# G Protein–Coupled Receptors in *Anopheles gambiae*

**Catherine A. Hill,[1]* A. Nicole Fox,[2]* R. Jason Pitts,[2]* Lauren B. Kent,[3] Perciliz L. Tan,[1] Mathew A. Chrystal,[1] Anibal Cravchik,[4] Frank H. Collins,[1] Hugh M. Robertson,[3] Laurence J. Zwiebel[2]†**

**We used bioinformatic approaches to identify a total of 276 G protein–coupled receptors (GPCRs) from the *Anopheles gambiae* genome. These include GPCRs that are likely to play roles in pathways affecting almost every aspect of the mosquito's life cycle. Seventy-nine candidate odorant receptors were characterized for tissue expression and, along with 76 putative gustatory receptors, for their molecular evolution relative to *Drosophila melanogaster*. Examples of lineage-specific gene expansions were observed as well as a single instance of unusually high sequence conservation.**

Several classes of proteins with deduced structural characteristics consistent with GPCRs (AgamGPCRs) mediate sensory pathways that directly impact the ability of the malaria vector mosquito *A. gambiae* to transmit disease, a quantifiable term known as vectorial capacity (*1*). For example, olfaction plays a major role in the strong preference for human hosts (anthropophily) that helps to make *A. gambiae* and other vector mosquitoes so adept in the transmission of diseases with catastrophic effects on global public health such as malaria, dengue, West Nile encephalitis, and yellow fever (*2*). GPCRs comprise the largest category of proteins in animal genomes and are involved in chemo- and photoreception, hormonal physiology, synaptic function, and a variety of other processes (*3*). In addition to other considerations, GPCRs are characterized by the presence of a central domain encoding seven-transmembrane (TM) regions (*4*). These receptors link ligands and downstream effectors as well as amplify and integrate other cellular signals (*5*). Here we describe the GPCR repertoire in the genome of *A. gambiae* along with detailed analyses of the opsin, methuselah, odorant, and gustatory receptors.

Table 1 summarizes the 276 AgamGPCRs we have identified (*6*), 271 of which have not been reported previously. In addition, a Quasi Periodic Feature Classifier (QFC) algorithm (*7*) was used to confirm homology-based bioinformatic searches and to identify, on the basis of amino acid physico-chemical information, additional proteins containing multiple TM domains that might be additional GPCRs. The QFC identified 2207 proteins of which 22 displayed features consistent with seven-TM spanning proteins and are potentially novel AgamGPCRs. The complete list of AgamGPCRs, including their predicted amino acid sequences, is available as supporting online material (table S1). AgamGPCRs represent ~1.6% of the proteome predicted from the 278 megabase *A. gambiae* genome sequence (*8*), which is comparable to the percentage of GPCRs predicted in *Drosophila melanogaster* and mammalian proteomes but less than *Caenorhabditis elegans*, where GPCRs constitute ~6% of the proteome (*9*).

In addition to the four generally accepted classes of GPCRs (*3*), we propose a new class consisting of the highly divergent insect chemosensory receptors that constitute the majority (155) of AgamGPCRs. In many cases, the names given to the *A. gambiae* genes and proteins refer to sequence similarities to functionally characterized GPCRs from other organisms. Though coupling to G proteins and the ligands for AgamGPCRs remain experimentally unverified and could differ from these counterparts, the availability of the complete repertoire of these sequences should facilitate further studies.

