

## **Evolution of the Gene Lineage Encoding the Carbon Dioxide Receptor in Insects**

Authors: Robertson, Hugh M., and Kent, Lauren B.

Source: Journal of Insect Science, 9(19) : 1-14

Published By: Entomological Society of America

URL: <https://doi.org/10.1673/031.009.1901>

---

The BioOne Digital Library (<https://bioone.org/>) provides worldwide distribution for more than 580 journals and eBooks from BioOne's community of over 150 nonprofit societies, research institutions, and university presses in the biological, ecological, and environmental sciences. The BioOne Digital Library encompasses the flagship aggregation BioOne Complete (<https://bioone.org/subscribe>), the BioOne Complete Archive (<https://bioone.org/archive>), and the BioOne eBooks program offerings ESA eBook Collection (<https://bioone.org/esa-ebooks>) and CSIRO Publishing BioSelect Collection (<https://bioone.org/csiro-ebooks>).

Your use of this PDF, the BioOne Digital Library, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at [www.bioone.org/terms-of-use](http://www.bioone.org/terms-of-use).

Usage of BioOne Digital Library content is strictly limited to personal, educational, and non-commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

---

BioOne is an innovative nonprofit that sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.



## Evolution of the gene lineage encoding the carbon dioxide receptor in insects

Hugh M. Robertson and Lauren B. Kent

Department of Entomology, University of Illinois at Urbana-Champaign, 505 S. Goodwin Ave., Urbana, IL 61801, USA

### Abstract

A heterodimer of the insect chemoreceptors Gr21a and Gr63a has been shown to be the carbon dioxide receptor in *Drosophila melanogaster* (Meigen) (Diptera: Drosophilidae). Comparison of the genes encoding these two proteins across the 12 available drosophilid fly genomes allows refined definition of their N-termini. These genes are highly conserved, along with a paralog of Gr21a, in the *Anopheles gambiae*, *Aedes aegypti*, and *Culex pipiens* mosquitoes, as well as in the silk moth *Bombyx mori* and the red flour beetle *Tribolium castaneum*. In the latter four species we name these three proteins Gr1, Gr2, and Gr3. Intron evolution within this distinctive three gene lineage is considerable, with at least 13 inferred gains and 39 losses. Surprisingly, this entire ancient gene lineage is absent from all other available more basal insect and related arthropod genomes, specifically the honey bee, parasitoid wasp, human louse, pea aphid, waterflea, and blacklegged tick genomes. At least two of these species can detect carbon dioxide, suggesting that they evolved other means to do so.

**Keywords:** olfaction, gustatory receptor, smell, intron evolution

**Correspondence:** hughrobe@uiuc.edu, lkent@uiuc.edu

**Received:** 3 September 2007 | **Accepted:** 23 February 2008 | **Published:** 13 May 2009

**Copyright:** This is an open access paper. We use the Creative Commons Attribution 3.0 license that permits unrestricted use, provided that the paper is properly attributed.

**ISSN:** 1536-2442 | Vol. 9, Number 19

#### Cite this paper as:

Robertson HM, Kent LB. 2009. Evolution of the gene lineage encoding the carbon dioxide receptor in insects. 14pp. *Journal of Insect Science* 9:19, available online: [insectscience.org/9.19](http://insectscience.org/9.19)

## Introduction

Many insects are capable of sensitive detection of elevated levels of carbon dioxide (CO<sub>2</sub>) (e.g. Stange 1996; Stange and Stowe 1999). Perhaps most famously, many hematophagous insects and other arthropods such as mosquitoes (e.g. Gillies 1980; Bowen 1991), tsetse flies (e.g. Vale and Hall 1985; Gibson and Torr 1999), and ticks (e.g. Holscher et al. 1980) use elevated CO<sub>2</sub> concentrations to locate their vertebrate hosts. Several moths employ CO<sub>2</sub> gradients in evaluating floral quality (e.g. Stange et al. 1995; Thom et al. 2004), while social insects like honeybees and ants regulate potentially lethal CO<sub>2</sub> concentrations in their social colonies (e.g. Seeley 1974; Kleineidam and Tautz 1996). The importance of carbon dioxide perception in other flies like *Drosophila melanogaster* (Meigen) (Diptera: Drosophilidae) has also been recognized (de Bruyne et al. 2001), where an increase in CO<sub>2</sub> released by stressed flies elicits avoidance behavior (Suh et al. 2004), and with this came the potential to discover the molecular nature of the elusive carbon dioxide receptor. DmGr21a is expressed in the ab1C olfactory receptor neurons (ORNs) in the ab1 sensilla on the antennae of *D. melanogaster* that are sensitive to CO<sub>2</sub> (de Bruyne et al. 2001; Suh et al. 2004), thereby implicating this member of the nominal gustatory receptor family within the insect chemoreceptor superfamily (Clyne et al. 2000; Scott et al. 2001; Dunipace et al. 2001; Robertson et al. 2003). Recently Jones et al. (2007) and Kwon et al. (2007) reported that the related receptor DmGr63a is also expressed in these neurons, and showed that both are required for detection of carbon dioxide when mis-expressed in an ORN without receptors. This breakthrough in identification of the heterodimeric nature of the carbon dioxide receptor in *Drosophila* flies has general implications, because this pair of Grs is highly conserved in the African malaria vector *Anopheles gambiae* mosquito, where they were named AgGr22 and AgGr24 (Hill et al. 2002). Indeed, Jones et al. (2007) show that AgGr22 and AgGr24 are also co-expressed in a set of ORNs, but this time in sensilla on the maxillary palps of *Anopheles* mosquitoes, the organ known to mediate their perception of carbon dioxide (e.g. Bowen 1991). Finally, Lu et al. (2007) show that this pair of receptors is sufficient for CO<sub>2</sub> perception in this mosquito.

Here we describe the evolution of these two Gr genes and proteins, and a third related gene/protein that was lost from the drosophilid fly lineage, in the available endopterygote (holometabolous) insect genomes. This third gene is also expressed in the same set of ORNs in sensilla on the palps of *Anopheles* mosquitoes and greatly improves the sensitivity to CO<sub>2</sub> when co-expressed with the other two proteins in *D. melanogaster* neurons (Lu et al. 2007). Remarkably, this entire three-gene lineage is not present in the available more basal hymenopteran genome sequences (honey bee and parasitoid wasp), nor is it present in any other available more basal insect or related

arthropod genome sequences, despite the ability of at least two of these arthropods to detect carbon dioxide.

## Materials and Methods

The Gr21a and Gr63a orthologs were retrieved from the 12 *Drosophila* genome sequences available at FLYBASE and GENBANK, as were orthologs of the AgGr22-24 genes from the *Aedes aegypti*, *Bombyx mori*, and *Tribolium castaneum* genome assemblies, using TBLASTN searches. All other available unpublished draft arthropod genome sequences were also searched using TBLASTN, as were the entire sets of raw reads at the Trace Archive using the NCBI BLAST CLIENT software. Gene models were built manually in the text editor of PAUP\*v4.0b10 (Swofford 2002). The gene models for BmGr1 and BmGr2 could not be confidently completed for their N-terminus because the expected upstream N-terminal coding exon(s) could not be identified due to the expected high divergence of the encoded amino acids. The N-terminal region of BmGr3 was constructed from a combination of the available contigs from both the Japanese and Chinese genome assemblies, both of which have frameshifting errors. The amino acid sequences are available in the Supplement.

Proteins were aligned in CLUSTALX (Jeanmougin et al. 1998) using default parameters, and phylogenetic analysis was performed using corrected distance methods. Corrected distances were calculated in TREE-PUZZLE v5.0 (Schmidt et al. 2002) using the BLOSUM62 amino acid matrix in their maximum likelihood model, and distance trees were estimated in PAUP\*v4.0b10 using tree-bisection-and-reconnection branch swapping. Support for branches was obtained from 1000 bootstrap replications of uncorrected distance analysis. Inferred intron gains and losses were mapped onto the phylogenetic tree using simple parsimony and equal weighting of gains and losses. Kyte-Doolittle hydropathy plots were produced in DNA Strider v1.1 (Marck 1988).

## Results and Discussion

### Gr21a/63a in other drosophilid flies

Highly conserved orthologs of the DmGr21a and DmGr63a genes are present in the *Drosophila pseudoobscura* genome sequence (Richards et al. 2005; Robertson 2009) and the other 10 newly available *Drosophila* genome sequences (Drosophila 12 Genomes Consortium 2007), as expected from the presence of conserved orthologs in *An. gambiae* (Hill et al. 2002). Alignments of the genes and their encoded protein products in these 12 *Drosophila* species allowed refinement of their uncertain N-termini. Both DmGr21a and DmGr63a have extended potential ORFs in their first coding exons beyond a potential start codon that is conserved in all the other species, adding 47 and 23 additional amino acids respectively. These N-

terminal extensions were originally annotated as part of these proteins in the absence of comparative information, and at least for DmGr21a are supported by the existence of a single full-length cDNA clone from the Berkeley Drosophila Genome Project (BDGP) (GenBank BT025007.1), which has a 674 bp 5' UTR in front of this most upstream available start codon.

For Gr21a this possible 47 amino acid extension is present in the sibling species *D. simulans* and *D. sechellia*, as well as the more distantly related *D. yakuba*. There is, however, a single base pair insertional frameshift in this extension in *D. erecta* (confirmed by all eight reads that cover this region), which eliminates the possibility of this N-terminal extension being functional in this species. No such extension is possible for *D. ananassae* in which this upstream sequence is considerably diverged, nor is any extension possible for the more divergent *D. willistoni*, *D. mojavensis*, *D. virilis*, or *D. grimshawi* genes. Remarkably, an N-terminal extension of 55 amino acids is also possible for *D. pseudoobscura*, however the amino acid sequence is completely different from the *D. melanogaster* extension. Furthermore, although the DNA sequence in this region is similar in the sibling species *D. persimilis*, a 2 bp insertional frameshift eliminates the possibility of it being functional in this species, and suggests that it is unlikely to be functional in *D. pseudoobscura* either.

For Gr63a the possible 23 amino acid N-terminal extension is similarly shared with the sibling species *D. simulans* and *D. sechellia*, as well as both *D. yakuba* and *D. erecta*, however in *D. simulans* there is an alternate allele represented by two of the ten reads that cover the region which has a four-base frameshifting deletion in this extension. Again, no extension is possible for the more divergent *D. ananassae*, *D. willistoni*, *D. mojavensis*, *D. virilis*, or *D. grimshawi* genes, however once again *D. pseudoobscura* and *D. persimilis* share a possible extension of 11 amino acids of completely different sequence from DmGr63a.

