y: **Topogenic signals in integral membrane proteins.** *European Journal of Biochemistry* 1988, **174:**671-678.
36. Arai M, Mitsuke H, Ikeda M, Xia JX, Kikuchi T, Satake M, Shimizu T: **ConPred II: a consensus prediction method for obtaining transmembrane topology models with high reliability.** *Nucleic Acids Res* 2004:W390-393.
37. Sato K, Pellegrino M, Nakagawa T, Vosshall LB, Touhara K: **Insect olfactory receptors are heteromeric ligand-gated ion channels.** *Nature* 2008, **452(7190):**1002-1006.
38. Wicher D, Schafer R, Bauernfeind R, Stensmyr MC, Heller R, Heinemann SH, Hansson BS: **Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels.** *Nature* 2008, **452(7190):**1007-1011.
39. Thorne N, Amrein H: **Atypical expression of Drosophila gustatory receptor genes in sensory and central neurons.** *J Comp Neurol* 2008, **506(4):**548-568.
40. Zhao GQ, Zhang Y, Hoon MA, Chandrashekar J, Erlenbach I, Ryba NJ, Zuker CS: **The receptors for mammalian sweet and umami taste.** *Cell* 2003, **115(3):**255-266.
41. Sorek R: **The birth of new exons: mechanisms and evolutionary consequences.** *RNA* 2007, **13(10):**1603-1608.