In many cases, a putative AgamGPCR ortholog has been identified for *D. melanogaster* GPCRs; exceptions include 10 *D. melanogaster* class A orphan receptors, 1 glycoprotein hormone receptor, 5 peptide receptors, 2 HE6-like receptors, and 1 γ-aminobutyric acid–B ($GABA_B$) receptor (their orthologs might have diverged sufficiently that they cannot be identified based on amino acid similarity; see table S1). In contrast to the generally orthologous relationships in most families, the methuselah and opsin families show considerable species-specific gene expansion. The methuselah family has 15 members in *D. melanogaster*, but only 7 have been identified in *A. gambiae*. Whereas there are single *A. gambiae* orthologs for the Dmelmth-like genes 1, 5, and 14, the rest of the family in *D. melanogaster*, including the canonical methuselah receptor (*10*), corresponds to a cluster of just three genes in *A. gambiae* (fig. S1). In contrast, *A. gambiae* displays an expansion of one subfamily of opsins (fig. S2). There are orthologs of DmelRh5 and -7, as well as the DmRh3/4 ultraviolet (UV) receptor pair, but no apparent ortholog of the intermediate wavelength DmRh1/2 receptors. In addition, *A. gambiae* has seven genes (four of which are almost identical) that are orthologous to the single green receptor DmelRh6. Similar expansions of long-wavelength photoreceptors have been reported in lepidopterans (*11*). Remarkably, *A. gambiae* has two genes encoding highly divergent opsins resembling the nonvisual pineal, vertebrate ancient (VA), and teleost multiple tissue (TMT) opsins of vertebrates as discussed in fig. S2 (*12*).

In the insect chemosensory receptor class, we identified 79 candidate odorant receptors (GPRors), 5 of which were previously characterized (*13*, *14*), and 76 candidate gustatory receptors (GPRgrs). In *D. melanogaster*, some of the gustatory receptors (Grs) are encoded by alternatively spliced genes (*15*) and there are five such putatively alternatively spliced AgGPRgr genes. Of these, one (GPRgr9) has 14 $NH_2$-terminal exons each hypothesized to be alternatively spliced to two short COOH-terminal exons encoding TM7 regions (fig. S3). We examined the molecular evolution of these two families in *A. gambiae* and *D. melanogaster* (Fig. 1 and fig. S4). Here, the only unequivocal orthologous pair of odorant receptors (Ors) is AgamGPRor7 and DmelOr83b at the base of the tree. DmelOr83b is unusual because it is expressed in most olfactory receptor neurons in the antennae and palps, in addition to the single specific Or expressed in each neuron (*16*, *17*). Therefore, it has been considered to be unlikely to function as a typical Or recognizing particular ligands, but rather to have another function common to all olfactory chemosensory neurons, perhaps as a dimerization partner with all other Ors (*17*). The high amino acid sequence conservation of DmelOr83b and AgamGPRor7 (78% identity) supports a distinct and important role for this protein in insect olfaction.

Seven other groupings of AgamGPRors and DmelOrs have bootstrap support, but most involve complicated relationships of multiple genes in one or both species, suggesting that at a minimum there has been extensive gene duplication in each lineage.

[1]Department of Biological Sciences, University of Notre Dame, Notre Dame, IN 46556, USA. [2]Department of Biological Sciences, Program in Developmental Biology and Center for Molecular Neuroscience, Vanderbilt University, Nashville, TN 37235, USA. [3]Department of Entomology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [4]Celera Diagnostics, 45 West Gude Drive, Rockville, MD 20850–1232, USA.

*These authors contributed equally to this work.
†To whom all correspondence should be addressed. E-mail: l.zwiebel@vanderbilt.edu