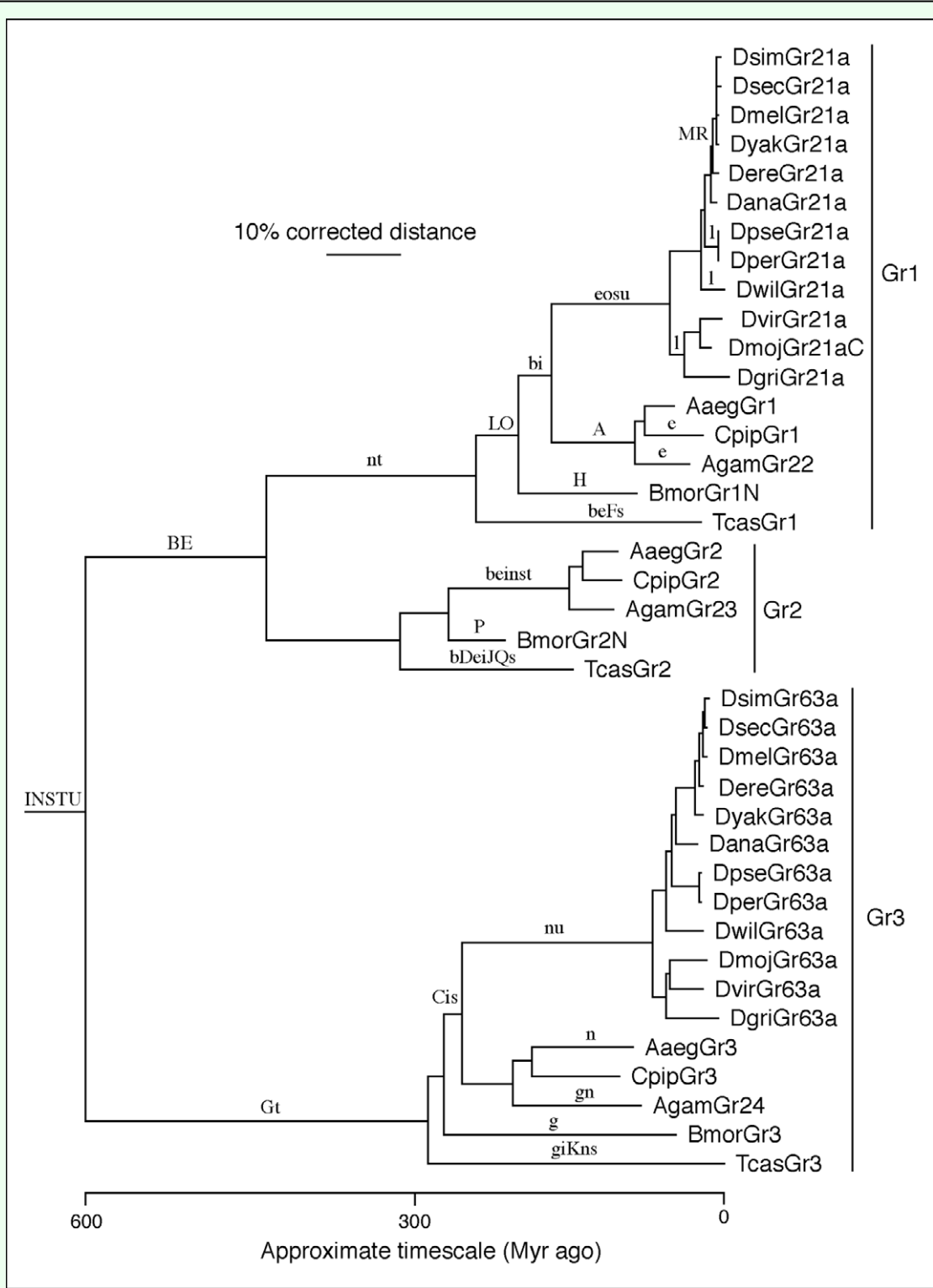
Given the high conservation of the rest of these proteins, we believe these N-terminal extensions are not present in these proteins in these flies, thus the true N-terminus of DmGr21a is likely to be MSFWAV and that of DmGr63a is MANYR (full amino acid sequences for all species are in the Appendix). If correct, this would require that the translation machinery ignore the first two available AUG start codons in the extended ORF of DmGr21a, but this seems feasible as it must already ignore another 14 potential AUG start codons in the 674 bp 5' UTR. Thus two of the versions of this protein available in GenBank likely have incorrectly elongated N-termini. Accession ABE01237.1 encoding "IP03362p" is from the BDGP cDNA with 47 additional amino acids on the N-terminus, while ABK97615.1 encoding "gustatory receptor 21a" is from Jones et al. (2007) and appears not to represent an experimentally determined cDNA but rather is a conceptual coding sequence from

the start to stop codons and encodes eight additional amino acids on the N-terminus. A third version in GenBank, represented by accession AAF51461.2, is our conceptual CDS from FlyBase, "CG13948-PA". Unfortunately the only way to confirm our inference would be to obtain N-terminal peptide sequence from these proteins as they are expressed in fly antennal neurons. It is also possible that these N-terminal extensions are sometimes present and/or do not affect function of the proteins, after all, the slightly extended version was employed by Jones et al. (2007) in their experiments identifying these proteins as a heterodimeric receptor for carbon dioxide.

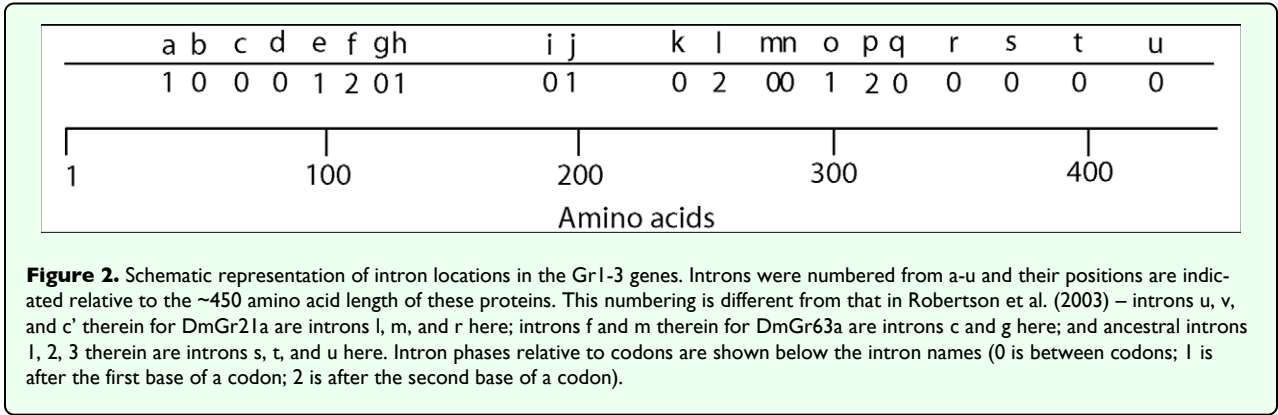
Another somewhat unusual feature of these *Drosophila* genes is the different introns in the Gr21a lineage (the Gr63a genes in these drosophilids all have the same two introns). The divergent lineage of *D. mojavensis/virilis/grimshawi*, as well as *D. willistoni* and the *D. pseudoobscura/persimilis* pair have no introns, while *D. ananassae* has one intron, and the *melanogaster* subgroup species (*D. yakuba/erecta/simulans/sechellia/melanogaster*) have two more introns. This interesting intron evolution is explored further below.

### Three genes in mosquitoes, silk moth, and flour beetle

In addition to the conserved orthologs of DmGr21a and DmGr63a in *An. gambiae* (AgGr22 and AgGr24), the *An. gambiae* genome contains a sister gene for DmGr21a/AgGr22, named AgGr23 (Hill et al. 2002). Like the DmGr21a/AgGr22 and DmGr63a/AgGr24 proteins, this third gene/protein lineage is also highly conserved in the yellow fever mosquito *Aedes aegypti* (Kent et al. 2008), the house mosquito *Culex pipiens* (HMR unpublished results), the silkworm *Bombyx mori* (Wanner and Robertson 2008), and the red flour beetle *Tribolium castaneum* (Tribolium Genome Sequencing Consortium 2008) (Figure 1). Although Lu et al. (2007) include the *Bombyx* and *Tribolium* proteins in their supplementary material and name them BmGr22-24 and TcGr22-24, we feel this naming convention is awkward, and propose to name these genes Gr1-3 in each of these species, in recognition of the fact that they are the only highly conserved and strictly orthologous lineages of Grs across these insects that are as old as 270 million years. We hope that this naming convention will be employed in future identifications of these three genes in non-drosophilid insects to minimize the nomenclatural confusion already engendered by their naming in *D. melanogaster* by cytological location (Clyne et al. 2000; Scott et al. 2001; Dunipace et al. 2001; Robertson et al. 2003) and in *An. gambiae* according to their order of discovery (Hill et al. 2002). Alignment of these three proteins from all these insects reveals that while their N- and C-terminal regions have different lengths and rather divergent sequences, the core transmembrane regions are well conserved and fully alignable, with at least 25% amino acid identity. The



**Figure 1.** Phylogenetic relationships of the Gr1-3 proteins in the twelve drosophilid flies, three mosquitoes, the silk moth, and the flour beetle. This is a corrected distance tree so cumulative distances along the longest branches total as much as 185%, while the actual distances are never larger than 78%. Species names are abbreviated to the first letter of the genus in capitals and the first three letters of the species name in lower case. The extreme N-termini of the BmorGr1 and BmorGr2 proteins could not be identified in the genome sequence, indicated by the suffix N after the protein name, but this does not affect the alignment used here. Similarly, a few amino acids are missing from the DmojGr21a C-terminus. Bootstrap support from 1000 replications of full heuristic uncorrected distance analysis was over 75% for all nodes except within the drosophilid flies. Inferred intron gains and losses are indicated by upper and lower case letters, respectively, on the relevant branches. Note that intron t, although only present in the Gr2 lineage, is inferred to be a much older intron because it is shared by other Grs (see Figure 2 legend).



only minor length differences are in five of the intra- and extra-cellular loops between the trans-membrane domains (the last, apparently intra-cellular loop, between TM6 and TM7 does not vary in length). The first and last TM domains are the most highly conserved in amino acid sequence, as is true for the entire Gr family (Clyne et al. 2000). From the results of Lu et al. (2007) who expressed all three proteins in an “empty neuron” system in *D. melanogaster* antennae, it appears that all three contribute to the detection of CO<sub>2</sub> in other insects.