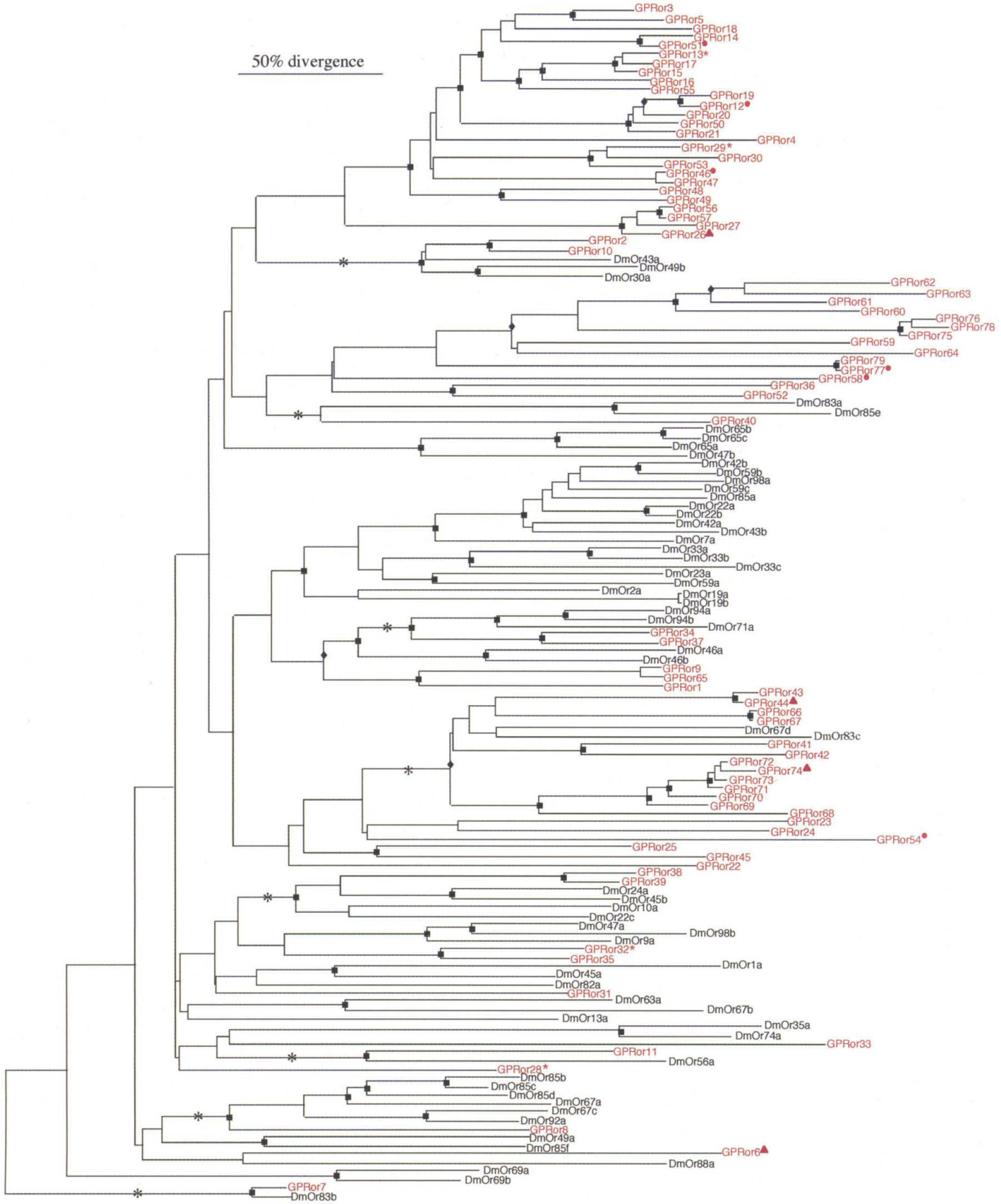


**Fig. 1.** Phylogeny of the GPRors and DmelOrs. AgamGPRor names are in red type; DmelOrs (*16*, *17*), in black type. The tree is rooted with the conserved DmelOr83b and AgamGPRor7 orthologous pair declared as the outgroup based on the location of DmelOr83b in analyses of the entire chemoreceptor superfamily. Bootstrap support is indicated with a diamond on the relevant branch point for 60 to 80% and a square for 80 to 100%. Branches of possibly orthologous groupings of GPRors and DmelOrs are highlighted with an asterisk. Tissue expression of GPRors, as determined by RT-PCR, is as follows: no symbol indicates olfactory tissue-specific expression; red asterisks denote expression in both olfactory tissues and legs; red-filled triangles indicate undetectable by RT-PCR; red-filled circles were not tested. The scale bar is 50% change in corrected distances.