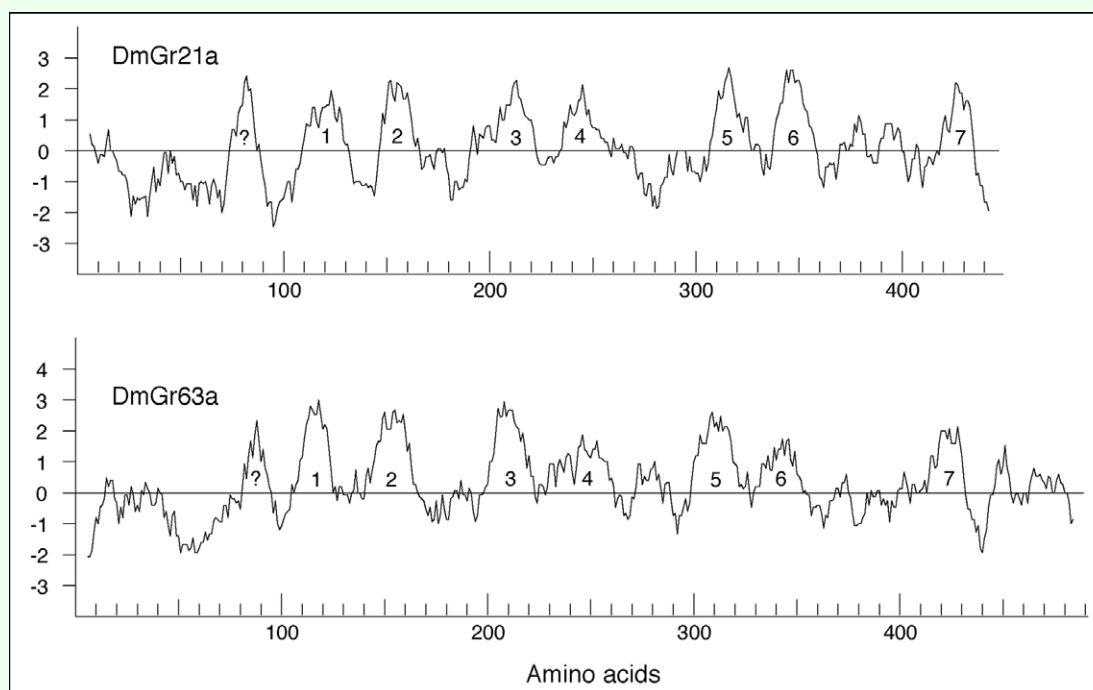
**Intron evolution**

As noted above, the intron content of the Gr21a gene lineage in the drosophilid flies is unusual. We attempted to place these introns in a broader context by considering the presence and absence of introns across all three gene lineages in all of these insects. Remarkably there are 21 different intron placements in the coding regions of this set of three genes (Figure 2; Appendix). It seems unlikely that the common ancestor of these three genes contained all 21 introns, indeed many are quite close to each other (e.g. introns m and n are one codon apart). Inference of intron losses and gains is clearly evident within many gene lineages, both when traced through species history (e.g. Roy and Gilbert 2005), and when traced in large paralogous gene families (e.g. Robertson 1998, 2000; Roy and Penny 2007). Indeed when considering the molecular evolution of the entire chemoreceptor superfamily as represented in *D. melanogaster*, Robertson et al. (2003) concluded that ~57 introns gains had occurred in the superfamily and ~48 intron losses were postulated to explain the current distribution of introns in the ~120 genes.

Mapping of the intron locations on the phylogenetic tree of the genes in Figure 1 leads to the inference that the common ancestral gene lineage had at least five introns. This is two more than was inferred for this branch in Robertson et al. (2003), suggesting that the inference therein of just three ancestral introns for the entire superfamily was an underestimate. Indeed, given the 2-3-fold excess of intron losses over intron gains (see below), introns b, e, and g might also be ancestral to the gene

lineage and subsequently lost in the Gr3 or Gr1/2 lineages, respectively. If b, e, and g are indeed older then the common ancestor gene would have had eight introns roughly equally spaced along its length. Thirteen intron gains (conservatively assuming that b, e, and g are older) are postulated to explain the current distribution of introns in these three genes, including most remarkably the two novel introns present in the Gr21a genes in the *melanogaster* subgroup species (introns m and r). One of these (m) is just one codon apart from intron n, and an alternative explanation might be that these are the same intron which has “slid” one codon in this highly conserved region of the gene/protein, however this model would require postulation of seven independent losses of the intron in the Gr21a lineage from beetles up to the *melanogaster* species grouping, which seems unlikely. Nevertheless, the complete absence of introns from this lineage in most of the drosophilids requires postulation of three independent losses of intron l.

Recent intron gains are unusual in animal genomes (e.g. Roy and Gilbert 2005). This three-gene lineage appears to have acquired at least 13 introns in the past ~600 Myr, for a rate of approximately 2.5 per billion years across all gene and species lineages (a total of roughly 5 billion years of evolution). Given an average coding length of 1200 bp, this is roughly  $2 \times 10^{-12}$  gains per possible insertion site per year, which is within the range of rates calculated by Roy and Gilbert (2005). The two most recent gains of introns m and r in the Gr21a gene at the base of the *melanogaster* species group are nevertheless at least 10, and as much as 20, Myr old (the time between the split of *D. ananassae* and the split of *D. yakuba/erecta*), and their short sequences of 53–57 bp are only weakly similar across these five species and bear no convincing similarity to other sequences in these genomes that might indicate their origin. Unfortunately the rapid rate of neutral evolution in these short-generation flies has obliterated any hint of the origins of these introns, and even global analyses of intron evolution in these fly genomes do not reveal additional intron gains



**Figure 3.** Kyte-Doolittle hydropathy plots for DmGr21a and DmGr63a. The seven transmembrane domains that align well throughout the Gr and Or families are numbered and the potential eighth TM domain at the N-terminus is indicated with a question mark.

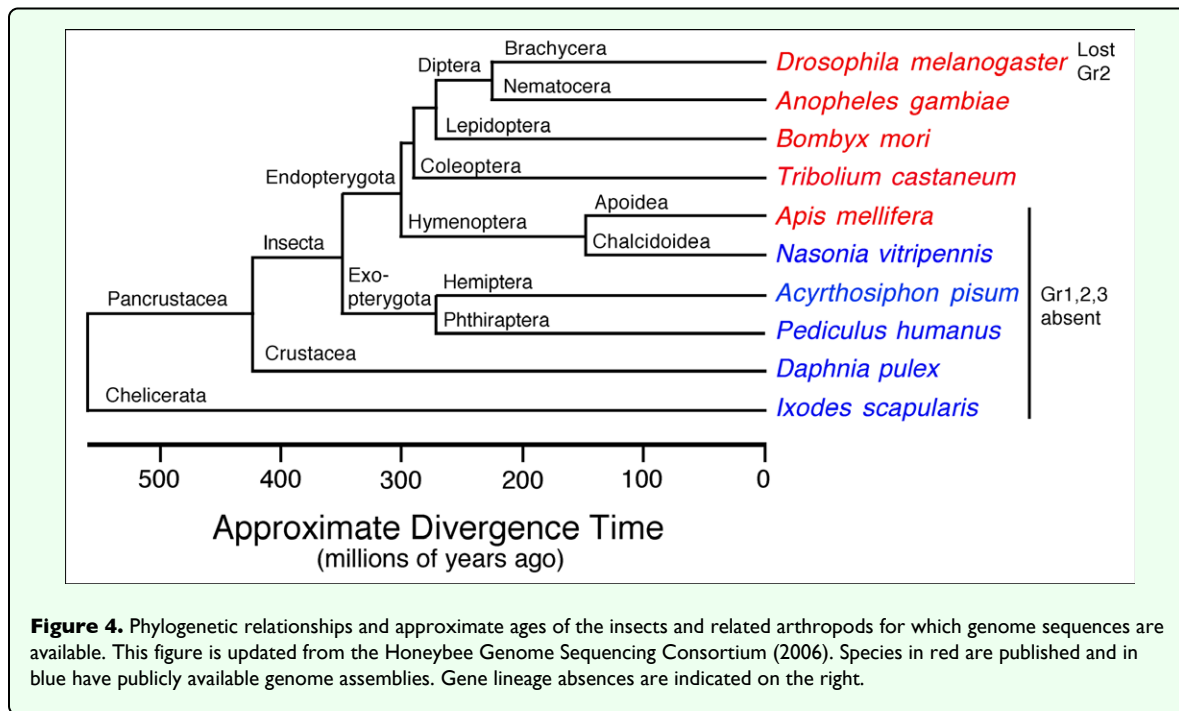
whose origin is discernable (Coulombe-Huntington and Majewski 2007).

At least thirty-nine intron losses are postulated on the tree, in keeping with many other estimates of considerably higher rates of intron loss than gain (e.g. Roy and Gilbert 2005; Roy and Penny 2007). The roughly equal rates of intron gain and loss estimated across the entire superfamily in *D. melanogaster* (Robertson et al. 2003) is likely to have resulted from underestimates of intron loss given the poor resolution of the superfamily tree and failure to recognize that several more introns might be ancestral to the superfamily. There appears to be a considerably higher rate of intron losses on the branches leading to the dipteran and beetle genes, with just one loss ascribed to a moth gene lineage. This bias towards intron loss in the dipteran and beetle genomes is consistent with their smaller genome size, and has been noted in genome-wide comparisons for the Diptera (e.g. Raible et al. 2005). As a result, no gene retains all five (or eight if b, e, and g are older) ancestral introns.

### Secondary structure

The insect chemoreceptor proteins had been considered to be members of a novel class of seven-transmembrane (7TM) G-protein-coupled receptors (GPCRs) (e.g. Hill et al. 2002; Benton et al. 2006), however their complete sequence divergence from all other known 7TM GPCR classes and their apparent functioning in extremely heterologous expression systems, e.g. frog oocytes (e.g. Wanner et al. 2007), has suggested that they might be a

completely different class of membrane proteins. This possibility was supported by the finding of Benton et al. (2006) that the membrane topology of two odorant receptors is the reverse of that expected of a GPCR, with the N-terminus intracellular, a result confirmed for DmOr83b by Lundin et al. (2007). Wistrand et al. (2006) came to a similar conclusion, finding that the insect Ors and Grs do not have a membrane topology typical of all the other GPCR classes, indeed using the “positive-inside” rule of von Heijne (1989) they find that they likely have the opposite membrane polarity of the GPCRs. Benton et al. (2006) and Wistrand et al. (2006) did find that most *Drosophila* Ors have seven predicted TM domains, but some have eight. Kyte-Doolittle hydropathy plot analysis of many members of the Gr family, including DmGr21a and DmGr63a (Figure 3), reveals that they may have eight transmembrane domains, most of which are also recognized as potential TM domains by several TM prediction programs, including DAS-TMfilter (Cserzo et al. 2002) and PolyPhobius (Käll et al. 2005) with the first and last candidate TM domains being somewhat equivocal in the various Gr1-3 proteins. This ambiguity is particularly well displayed by the ConPredII server which uses the results of nine different prediction programs including the above two and predicts seven or eight TM domains for these proteins, with roughly half having the N-terminus inside the cell (Arai et al. 2004). If they indeed have eight TM domains, then the N-terminus could be internal yet the remaining membrane topology would be the same as GPCRs, however that



would not fit with the findings of Wistrand et al. (2006) about the “positive-inside” rule. ConPredII calculates results for the “positive-inside” rule and for most of these Gr1-3 proteins strongly suggests the opposite topology to that of the GPCRs, in agreement with Wistrand et al. (2007). Resolution of this conundrum of the secondary structure and membrane topology of these insect chemoreceptors will require additional experimental study of both Ors and Grs.

#### Absence of this gene lineage from other available insect and arthropod genomes

This three-gene lineage is entirely absent from the genome of the honey bee, *Apis mellifera*, representing the more basal Hymenoptera (Robertson and Wanner 2006). It is possible that a draft genome sequence might be missing a few genes, perhaps because they reside in poorly cloned, sequenced, and assembled heterochromatic domains like the pericentromeric regions, however it is unlikely that all three of these genes would be absent from the draft genome assembly, which is of high quality (Honeybee Genome Sequencing Consortium 2006; Robertson et al. 2007). Furthermore, we searched all the raw traces from this genome using TBLASTN without finding a single read that encoded amino acid sequence with convincing matches to any of these three proteins. Honey bees have receptor neurons in their antennae that can detect carbon dioxide (Lacher 1964, cited in Winston 1991), and monitor its levels in their hives, responding to elevated levels by aerating the hive (e.g. Seeley, 1974).

Furthermore, examination of all other available basal insect and related arthropod genome sequences by TBLASTN searches of their publicly available draft

assemblies, as well as all raw reads available in the Trace Archive at NCBI, reveals that this entire gene lineage is also missing from all of them (parasitoid wasp *Nasonia vitripennis*, pea aphid *Acyrtosiphon pisum*, body louse *Pediculus humanus*, waterflea *Daphnia pulex*, and blacklegged tick *Ixodes scapularis*) (Figure 4). The conservation of this protein lineage implies that it should be present in most of the above lineages, most of which are younger than 500 Myr (e.g. Glenner et al. 2006). Indeed, when an approximate timescale is plotted along the base of the tree in Figure 1, calibrated by the 250 Myr estimate for the split of the *Drosophila* flies and the mosquitoes within the Diptera, it appears that at least two, and perhaps all three, genes should be present in most or all of these arthropods. It seems unlikely that all three proteins independently became highly conserved only after the divergence of the Hymenoptera from the basal insect lineage. Instead this protein lineage likely exists in at least some basal insects and perhaps related arthropods, but was independently lost from each of the species currently targeted for genome sequencing.

There is little information on the ability of *Nasonia* wasps, *Acyrtosiphon* aphids, *Pediculus* lice, or *Daphnia* waterfleas, or related species, to detect carbon dioxide (but see Stross 1971). Indeed the biology of each of these arthropod groups suggests that they might no longer need to detect carbon dioxide and simply lost these receptors. However, like other ticks the blacklegged tick uses CO<sub>2</sub> as a cue for the presence of vertebrate hosts (e.g. Holscher et al. 1980; Schulze et al. 1997; McMahon and Guerin 2002), so like honey bees they must use another method to detect carbon dioxide. This might involve other members

of the chemoreceptor superfamily or it might involve a quite different mechanism like that reported in mammals (Hu et al. 2007) or in the gustatory system of *Drosophila* flies (Fischler et al. 2007). In either case it is remarkable that multiple methods of perceiving carbon dioxide appear to have evolved in arthropods.

## Acknowledgments

We thank the five US genome sequencing centers for making raw traces and draft assemblies available for the ten drosophilid and other insect and arthropod genomes before publication, and Scott Roy and two anonymous reviewers for comments on draft versions of the manuscript. This work was funded by NIH grant AI56081.

## References

- Arai M, Mitsuke H, Ikeda M, Xia JX, Kikuchi T, Satake M, Shimizu T. 2004. ConPred II: a consensus prediction method for obtaining transmembrane topology models with high reliability. *Nucleic Acids Research* 32 (Web Server issue): W390-293.
- Benton R, Sachse S, Michnick SW, Vosshall LB. 2006. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors in vivo. *PLoS Biology* 4(2): e20
- Bowen MF. 1991. The sensory physiology of host-seeking behavior of mosquitoes. *Annual Review of Entomology* 36: 139-158.
- Clyne PJ, Warr CG, Carlson JR. 2000. Candidate taste receptors in *Drosophila*. *Science* 287(5459): 1830-1834.
- Coulombe-Huntington J, Majewski J. 2007. Intron loss and gain in *Drosophila*. *Molecular Biology and Evolution* 24(12): 2842-2850.
- Cserzo M, Eisenhaber F, Eisenhaber B, Simon I. 2002. On filtering false positive transmembrane protein predictions. *Protein Engineering* 15(9): 745-752.
- de Bruyne M, Foster K, Carlson JR. 2001. Odor coding in the *Drosophila* antenna. *Neuron* 30(2): 537-552.
- Drosophila* 12 Genomes Consortium 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450(7167): 203-218.
- Dunipace L, Meister S, McNealy C, Amrein H. 2001. Spatially restricted expression of candidate taste receptors in the *Drosophila* gustatory system. *Current Biology* 11(11): 822-835.
- Fischler W, Kong P, Marella S, Scott K. 2007. The detection of carbonation by the *Drosophila* gustatory system. *Nature* 448(7157): 1054-1057.
- Gibson G, Torr SJ. 1999. Visual and olfactory responses of haematophagous Diptera to host stimuli. *Medical and Veterinary Entomology* 13(1): 2-23.
- Gillies MT. 1980. The role of carbon dioxide in host-finding by mosquitoes (Diptera: Culicidae): a review. *Bulletin of Entomological Research* 70(4): 525-532.
- Glenner H, Thomsen PF, Hebsgaard MB, Sorensen MV, Willerslev E. 2006. The origin of insects. *Science* 314(5807): 1883-1884.
- Hill CA, Fox AN, Pitts RJ, Kent LB, Tan PL, Chrystal MA, Cravchik A, Collins FH, Robertson HM, Zwiebel IJ. 2002. G protein-coupled receptors in *Anopheles gambiae*. *Science* 298(5591): 176-178.
- Holscher KH, Gearhart HL, Barker RW. 1980. Electrophysiological responses of three tick species to carbon dioxide in the laboratory and field. *Annals of the Entomological Society of America* 73(3): 288-292.
- Honeybee Genome Sequencing Consortium 2006. Insights into social insects from the genome of the honeybee *Apis mellifera*. *Nature* 443(7114): 931-949.
- Hu J, Zhong C, Ding C, Chi Q, Walz A, Mombaerts P, Matsunami H, Luo M. 2007. Detection of near-atmospheric concentrations of CO<sub>2</sub> by an olfactory subsystem in the mouse. *Science* 317(5840): 953-957.
- Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ. 1998. Multiple sequence alignment with Clustal X. *Trends in Biochemical Sciences* 23(10): 403-405.
- Jones WD, Cayirlioglu P, Kadow IG, Vosshall LB. 2007. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila*. *Nature* 445(7123): 86-90.
- Käll L, Krogh A, Sonnhammer EL. 2005. An HMM posterior decoder for sequence feature prediction that includes homology information. *Bioinformatics* 21(Suppl. 1): i251-257.
- Kent LB, Walden KKO, Robertson HM. 2008. The Gr family of olfactory and gustatory receptors in the yellow fever mosquito, *Aedes aegypti*. *Chemical Senses* 33(1): 79-93.
- Kleineidam C, Tautz J. 1996. Perception of carbon dioxide and other "air-condition" parameters in the leaf cutting ant *Atta cephalotes*. *Naturwissenschaften* 83(12): 566-568.
- Kwon JY, Dahanukar A, Weiss LA, Carlson JR. 2007. The molecular basis of CO<sub>2</sub> reception in *Drosophila*. *Proceedings of the National Academy of Sciences USA* 104(9): 3574-3578.
- Lacher V. 1964. Elektrophysiologische Untersuchungen an einzelnen Rezeptoren für Geruch, Kohlendioxyd, Luftfeuchtigkeit und Temperatur auf den Antennen der Arbeitsbiene und Dohne (*Apis mellifera* L.). *Zeitschrift fuer Vergleichende Physiologie* 48(6): 587-623.
- Lu T, Qiu YT, Wang G, Kwon JY, Rutzler M, Kwon H-W, Pitts RJ, van Loon JJA, Takken W, Carlson JR, Zwiebel IJ. 2007. Odor coding by maxillary palp neurons of the malaria vector mosquito *Anopheles gambiae*. *Current Biology* 17(18): 1533-1544.
- Lundin C, Käll L, Kreher SA, Kapp K, Sonnhammer EL, Carlson JR, Heijne G, Nilsson I. 2007. Membrane topology of the *Drosophila* OR83b odorant receptor. *FEBS Letters* 581(29): 5601-5604.
- Marck C. 1988. 'DNA Strider' a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. *Nucleic Acids Research* 16(5): 1829-1836.
- McMahon C, Guerin PM. 2002. Attraction of the tropical bont tick, *Amblyomma variegatum*, to human breath and to the breath components acetone, NO and CO<sub>2</sub>. *Naturwissenschaften* 89(7): 311-315.