Given the paucity of available data, it is difficult to make strong arguments about the ligand specificities of any of these AgamGPRors on the basis of their similarity to DmelOrs, although ligands have been identified for DmelOr43a (*18*, *19*), a member of a potentially orthologous group of *A. gambiae* and *D. melanogaster* Ors.

The dominant feature of the molecular evolution of AgamGPRors and DmelOrs is the expansion of subfamilies unique to each dipteran lineage. For example, there is a large subfamily of 27 AgamGPRors with no close *D. melanogaster* relatives and a large subfamily of 18 DmelOrs with no close *A. gambiae* relatives. AgamGPRor family members are dispersed throughout the three chromosomes of *A. gambiae* (see fold-out chart associated with this issue). Most GPRors are tightly linked as pairs, triplets, or larger clusters of up to nine genes, whereas 17 GPRors exist as single genes. In contrast, DmelOrs are never found in linkages of more than three (triplets) (*16*). Moreover, these large clusters of GPRor genes are exclusively made up of sets of recently duplicated GPRor genes as evidenced by their high degree of sequence identity to each other, a lack of putative DmelOr orthologs, and conserved transcription orientation. This pattern of lineage-specific gene subfamily expansion may reflect the ecological and physiological relevance of these receptors where they may be responsible for detection of chemicals uniquely important to each species, for example, rotting fruit odors for *D. melanogaster* and human host odors for *A. gambiae* (*2*).

Phylogenetic analysis of the Gr families in these two species (fig. S4) reveals similar patterns of largely lineage-specific gene subfamily amplification. There are just seven possibly orthologous pairs or clusters of Grs shared by the two species, but several of these are remarkably conserved (for example, AgamGPRgr22/DmelGr21a and AgamGPRgr24/DmelGr63a display 68 and 65% identity, respectively), suggesting they might be sufficiently conserved to recognize similar ligands. DmelGr21a and -63a are two of four DmelGrs expressed in antennal olfactory sensory neurons instead of gustatory neurons and may be a second lineage of functionally olfactory receptors in the superfamily (*20*, *21*). The only DmelGr for which the ligand is known is DmelGr5a, which recognizes trehalose (*22*, *23*) and is part of an orthologous cluster of eight Grs in each species that might be involved in perception of diverse sugars.

We used reverse transcriptase–polymerase chain reaction (RT-PCR) to investigate whether the putative GPRors display olfactory-specific expression (Fig. 1 and fig. S5). The majority of GPRors (64 of 79) show expression only in olfactory tissues, whereas four GPRors show additional expression in legs (five were undetectable by RT-PCR, and six were not tested). The observation that a small number of GPRors are expressed in legs in addition to olfactory tissues is reminiscent of *D. melanogaster* where a subset of DmelOrs is expressed in legs (*20*). No candidate GPRor-specific products were observed with head/proboscis or body cDNA templates.

Elucidation of signaling pathways downstream of AgamGPCRs is an ongoing undertaking that, along with the work in *D. melanogaster*, will improve our understanding of the functions of these receptors. Experimental verification of predicted GPCRs is an essential step in pursuing these studies that may facilitate the discovery of pharmacological targets ultimately leading to the development of innovative applications for invertebrate pest control. Moreover, the important role that these genes are likely to play in mediating host preference and other processes central to vectorial capacity suggests these studies will likely contribute to the development of a new generation of mosquito attractants and repellents, as well as other types of antimalarial programs.