- Raible F, Tessmar-Raible K, Osoegawa K, Wincker P, Jubin C, Balaovine G, Ferrier D, Benes V, de Jong P, Weissenbach J, Bork P, Arendt D. 2005. Vertebrate-type intron-rich genes in the marine annelid *Platynereis dumerilii*. *Science* 310(5752): 1325-1326.
- Richards S, Liu Y, Bettencourt BR, Hradecky P, Letovsky S, Nielsen R, Thornton K, Hubisz MJ, Chen R, Meisel RP, Couronne O, Hua S, Smith MA, Zhang P, Liu J, Bussemaker HJ, van Batenburg MF, Howells SL, Scherer SE, Sodergren E, Matthews BB, Crosby MA, Schroeder AJ, Ortiz-Barrientos D, Rives CM, Metzker ML, Muzny DM, Scott G, Steffen D, Wheeler DA, Worley KC, Havlak P, Durbin KJ, Egan A, Gill R, Hume J, Morgan MB, Miner G, Hamilton C, Huang Y, Waldron L, Verduzco D, Clerc-Blankenburg KP, Dubchak I, Noor MA, Anderson W, White KP, Clark AG, Schaeffer SW, Gelbart W, Weinstock GM, Gibbs RA. 2005. Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Research* 15(1): 1-18.
- Robertson HM. 1998. Two large families of chemoreceptor genes in the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae* reveal extensive gene duplication, diversification, movement, and intron loss. *Genome Research* 8(5): 449-463.
- Robertson HM. 2000. The large *srh* family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Research* 10(2): 192-203.
- Robertson HM. 2009. The insect chemoreceptor superfamily in *Drosophila yakuba* and *Drosophila pseudoobscura*: molecular evolution of ecologically-relevant genes over 10–25 million years. *Journal of Insect Science* 9:18, available online at <http://insectscience.org/9.18>
- Robertson HM, Reese JT, Milshina NV, Agarwala R, Solignac M, Walden KKO, Elisk CG. 2007. Manual superscaffolding of honey bee (*Apis mellifera*) chromosomes 12–16: implications for the draft genome assembly version 4, gene annotation, and chromosome structure. *Insect Molecular Biology* 16(4): 401-410.
- Robertson HM, Wanner KW. 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 HM, Warr CG, Carlson JR. 2003. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences USA* 100(Suppl. 2): 14537-14542.
- Roy SW, Gilbert W. 2005. Rates of intron loss and gain: implications for early eukaryotic evolution. *Proceedings of the National Academy of Sciences USA* 102(16): 5773-5778.
- Roy SW, Penny D. 2007. On the incidence of intron loss and gain in paralogous gene families. *Molecular Biology and Evolution* 24(8): 1579-1581.
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 18(3): 502-504.
- Schulze TL, Jordan RA, Hung RW. 1997. Biases associated with several sampling methods used to estimate abundance of *Ixodes scapularis* and *Amblyomma americanum* (Acari: Ixodidae). *Journal of Medical Entomology* 34(6): 615-623.
- 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.
- Seeley TD. 1974. Atmospheric carbon dioxide regulation in honey-bee (*Apis mellifera*) colonies. *Journal of Insect Physiology* 20(11): 2301-2305.
- Stange G. 1996. Sensory and behavioural responses of terrestrial invertebrates to biogenic carbon dioxide gradients. In: Stanhill G, editor. *Advances in Bioclimatology* 4, pp. 223-253. Springer.
- Stange G, Stowe S. 1999. Carbon-dioxide sensing structures in terrestrial arthropods. *Microscopy Research and Technique* 47(6): 416-427.
- Stange G, Monro J, Stowe S, Osmond CB. 1995. The CO<sub>2</sub> sense of the moth *Cactoblastis cactorum* and its probable role in the biological control of the CAM plant *Opuntia stricta*. *Oecologia* 102(3): 341-352.
- Stross RG. 1971. Photoperiod control of diapause in *Daphnia*. IV. Light and CO<sub>2</sub>-sensitive phases within the cycle of activation. *The Biological Bulletin* 140(1): 137-155.
- Suh GS, Wong AM, Hergarden AC, Wang JW, Simon AF, Benzer S, Axel R, Anderson DJ. 2004. A single population of olfactory sensory neurons mediates an innate avoidance behaviour in *Drosophila*. *Nature* 431(7010): 854-859.
- Swofford DL. 2001. *PAUP\*: Phylogenetic Analysis Using Parsimony and Other Methods, Version 4*. Sinauer Press.
- Thom C, Guerenstein PG, Mechaber WL, Hildebrand JG. 2004. Floral CO<sub>2</sub> reveals flower profitability to moths. *Journal of Chemical Ecology* 30(6): 1285-1288.
- Tribolium Genome Sequencing Consortium. 2008. The genome of the model beetle and pest *Tribolium castaneum*. *Nature*. 452(7190): 949-955.
- Vale GA, Hall DR. 1985. The role of 1-octen-3-ol, acetone and carbon dioxide in the attraction of tsetse flies, *Glossina* spp. (Diptera: Glossinidae), to ox odour. *Bulletin of Entomological Research* 75(2): 209-217.
- von Heijne G. 1989. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature* 341(6241): 456-458.
- Wanner KW, Nichols AS, Walden KKO, Brockmann A, Luetje CW, Robertson HM. 2007. A honey bee odorant receptor for the queen substance 9-oxo-2-decenoic acid. *Proceedings of the National Academy of Sciences USA* 104(36): 14383-14388.
- Wanner KW, Robertson HM. 2008. The gustatory receptor family in the silkworm moth *Bombyx mori* is characterized by a large expansion of a single lineage of putative bitter receptors. *Insect Molecular Biology* 17(6): 621-629.
- Winston ML. 1991. *The Biology of the Honey Bee*. Harvard University Press.
- Wistrand M, Kall L, Sonnhammer EL. 2006. A general model of G protein-coupled receptor sequences and its application to detect remote homologs. *Protein Science* 15(3): 509-521.

**Supplement.** Amino acid sequences of the carbon dioxide receptor proteins.

>DmelGr21a  
MSFWAVSRGLTPPSKVVPMLNPNQRQFLEDEVRYREKLKLMARGDAMEEVYVRKQETVDDPLELDKHDSFYQTTKSLVLVLFQ  
IMGVMPiHRNPPEKNLPRGTGYSWGSQVWMAIFIYSCQTTIVVLVLRERVKKFVTSPDKRFDEAIYNVIFISLLFTNFFLLPV  
ASWRHGPQVAIFKNMWTNYQYKFFKTTGSPVFPNLYPLTWSLCVFSWLLSIAINLSQYFLQPDFRLWYTFAYYPIIAMLNC  
FCSLWYINCNAFGTASRALSDALQTTIRGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMYGMYCLVIFFTTIIATYGSISE  
IIDHGATYKEVGFLFVIVFYCMGLLYIICNEAHYASRKVGDLDFQTKLLNINLTAVDAATQKEVEMLLVAINKNPPIMNLDGYA  
NINRELITTNISFMATYLVVLLQFKITEQRRIGQQQA

>DsimGr21a  
MSFWAVSRGLTPPSKVVPMLNPNQRQFLEDEVRYREKLKLMARGDAMEEVYVRKQETVDDPLELDKHDSFYQTTKSLVLVLFQ  
IMGVMPiHRNPPEKNLPRGTGYSWGSQVWMAIFIYSCQTTIVVLVLRERVKKFVTTPDKRFDEAIYNVIFISLLFTNFFLLPV  
ASWRHGPQVAIFKNMWTNYQYKFFKTTGSPVFPNLYPLTWSLCVFSWLLSIAINLSQYFLQPDFRLWYTFAYYPIIAMLNC  
FCSLWYINCNAFGTASRALSDALQTTIRGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMYGMYCLVIFFTTIIATYGSISE  
IIDHGATYKEVGFLFVIVFYCMGLLYIICNEAHYASRKVGDLDFQTKLLNINLTAVDAATQKEVEMLLVAINKNPPIMNLDGYA  
NINRELITTNISFMATYLVVLLQFKITEQRRIGQQQT

>DsecGr21a  
MSFWAVSRGLTPPSKVVPMLNPNQRQFLEDEVRYREKLKLMARGDAMEEVYVRKQETVDNPLELDKHDSFYQTTKSLVLVLFQ  
IMGVMPiHRNPPEKNLPRGTGYSWGSQVWMAIFIYSCQTTIVVLVLRERVKKFVTSPDKRFDEAIYNVIFISLLFTNFFLLPV  
ASWRHGPQVAIFKNMWTNYQYKFFKTTGSPVFPNLYPLTWSLCVFSWLLSIAINLSQYFLQPDFRLWYTFAYYPIIAMLNC  
FCSLWYINCNAFGTASRALSDALQTTIRGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMYGMYCLVIFFTTIIATYGSISE  
IIDHGATYKEVGFLFVIVFYCMGLLYIICNEAHYASRKVGDLDFQTKLLNINLTAVDAATQKEVEMLLVAINKNPPIMNLDGYA  
NINRELITTNISFMATYLVVLLQFKITEQRRIGQQQS

>DereGr21a  
MSFWAVSRGLTPPSKVVPMLNPNQRQFLEDEVRYREKLKLMARGDAMEEVYVRKPETVDDPLELDKHDSFYQTTKSLVLVLFQ  
IMGVMPiHRNPPEKNLPRGTGYSWGSQVWMAIFIYSCQTTIVVLVLRERVKKFITSPDKRFDEAIYNVIFISLLFTNFFLLPV  
ASWRHGHQVAIFKNMWTNYQYKFFKTTGSPVFPNLYPLTWSLCVFSWLLSIAINLSQYFLQPDFRLWYTFAYYPIIAMLNC  
FCSLWYINCNAFGTASRALSDALQTTIRGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMYGMYCLVIFFTTIIATYGSISE  
IIDHGATYKEVGFLFVIVFYCMGLLYIICNEAHYASRKVGDLDFQTKLLNINLTAVDSATQKEVEMLLVAINKNPPIMNLDGYA  
NINRELITTNISFMATYLVVLLQFKITEQRRVGQQQP

>DyakGr21a  
MSFWAVSRGLTPPSKVVPMLNPNQRQFLEDEVRYREKLKLMARGDAMEEVYVRKPETVDDPLELDKHDSFYQTTKSLVLVLFQ  
IMGVMPiHRNPPEKNLPRGTGYSWGSQVWMAIFIYSCQTTIVVLVLRERVKKFVTSPDKRFDEAIYNVIFISLLFTNFFLLPV  
ASWRHGPQVAIFKNMWTNYQYKFFKTTGSPVFPNLYPLTWSLCVFSWLLSIAINLSQYFLQPDFRLWYTFAYYPIIAMLNC  
FCSLWYINCNAFGTASRALSDALQTTIRGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMYGMYCLVIFFTTIIATYGSISE  
IIDHGATYKEVGFLFVIVFYCMGLLYIICNEAHYASRKVGDLDFQTKLLNINLTAVDSATQKEVEMLLVAINKNPPIMNLDGYA  
NINRELITTNISFMATYLVVLLQFKITEQRRIGQQQQ

>DanaGr21a  
MSFWAVSRGGTPPSKVVPMLNPNQRQFLEDEVRYREKLKLMARGDAMEEVYVRKQETVDDPLELDKHDSFYHTTKSLVLVLFQ  
IMGVMPiHRNPVKNLPRGTGYSWGSQVWMAIFIYSCQTTIVVLVLRERVKKFVTSPDKRFDEAIYNVIFISLLFTNFFLLPV  
ASWRHGPQVAIFKNMWTNYQYKFFKTTGSPVFPNLYPLTWSLCVFSWLLSIAINLSQYFLQPDFRLWYTFAYYPIIAMLNC  
FCSLWYINCNAFGTASRALSDALQMTIRGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMYGMYCLVIFFTTIIATYGSISE  
IIDHGATYKEVGFLFVIVFYCMGLLYIICNEAHYASRKVGDLDFQTKLLNINLTAVDAATQKEVEMLLVAINKNPPIMNLDGYA  
NINRELITTNISFMATYLVVLLQFKITEQRRITQSQA

>DpseGr21a  
MSFWAVSRGLTPPSKVAPMLNPNQRQFLEDEMRYREKLKLVARGDAMDEVYVRKQETVDDPLELDRHDSFYQTTKSLVLVLFQ  
IMGVMPiHRNPVKNLPRGTGYSWGSQVWMAIFIYSCQTTVVVLVLRERVKKFVTSPDKRFDEAIYNVIFISLLFTNFFLLPV  
ASWRHGPQVAIFKNMWTNYQYKFFKTTGSPVFPNLYPLTWSLCIFSWVLSIAINLSQYFLQPDFRLWYTFAYYPIIAMLNC  
FCSLWYINCNAFGTASRALSDALQTTIRGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMYGMYCLVIFFTTIIATYGSISE  
IMDHGATYKEVGFLFVIVFYCMGLLYIICNEAHYASRKVGDLDFQTKLLNINLTSVDAATQKEVEMLLVAINKNPPIMNLDGYA  
NINRELITTNISFMATYLVVLLQFKITEQRRITSQAA

>DperGr21a  
MSFWAVSRGLTPPSKVAPMLNPNQRQFLEDEMRYREKLKLVARGDAMDEVYVRKQETVDDPLELDRHDSFYQTTKSLVLVLFQ  
IMGVMPiHRNPVKNLPRGTGYSWGSQVWMAIFIYSCQTTVVVLVLRERVKKFVTSPDKRFDEAIYNVIFISLLFTNFFLLPV  
ASWRHGPQVAIFKNMWTNYQYKFFKTTGSPVFPNLYPLTWSLCIFSWVLSIAINLSQYFLQPDFRLWYTFAYYPIIAMLNC  
FCSLWYINCNAFGTASRALSDALQTTIRGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMYGMYCLVIFFTTIIATYGSISE  
IMDHGATYKEVGFLFVIVFYCMGLLYIICNEAHYASRKVGDLDFQTKLLNINLTSVDAATQKEVEMLLVAINKNPPIMNLDGYA  
NINRELITTNISFMATYLVVLLQFKITEQRRITSQAA

>DwilGr21a  
MSFWAVSRGLSPPGATSKVAPMLNPNQRQFLEDEMRLREKMKLMARGDTNAMDDMYLRKPETVDDPLELDKHDSFYQTTKSL  
LVLFQIMGVMPiHRNPPVKNLPRGTGYSWTSRQVMWAFIYSCQTTIVVLVLRERVKKFVTSPOKRFDEAIYNVIFISLLFTN  
FLLPVASWRHGPQVAIFKNMWTNYQYKFFKTTGSPiVFPNLYPLTWALCVFSWLLSIGINLSQYFLQPDFRLWYTFAYYPII  
AMLNCFCSLWYINCNAFGTASHALS DALQATIRGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMYGMYCLVIFFTTIIATY  
GSISEIIDHGATYKEVGLFVIVFYCMGLLYIICNEAHYASRKVGLEDFTQKLLNINLTAVDAATQKEVEMLLVAINKNPPIMN  
LDGYANINRELITTNISFMATYLVVLLQFKITEQRRSQNTSTTIT

>DmojGr21aC  
MSFWAVSRGLTPQGVAPMLNPNQRQFLEDELRYREKLKMLAGGNAIDDVYVRKPDVTDDPLELDKHDSFYKTTKSLLVLFQ  
IMGVMPiHRNPPVRNMPRTGYSWRSKQVMWAVFIYSCQTTIVVLVLRERVKKFITSPDKRFDEAIYNVIFISLLFTNLLPV  
ASWRHGPQVAIFKNMWTNYQYKFFKTTGSPiVFPNLYPLTYALCVFSWLLSIAINLSQYFLQPDFRLWYTFAYYPIIAMLNC  
FCSLWYINCNAFGTASHALS DALQATIKGEKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMYGMYCLVIFFTSIIATYGSISE  
IIDHGATYKEVGLFVIVFYCMCLLYIICNEAHYASHKVGMEFQTKLLNINLTAVDTATQKEVDMLLVAINKNPPIMNLDGYA  
NINRELITTNISFMATYLVVLLQFKITEQRR

>DvirGr21a  
MSFWAVSRGLTPPGKVAPMLNPNQRQFLEDELRYREKLKMLAGGTITIEDGYVRKPDVTDDPFELDKHDAFYRATKSLLVLFQ  
IMGVMPiLRNPPVKNMPRTGYTWTSTKQAMWAMFIYAIQTTIVVLVLRERVKKFITSPDKRFDEAIYNVIFISLLFTNLLPI  
ASWRHGPQVAIFKNMWTNYQYKFFKTTGSTIVFPNLYTVTYVLCSSWLLSIAINLSQYFLQPDFSLWYTFAYYPIIAMLNC  
FCSLWYVNCNAFGTASRALSDALQATIRGDKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMYGMYCLVIFFTSIIAAYGSISE  
IIDHGATYKEVGLFVIVFYCMCLWLFIFCNEAHFASRKVGLEDFTQKLLNINLTAVDTATQKEVDMLLVAISKNPPTMNLGDGA  
NINRELITTNISFMATYLVVLLQFKITEQRRSNIA

>DgriGr21a  
MSFWAVSRGQTTPGKVAPMLNPNQRQFLEDEIRYREKLKMQEGGMSNVYVRKLETVDDPDELDKHDSFYQTTKSLLVLFQIM  
GVMPiHRNPKDKNLPRGTGYSWTSKQVMWAFIYSCQTTIVVMVLRERVKKFITSPDKRFDEAIYNVIFISLLFTNLLPVAS  
WRHGPQVAIFKNMWTNYQYKFFKTTGSPiVFPNLLPLTYALCVFSWLLSIAINLSQYFLQPDFRLWYTFAYYPIIAMLNCF  
SLWYINCNAFGTASHALS DALQDTIKGDKPAQKLTEYRHLWVDLSHMMQQLGRAYSNMYGMYCLVIFFTSIIATYGSISEIL  
DHGATYKEVGLFVIVFYCMTLLYIICNEAHYASEKVGLEDFTQKLLNINLTAVDTATQKEVDMLLVAINKNPPIMNLDGYANI  
NRELITTNISFMATYLVVLLQFKITEQRRANLVN

>AgamGr22  
MIHTQMEDAQYEIRHQVLNPNQRQQLLEDRRRIKEQLHQLEQDNESPTHMYRRKLKIASDVNLLDQHDSFYHTTKSLLVLFQI  
MGVMPiMRSPKGVMDPRTTFTWCSKAFLWAYFIYACETIVVLVVARERINKFISTSDKRFDEVIYNIIFMSIMVPHFLLPVA  
SWRNGSEVAKFKNMWTDYQYKYLIVTGKPIVFPKLYPITWTLCTIVSWSLSLVIILSQQYLLQPDFQFCHTFAYYHIIAMLNGF  
CSLWVFNCTAFGTASKAFAKELTDVLATERPAAKLTEYRHLWVDLSHMMQQLGKAYSNMGIYCLVIFFTTIIATYGSLS  
IEHGATYKEVGLFVIVFYCMTLLFIICNEAHHASKRVGLNFQERLLNVNLTAVDKATQKEVEMFLVAIDKNPPTMNLGDYAN  
INRGLITSNISFMATYLVVLMQFKLTLLRQSAKNAFISALKANLSRIRSLDADKVNT

>AaegGr1  
MIHSQMEDSQYQIRQQILNPNQRQQLLEDNRRIKEQMQLQRDDASPSHMYIRKLEFQADVNLDDKHDSFYHTTKSLLVLFQI  
MGVMPiVRSPKGVNMPRTTFTWFSKAFIWAYFIYACETIVVLVAKERIKRFISTSDKRFDEVIYNIIFMSLLVPHFLLPVA  
SWRNGSEVAKFKNMWTDYQYKYLIVTGKPIVFPKLYPITWVLCVSVWAVSFVIMSQQYLLQPDFQLTHTFAYYHIIAMLNGF  
CSLWVFNCTAFGTASKAFQAELSNI LATEQPADKLTEYRHLWVDLSHMMQQLGKAYSNMGIYCLVIFFTTIIATYGALSEI  
IEHGATYKEVGLFVIVFYCMGLLFIICNEAHHASRRVGLNFQERLLNVNLTAVDKATQKEVEMFLVAIDKNPPTMNLGDYAN  
INRGLITSNISFMATYLVVLMQFKLTLLRQSARKALIPALRANLTKLEN

>CpipGr1  
MIHSQMEDAGYQIRQQVLNPNQRQQLLEDNRRIKEQMEQLQKENASPTRLYLRKMKVQADVNLDDHDSFYHTTKSLLVLFQI  
MGVMPiVRSPPGVNMPRTTFTNWGSRAFIWAYLIYAIETVYVVLVAKERINKFISNSDKRFDEVIYNIIFFSIMVPHFLLP  
SWRNGAEVAKFKNMWTDYQYKYLIVTGKPIVFPKLYPITWALCVSVWGVSAFVIMSQQYLLQPDFQLWHTFAYYHIIAMLNGF  
CSLWVFNCTAFGEASKAFQAELSNI FATDRPADKLTEYRHLWVDLSHMMQQLGKAYSNMGIYCLVIFFTTIIATYGALSEI  
IEHGATYKEVGLFVIVFYCMTLLFIICNEAHHASKRVGLNFQERLLNVNLTAVDKATQKEVEMFLVAIDKNPPTMNLGDYAN  
INRGLITSNVFSMATYLVVLMQFKLTLLRQSARKALIASLTNTNIAAAKTNPQ

>BmorGr1N  
DIYGPEITDKDDGALLDKHDSFYLNKTSLLVLFQIMGVMPiMRVPKSAQTTRRTTYNWISKATLWAYLVWGLECIIVVKVGQ  
ERLANFQIGSNKRFDEVIYNIIFLSILIPHFLPIASWRHGPQVAIFKNMWTNYQLKYLKITGKPIVFPNLYILTWGLCIFS  
WVLSFAVVLSSQHYLQDDFELWHSFAYYHIIAMLDGFCSLWYINCNAFGTASRLA INLHKALEAEHPALKLAQYRHLWVDLS  
HMMQQLGRAYSNMYGIYCMVIFFTTTISLYGALSEILEHGLSYKEMGLFVIVAYCMTLLFIICNEAYHASRKVGHEFDRL  
NVNLGAIDRSTQREVEMFLVAIAKNPPIMNLDGFTNINRELFTANISFMSTYLIVLMQFKLTLLRQGARKVTVAIVRAIFNT  
TITDNGAGGSDEDEQE

>TcasGr1  
MRNDHGSNTHLHPDDAIRRAKIVKVAASPTSANPDEEDPELLDRYDNFYQTTKSLLVLFQIMGVMPIERSGKGRTRTFRWLS  
STSIYAYFIFGAETIFVTMVFKERLYLILRPGKRFDEYIYGIIFLSILIPHFLLPVAAWTNGTEVAKFKNMWTRFQLKYYQV  
TGTPIIFHNLTLLITYSLCVISWAVGIGIMLAQYYLQADMLLWHTFGYYHILAMLNCLCSLWFINCTAKGRVAVWMCNNLHKA  
LESRNPAKILGAYRDLWVDLSHMMQQLGKAYSGMYSMYCLLILLTTIVASYGSVTEIMDQGISFKEAGLFMIAFYCMTLLYI  
ICNEGHATRKMGPEFRERLLNVNLSAVDQKTRQEVHMFMAIEKNPPIMNLNGYANVRKLISSTVTSTIATYLVMLMQFRL  
TLMRNAQLAARRAIANVSVSSGNTTMS