**Table 1.** *A. gambiae* (*Ag*) GPCRs. The number of receptors predicted in each class and family is shown in comparison to *D. melanogaster* (*Dm*). LGR, leucine-rich repeat GPCR.

| Receptor | *Ag* | *Dm* |
|---|---|---|
| *Class A: Rhodopsin-like* | | |
| Biogenic amine | 18 | 17 |
| Glycoprotein hormone/LGRs | 3 | 4 |
| Peptide | 25 | 30 |
| Purine | 1 | 1 |
| (Rhod)opsin | 12 | 8 |
| Orphan | 22 | 32 |
| *Class B: Secretin-like* | | |
| Calcitonin | 3 | 3 |
| Diuretic insect hormone | 2 | 3 |
| Growth hormone releasing hormone | 4 | 3 |
| HE6 like | – | 2 |
| Latrophilin | 1 | 1 |
| Methuselah-like | 7 | 15 |
| Orphan | 4 | 2 |
| *Class C: Metabotropic glutamate-like* | | |
| Metabotropic glutamate | 5 | 5 |
| $GABA_B$ | 3 | 4 |
| *Class D: Atypical GPCRs* | | |
| Frizzled/Smoothened | 8 | 6 |
| Boss | 1 | 1 |
| Orphan | 2 | 2 |
| *Class E: Chemosensory GPCRS* | | |
| Odorant | 79 | 62 |
| Gustatory | 76 | 69 |

## References and Notes


1. G. Macdonald, *The Epidemiology and Control of Malaria* (Oxford Univ. Press, London, 1957).
2. W. Takken, B. G. Knols, *Annu. Rev. Entomol.* **44**, 131 (1999).
3. T. Brody, A. Cravchik, *J. Cell. Biol.* **150**, F83 (2000).
4. J. Bockaert, J. Pin, *EMBO J.* **18**, 1723 (1999).
5. H. G. Dohlman, J. Thorner, M. G. Caron, R. J. Lefkowitz, *Annu. Rev. Biochem.* **60**, 653 (1991).
6. Materials and methods are available as supporting online material on *Science* Online.
7. J. Kim, E. N. Moriyama, C. G. Warr, P. J. Clyne, J. R. Carlson, *Bioinformatics* **16**, 767 (2000).
8. R. A. Holt *et al.*, *Science* **298**, 129 (2002).
9. H. M. Robertson, *Chem. Senses* **26**, 151 (2001).
10. A. P. West *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 3744 (2001).
11. A. D. Briscoe, *Mol. Biol. Evol.* **18**, 2270 (2001).
12. P. Moutsaki *et al.*, *J. Exp. Biol.*, in press.
13. A. N. Fox, R. J. Pitts, H. M. Robertson, J. R. Carlson, L. J. Zwiebel, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14693 (2001).
14. A. N. Fox, R. J. Pitts, L. J. Zwiebel, *Chem. Senses* **27**, 453 (2002).
15. P. J. Clyne, C. G. Warr, J. R. Carlson, *Science* **287**, 1830 (2000).
16. P. J. Clyne *et al.*, *Neuron* **22**, 327 (1999).
17. L. B. Vosshall, A. M. Wong, R. Axel, *Cell* **102**, 147 (2000).
18. K. F. Storkuhl, R. Kettler, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9381 (2001).
19. C. H. Wetzel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9377 (2001).
20. K. Scott *et al.*, *Cell* **104**, 661 (2001).
21. L. Dunipace, S. Meister, C. McNealy, H. Amrein, *Curr. Biol.* **11**, 822 (2001).
22. A. Dahanukar, K. Foster, W. M. van der Goes van Naters, J. R. Carlson, *Nature Neurosci.* **4**, 1182 (2001).
23. K. Ueno *et. al.*, *Curr. Biol.* **18**, 1451 (2001).


24. We thank P. Russell and M. Hillenmeyer for technical assistance. This investigation received financial support from the United Nations Development Programme/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases (TDR V30/181/208 to L.J.Z.) and the NIH (R01 DC04692-01 to L.J.Z., F31 DC05265-01A1 to A.N.F., U01AI50687 to Celera, U01AI48846 to F.H.C.), and funds from the French Government to Genoscope. Support to Celera, Genoscope, and F.H.C. was provided on behalf of the *Anopheles gambiae* Genome Consortium.


**Supporting Online Material**

www.sciencemag.org/cgi/content/full/298/5591/176/DC1
Materials and Methods
SOM Text
Table S1
Figs. S1 to S5

16 July 2002; accepted 5 September 2002