>AgamGr23  
MVIKESFEFDDSLGYALLRRDMGTVWDTAKDERMVNGTMDPELIQRAKERAVRAQLNSADGDTCEHDQFYRDHKLLLVLFRG  
LAVMPITRSVPGRITFSWRSAAISIYAFCFYLVSTVIVLVVGYERIKVFQTTTKFDEYIYGILFVIFLVPHFWIPFVGWGVAK  
QVAIYKTMWGAQVRYRVTGTSLQFPHLKLLIVFLSIGCLVCAIVFLLSLSFLLEGFALWHTSAYYHIIITMLNMNSALWYI  
NSRGIRVASSSLSRCFRQDVAIECTAAMISRYRFLWNLSELLQALGNAYARTYSTYCLFMFVNITVAIYGALSEIIDHGFG  
FSFKEIGLIVDVTYVCSTLLFIFCDCSHNATLQVAQGVQDTLLSINLLKVDQPTQKEIDLFIQAIEMNPAIVSLKGYAEVNRE  
LLTSSIIATIAIYLVLLQFKLSLISQQIPVEIIENVKLLQKQ

>AaegGr2  
MVIKSEFEDSLNYALLRGDMGTTWDINKDERMMNGTLDPELIQRAKERAIRAQLNSADGDTCELHDQFYRDHKLLLVLFRA  
LAVMPILRSSPGRITFDWRSWASIYAYCFYVSTVIVLVVGYERIKILQDTKKFDEYIYGVLFIIFLVPHFWIPFVGWGVAK  
HVAVYKTMWGAQVRYRVTGTNLQFPHLKILIVMFSIGCLVCAIVFLLSLSFLLEGFALWHTSAYYHIIITMLNMNSALWYI  
NCRGIRVASSSLSDRFRKDVAIECTAAMISQYRFLWNLSELLQALGNAYARTYSTYCLFMFANITIAIYGALSEVIDHGFG  
FSFKEIGLIVDVTYVCSTLLFIFCDCSHNATLQVAQGVQDTLLGINLLKVDHPTQKEIDLFIQAIEMNPAIVSLKGYAEVNRE  
LLTASIIATIAIYLVLLQFKLSLISQQMPIELMEIKHSHKG

>CpipGr2  
MVIKDSDFDESLNYALLRGDMGAIWDTTKDQRLMNGTMNPPELIQSKERAIRAQLNSADGDTAETHDQFYRDHKLLLVLFRA  
LAVMPILRSSPGRITFDWRSWASIYAYCFYFLSTIVVLVVGYERIKVLQETKKFDEYIYGVLFIIFLVPHFWIPFVGWGVAK  
HVAVYKTMWGAQVRYRVTGTNLQFPHLKVLIVIFSIGCLICAIVFLLSLSFLLEGFLLWHTTAYYHIIITMLNMSSALWYI  
NCRGIRVASSSLSDRFRKDVAIECTAAIISQYRFLWNLSEMLQALGNAYARTYSTYCLFMFVNITIAIYGALSEVIDHGFG  
FSFKEIGLIVDVTYVCSTLLFIFCDCSHNATLQVAQGVQDTLLSINLLKVDLPTQKEIDLFIQAIEMNPAIVSLKGYAEVNRE  
LLTSSIIATIAIYLVLLQFKLSLISQQMPVDLLENLQKAHAH

>BmorGr2N  
KEQEQRDLSSQDGTCEIHDQFYRDHKLLLVLFRALAVMPITRSRPGTITFSWKSTATIYAVCFYIAATAVVLIVGYERIQ  
ILQSIKRFDDYIYAILFIVFLVPHFWIPFVGWGVAVQVAIYKTNWGKFQVRYRVTGENLKFNPCLKTLIVIIISVGCLLAVC  
FLLSLCALLDGFLKHTSAYYHIIITMINMNCALWYINCKAIKIASQSLSECFQRDVEDIECSAQLIARYRYLWNLSELLQSL  
GNAYARTYSTYCLFMFANITIAVYGALSEIVDHGFGFTFKEVGLFVDAAYCSTLLFVFADCSHKSTLKVAAGVQDTLLSIDV  
LAVDRPTQKEIDHFIQAIEMNPAFVSLKGYAHVNRELLTSAISMITIYLVLLQFKISLPKEPHGTGQ

>TcasGr2  
MEISDLAQLYGNELHIKQISKWLGRSARAQEIQRSELDKDGHVIDEHDQFFRDHKLLLVLFRLVGMPIQRGEIGRITFG  
WTSIPMLYAYVYVVTTLVVLVGYERFDILLNKSKKFDEYIYSIIFIYIYLIPHFFIPFVGWGVAYEVCYKNSWGGFQLHY  
YKITGNLQFPLSTLIIISLGLILAVVFLTLTSLALLEGFTLYHTTAYLHIIITMINMNCALWYINCRVGNASTALAESF  
QNDVDRNCSAYIIAHYRVLWLSLSDLLQKMGNAYARTYSTYSLFMMANITVAVYGFTSEIVDHGIRFSFKEIGLLVDSTYCL  
FLLFVFCDCSHQASLNIARRVQVTLQVNLQVDPATRKEDIFLVAIQMNPVKVSLKGYTVVNRELVTASVATIAIYLVIL  
LQFKISLLNMRG

>DmelGr63a  
MANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSIGLAERLDADYEAPPLDRKKSSDSTASNNPEFKPSVFYRNIDPINW  
FLRIIGVLPPIVRHGPAPAKFEMNSASFYISVVFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILW  
YEARKIAKLFNDWDDFEVLYYQISGHSPLPLKLRQKAVYIAIVLPILSVLSVVITHVTMSDLNINQVVPYCILDNLTLMLGAW  
WFLICEAMSIHAHLAERFQKALKHIGPAAMVADYRVLWRLSKLTRDTGNALCYTFVFMSSLYLFFIITLSIYGLMSQLSEG  
FGIKDIGLTITALWNIGLLFYICDEAHYASVNVRTNFQKLLMVELNWMNSDAQTEINMFLRATEMNPNSTINCGGFFDVNRT  
LFGKLLTTMVITYLVLLQFQISIPTDKGDSEGANNTIVVDFVMSLDNDMSLMGASTLSTTTVGTTLPPPIMKLGKRG

>DsimGr63a  
MANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSIGLAERLDADYEAPPLDRKKSSDSTASNNPEFTSPVFYRNIAPVNW  
FLRIIGVLPPIVRHGPAPAKFEMNSASFYISVVFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILW  
YEARKIAKLFNDWDDFEVLYYQISGHSPLPLKLRQKAVYIATVLPILSVLSVVITHITMSDLNINQVVPYCILDNLTLMLGAW  
WFLICEAMSIHAHLAERFQKALKHIGPAAMVADYRVLWRLSKLTRDTGNALCYTFVFMSSLYLFFIITLSIYGLMSQLSEG  
FGIKDIGLTITALWNIGLLFYICDEAHYASVNVRTNFQKLLMVELNWMNSDAQTEINMFLRATEMNPNSTINCGGFFDVNRT  
LFGKLLTTMVITYLVLLQFQISIPTDKGDSEGANNTIVVDFVMSLDNDMSLMGASTLSTTTVGTTLPPAPIMKLGKRG

>DsecGr63a  
MANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSIGLAERLDADYEAPPLDRKKSSDSTASNNPEFTPSVFYRNIAPVNW  
FLRIIGVLPPIVRRGPARAKFEMNSASFIYSVVFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILW  
YEARKIAKLFNDWDDFEVLYYQISGHSPLPLKLRQKAVYIATVLPILSVLSVVIHITMSDLNINQVVPYCILDNLTLAMLGAW  
WFLICEAMSIHAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSSLYLFFIITLSIYGLMSQLSEG  
FGIKDIGLTITLWNIIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATERNPSTINCGGFFDVNRT  
LFGKLLTMTVLYLVLLQFQISIPDKGDEGANNITVDFVMDSLDNDMSLMGASTPSTTTVGTTLPPAPIMKQKGRRG

>DyakGr63a  
MANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSIGLAERLDADYEAPPLERKKSSESTASNNPEFTPSVFYRNIAPVNW  
FLRIIGVLPPIVRNGPARARFEMNSASFIYSVVFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILW  
YEARKIAKLFNDWDDFEVLYYQISGHSPLPLKLRQKAVYIATVLPILSVLSVVIHITMSDLNINQVVPYCILDNLTLAMLGAW  
WFLICEAMSIHAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSSLYLFFIITLSIYGLMSQLSEG  
FGIKDIGLTITLWNIIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATERNPSTINCGGFFDVNRS  
LFGKLLTMTVLYLVLLQFQISIPDKGDEGATNITVDFVMDSLDNDMSLMGVSTPSTTTAGTTLPPIPMKQKGRKG

>DereGr63a  
MANYYRRKKGDAVFLNAKPLNSANAQAYLYGVRKYSIGLAERLDADYEAPPLDRKKSSSESTASNNPEFTPSVFYRNIAPVNW  
FLRIIGVLPPIVRHGPARAKFEMNSASFIYSVVFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMIIPILW  
YEARKIAKLFNDWDDFEVLYYQISGHSPLPLKLRQKAVYIATVLPILSVLSVVIHITMSDLNINQVVPYCILDNLTLAMLGAW  
WFLICEAMSIHAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSSLYLFFIITLSIYGLMSQLSEG  
FGIKDIGLTITLWNIIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATERNPSTINCGGFFDVNRT  
LFGKLLTMTVLYLVLLQFQISIPDKGDEGATNITVDFVMDSLDNDMSLMGASTPSTTTAGTSPPPIMKQKGRKG

>DpseGr63a  
MANYYRRKKDAVFHNAKPLNSGNAQAYLYGVRKYSIGLAERLDADYQPPPSDRKKSSDSTGSNNPEFTPSVFYRNIAPVNW  
LRIIGVLPPIVRRGPARAKFEMSSASFIYSVVFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMVPIPLWW  
EARKIAKLFNDWDDFEVLYYQISGHSPLPLRLRQKALYIAIVLPILSVLSVVIHITMSDLNINQVVPYCILDNLTLAMLGAW  
FLICEAMSTTAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNAMCYTFVFMSSLYLFFIITLSIYGLMSQLSEGF  
GIKDIGLTITLWNIIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATERNPSTINCGGFFDVNRSL  
FKGLLTMTVLYLVLLQFQISIPDKGSDGGTNTITVDMMLDMSLGNDMTILSASSSTTHSTATSSTTPPPTSAKHGRGHR  
G

>DperGr63a  
MANYYRRKKDAVFHNAKPLNSGNAQAYLYGVRKYSIGLAERLDADYQPPPSDRKKSSDSTGSNNPEFTPSVFYRNIAPVNW  
LRIIGVLPPIVRRGPARAKFEMSSASFIYSVVFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIMVPIPLWW  
EARKIAKLFNDWDDFEVLYYQISGHSPLPLRLRQKALYIAIVLPILSVLSVVIHITMSDLNINQVVPYCILDNLTLAMLGAW  
FLICEAMSTTAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNAMCYTFVFMSSLYLFFIITLSIYGLMSQLSEGF  
GIKDIGLTITLWNIIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATERNPSTINCGGFFDVNRSL  
FKGLLTMTVLYLVLLQFQISIPDKGSDGGTNTITVDMMLDMSLGNDMTILSASSSTTHSTATSSTTPPPASAKHGRGHR  
G

>DanaGr63a  
MASYYRRKKPDAVFLNAKPLNSANAQAYLYGVRKYSIGLAERLDADYEAPPVDRKKSSDSTASNNPEFTPSVFYRNIAPVNW  
FLRIIGVLPPIVRRGPARAKFEMNSASFIYSVVFYILLSCYVSIVANNRIHVVRSLSGPFEEAVIAYLFLVNILPIMIIPILW  
SEARKIARLFNDWDDFEVLYYQISGHSPLPLKLRQKAVYIAIVLPILSVLSVVIHITMSDLNINQVVPYCILDNLTLAMLGAW  
WFLICEAMSTTAHLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSSLYLFFIITLSIYGLMSQLSEG  
FGIKDIGLTITLWNIIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATERNPSTINCGGFFDVNR  
LFGKLLTMTVLYLVLLQFQISIPDKGDEGNSNMTVVDLLMDSLDNDMTLMGSTATTASTTRVSTSLAPPTTKSTRGRKG

>DmojGr63a  
MASYYRRKKPDMVFLNAKPLNSGNAQAYLHGVRKYSIGLAERLDSYIIPPNDRKSSSVSTIASNNPDFTPSVFYRNIAPVN  
WFLRIIGVLPPIVRRGPRAKFALNSAFAFIYSVVFVLLAFYVGYVANKRIHAVRSLSGPFEEAVIAYLFLVNILPIIPIIL  
WWEARKIARLFNDWDDFEVLYYQISGHSPLPLNRQKAVYIAIVLPILSILSVVIHITMSDLNINQVVPYCILDNLTLAMLGA  
WWFLICEAISTTAYLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSSLYLFFIITLSIYGLMSQLSE  
GFGIKDIGLTITLWNIIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATERNPSTINCGGFFDVNR  
TLFGKLLTMTVLYLVLLQFQISIPDKGDEGNSNMTVVDLLMDSLDNDMTLLGAPSTVATPSTTTTAVPPTVNRSGRGRK  
G

>DvirGr63a  
MASYYRRKKADTVFLNAKPLNSGNAQAYLQGVKYSIGLAERLDNDYIIPANDKKRGSISTVGSNNPDFTPSVFYRSIAPVN  
WFLRIIGVLPPIVRRGPRAKFALNSAFPVYSVVFVLLACYVGYVANNRIHIVRSLSGPFEEAVIAYLFLVNILPIIPIIL  
WLEAKKIALLFNDWDDFEVLYYQISGHSPLPLNRQKAIYIAILLPILSVLSVVIHITMSDFNLNQVVPYCILDNLTLAMLGG  
WWFLICEAISTTAYLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDMGNALCYTFVFMSSLYLFFIITLSIYGLMSQLSE  
GFGIKDIGLTITLWNIIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATERNPSTINCGGFFDVNR  
TLFGKLLTMTVLYLVLLQFQISIPDKGDEGAGNSNVTVADMLDMSLGNDMTLLGTPSSTLAPTPTTTPVGRSGRGRKG

>DwilGr63a  
MANYRRKKPDVFLNAKPINSANAQAYLYGVRKYSIGLAERLDSYQPPPIERKKSTASTGSNNPEFTPSVFYRNIAPVNW  
FLRIIGVLPPIVRRGPARAKFEMNSAAFFYSVVFMLLACYVGYVANNRIHVVRSLSGPFEEAVIAYLFLVNILPIMIIPILW  
WEAKKIARLFNDWDDFEVLYYQISGHSPLHLRQKALYIAIILPILSVLSVVITHITMSDLNINQVVPYCILDNLTLAMLGAW  
WFLICEAMSNTAHLAERFQKALKHVGPAAVADYRVLWLRLSKLTRDTGNAMCYTFVFMSSLYLFFIITLSIYGLMSQLSEG  
FGIKDIGLTITLWNIGLLFYICDEAHYASVNVRTNFQKKLLMVELNWMNSDAQTEINMFLRATEMNPNSINCGGFFDVNRS  
LFKGLLTMTVTYLVVLLQFQISIPDTKDGDEASTNVTVDMLMDSLNDMTLLGPTSTAGTTATMRAAATTTTMTPTVKQG  
RAGRRG

>DgriGr63a  
MASYYRRKKPDTVFLNANPINSSNAQAYLQGVKYSIGLAERLDSGYQKPSNDRKRSSVTTVDSQSLGTTFTPSVFYRNIAPV  
NWFLRIIGVLPPIVRSGPSRAKFALNTAPFLYSVIFFTLLACYVGYVAKNRIHVVRSLSGPFEEAVIAYLFLVNILPIMIIPILW  
LWWEARKIARLFNDWDDFEVLYYQISGHSPLPLNRQKATYIAIGLPIISVLSVVIHMTMSDLNINQVVPYCILDNLTLAMLG  
AWWFIICEAISTTAYLLAERFQKALKHIGPAAMVADYRVLWLRLSKLTRDTGNALCYTFVFMSSLYLFFIITLSIYGLMSQLS  
EGFGIKDIGLTITLWNIGLLFYICDQAHYASVNVRTNFQKKLLMVELNWMNMMDAQTEINMFLRATEMNPNSINCGGFFDND  
RSLFKGLLTMTVTYLVVLLQFQISIPDTKDGDDGDNANMTVIDLLMDSMNNDMTVVGHSSSTTGTTAAPTNTTTTTPVN  
RSGGRGRKG

>AgamGr24  
MRIERSVHEPKRNRNVFLDVKPIADDANVNVPPRQAARRNATVFNNRVGFPLTPKEAFVDAVPADQTCMVFESSKPIYLV  
LRAIGVLPYTRLPSSGGTAFVLASPSMTYCVLFFLLLTVYIAFILLNRIEIVRTLEGRFEESVIAYLFIIVNLPILIIPLMWY  
ESRKVSVVNGWDFETVYRETTGRALELRRTKAQVIAILLPILCSLSVAITHVTMVDFKLLQVIPPYCVLDTITYMMGGYW  
YMACETLSITAKILAEDFQRALRHVGPAAKVSEYRSLWLRLSKLARDTGFTCYTFTFICLYLFFIITLSIYGLMSQISDGF  
GVKDIGLAVTAFCSVGLLFYICDEAHYASFNVRTNFQKKLLMVELSWMNTDAQTEINMFLRATEMNPSSINLGGFFDVNRTL  
FKSLLATMTVTYLVVLLQFQISIPDEPSAMLMHSNSSHS

>AaegGr3  
MNLNQDPIQYINLNNNARTVFLDVKPIYNEEKRKVSNGFNRRIGFPPISSRRVFGLESEFNTRSDIVYGTTPKIYNVLRMLG  
VFPFSRPSPGVTLFACASPAMAYCGVLFVTLMAVVIYITILRVHIVRTLEGRFEEAVIAYLFIIVNLPVLIIPLMWYETRKY  
SSLLNQWVDFEATYRKTAGRELELSFRTKALLIAILLPVLSCLAVIITHVTMVEFQLVQVIPPYCILDTLTYMMGGYWMTCE  
TLSITANILAEDFQRALRHVGPAAMVSEYRSLWLRLSKLARETGSSTCYTFTFICLYLFFIITLSIYGLMSQISEGFGIKDI  
GLAVTAFCSVGLLFYICDEAHYASFNVRTNFQKKLLMAELSWMNSDAQTEINMFLRATEMNPSSINLGGFFDVNRTLFSKLL  
ATMTVTYLVVLLQFQISIPDDSSMLVMHNMGTGSYRE

>CpipGr3  
MSIFPDTLRYIEVEPDPKTRAVFLDAKPAYLDHQHQRQTNGFGNRVGFAGPPREAFDGGGVVKSIIYDSSKPIYNV  
LRLLGVPFFMRPTAGMTMFACASPAMAYSVVFLVVLTIYVVFIMISRIDIIVRTLEGRFEEAVIAYLFIIVNLPVLIIPLMWY  
ETRKVCNLLNNWVDFEVLYQKTAGRELALNLKNKSLIAVLLPVLSCLASVITHVTMVEFQLVQVIPPYCILDTLTYMMGGYW  
YLACETLSTTANVLAEDFQMALRHVGPAAMVSEYRSLWLRLSKLARDTGFTCYTFTFICLYLFFIITLSIYGLMSQISEGF  
GIKDIGLAVTAFCSIGLLFFYICDEAHYASFNVRTNFQKKLLMVELTWMNSDAQTEINMFLRATEMNPSSINLGGFFDVNRTL  
FKSLLATMTVTYLVVLLQFQISIPDDPTAMLKQNSTAAH

>BmorGr3  
MSFEIKNNFFRTSVPIPNFGFPVQTEAKSKNKPIFLDVSPAPTPKVNSPNAIIPMKNLIDPFINKDIIYENIKPVFMVLRIM  
GVLPLTRTTSGVNEFHFIISAMVYSLTVFIILVSYISYLSLHKVQIVRNSEKGFEEAVIEYLFVYLFPLTVVPILWYETRK  
IANVLNGWVQFEVYTKQLSNRILPVKLYKKSLLIAIIIPILSTTSVIVTHVTMVHFKTSQIIPYVFLIILTYMLGGYWYLLC  
EILSLCANVLAEDFQALRHVGPAAGVAKYRALWLRLSKLARNTGVANCYTFTFVNLYLFLIITLSIYGLLSKISEGFGTKD  
IGLALTALCSVFLFFYICDEAHYASHNVRTNFQKKLLMVELSWMNTDAQTEVNMFRLATEMNPSSQISLGGFFDVNRTLFSKLL  
LATMTVTYLVVLLQFQISIPDATQPEIPTNIDHVNITDTTTEASSPISTLMSAFAKRKND

>TcasGr3  
MYHQDQAVSILGEAIPKRRSVFLESVNSADSFKASKVGPAPPIKFINKSSTDKFGNGAIYEVLPKIYALMRIVGIFPIKNT  
EPGMFRVAPELLGYSVVVFVVMGYIGFIEWDKVEIVRSQEGRFEEAVIDYLFVYLLPIIINPLVLYEARLANVVDWVN  
FERIYYKLTKKKLSVFFGNKPVILTVPVLLACGVMVTHITMAHFKIIQVVPYCYINCLIIYLIGGFWMQCDVVGKVASQL  
AEDFQMALKHVGPSSQVADYRSLWMLLSKLIRDVGNASGYTVTFLCLYLFLIITLTIIYGLLSQLQAGFSTKDIGLTINAGLA  
IFILYFICDEAHYASNCLRVQFQKKLLLVLSWMNDEAQEQEINMFLKATEMSPTDISLVGGFFDVNRNLFKSLLATMTVTYLVV  
LLQFQISIPPEEASPTNSTTTTTPN