Chem. Senses 33: 79–93, 2008
doi:10.1093/chemse/bjm067
Advance Access publication October 10, 2007

The Gr Family of Candidate Gustatory and Olfactory Receptors in the Yellow-Fever Mosquito *Aedes aegypti*

Lauren B. Kent, Kimberly K.O. Walden and Hugh M. Robertson

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

Correspondence to be sent to: Hugh M. Robertson, Department of Entomology, University of Illinois at Urbana-Champaign, 320 Morrill Hall, mc-118, 505 S. Goodwin, Urbana, IL 61801 USA. e-mail: hughrobe@life.uiuc.edu

Abstract

The gustatory receptor (Gr) protein family contains most of the diversity in the insect chemoreceptor superfamily, including within it not only taste receptors but select olfactory receptors as well. Manual annotation of the Gr family in the genome sequence of the yellow-fever mosquito, *Aedes aegypti*, yielded a total of 114 potential proteins encoded by 79 genes. In the sequenced genome, 23 of these genes and protein isoforms are pseudogenic, leaving 91 putatively functional Grs. Comparison with our previously published set of 76 Grs encoded by 52 genes in the distantly related *Anopheles gambiae* mosquito revealed 13 new AgGrs encoded by 8 genes. Phylogenetic analysis reveals the conservation of carbon dioxide, sugar, and several orphan receptors in these 2 mosquitoes and *Drosophila* flies. On the other hand, most of these Grs are unique to mosquitoes and many are specific to the *Aedes* or *Anopheles* lineages, indicating their involvement in mosquito-specific aspects of both gustatory and olfactory perception. In particular, most instances of alternative splicing in orthologous loci appear to have evolved after the culicine–anopheline split ±150 million years ago.

Key words: chemoreceptor, gene annotation, gene phylogeny, molecular evolution, olfaction, taste

Introduction

Aedes aegypti ranks high among the mosquitoes of most critical medical significance, being the primary vector of yellow fever and dengue fever, which are responsible for roughly 200 000 and 50–100 million cases of each disease, respectively, worldwide per year (Mackenzie et al. 2004; Tomori 2004). Additionally, this mosquito is able to transmit a variety of other viral diseases as well as filarial worms, and it is also used as a laboratory model for avian malaria (e.g., Thathy et al. 1994; Morlais et al. 2003). In a reemergence and ongoing outbreak of chikungunya virus in India, in which *A. aegypti* is the presumed vector, 1.4 million cases of the disease were reported in 2006 alone (Pialoux et al. 2007). As a consequence of its anthropophily and facile adaptation to breeding in domestic environments, this mosquito has become an efficient disease vector, contributing to thousands of human deaths per year. In addition to its medical significance, *A. aegypti* has been frequently used as a model for physiological studies in insects—including chemoreception and phagostimulation research (e.g., Salama 1966; Lee 1974; Davis 1975; Werner-Reiss et al. 1999), thus laying the framework for application of molecular information unearthed by the recent sequencing of its genome (Nene et al. 2007).

The insect chemoreceptor superfamily is defined as the combination of the odorant receptor (Or) and gustatory receptor (Gr) families; however, the Ors are a single highly expanded lineage within the superfamily, whereas the Grs contain many highly divergent protein lineages that represent most of the diversity in the superfamily (Robertson et al. 2003). The Gr family was so named because most of those first identified in *Drosophila melanogaster* (DmGrs) are expressed in mouthparts and other structures with gustatory functions (Clyne et al. 2000); however, several of the DmGrs are expressed in olfactory organs such as the antennae and consequently are putative olfactory receptors (Scott et al. 2001; Suh et al. 2004; Fishilevich and Vosshall 2005). Indeed, the heterodimer of DmGr21a and DmGr63a was recently shown to be the carbon dioxide receptor in flies (Jones et al. 2007; Kwon et al. 2007). The Grs are defined by a conserved C-terminal motif, immediately after an ancient and common phase 0 intron (Clyne et al. 2000; Scott et al. 2001; Robertson et al. 2003). This motif is hh(G/A/S)

(A/S)hhTYhhhhhQF, where "h" is a hydrophobic residue; however, even the highly conserved TY and QF positions are substituted in some Grs. The insect chemoreceptors are generally considered to have 7 transmembrane (7TM) domains and to be G-protein–coupled receptors (GPCRs) (e.g., Hill et al. 2002); however, hydropathy plots and study of 2 Ors indicate that the transmembrane domain arrangements might not be the same as most 7TM proteins (Benton et al. 2006). Furthermore, these proteins have no sequence similarity to other GPCRs, specifically the vertebrate and nematode chemoreceptors that are members of the rhodopsin superfamily (e.g., Bargmann 2006) and therefore at best are a completely independent class of GPCRs (e.g., Hill et al. 2002) or perhaps function in a completely different way.

Materials and methods

The *A. aegypti* assembled genome sequence v1.0 available at VectorBase, the National Center for Biotechnology Information (NCBI), the Broad Institute, and The Institute for Genomic Research (Nene et al. 2007) was searched using the TBLASTN algorithm (Altschul et al. 1997) for matches to all the AgGrs from *Anopheles gambiae* annotated by Hill et al. (2002), as well as all 68 DmGrs (Robertson et al. 2003) and the 10 AmGrs from the honey bee *Apis mellifera* (Robertson and Wanner 2006). Additional searches were performed in which the EXPECT significance value was relaxed to 1000 to find highly divergent genes sharing only weak statistically insignificant similarity in their TM7 C-terminal regions. Gene models were built manually in the text editor of PAUP* v4.0b10 (Swofford 2002), using the AgGrs as guides when appropriate. The splice site predictor using Neural Network at the Berkeley Drosophila Genome Project (http://www.fruitfly.org/seq_tools/splice.html) was used to help predict splice sites. Regions causing pseudogenization of gene models were checked against the raw reads at the NCBI Trace Archive to establish whether they were polymorphic or misassemblies. Apparent pseudogenes were translated as best possible to produce a comparable protein product for phylogenetic analysis, for example, by ignoring in-frame stop codons in exons and judicious introduction of frameshifts to accommodate insertions and deletions. AaGr and AgGr proteins were aligned with CLUSTALX (Thompson et al. 1997), using multiple alignment mode with default settings. These alignments were used to detect potential problems with the gene models, which were then refined. All AaGr gene models were compared with those available in the 15 419 "high-confidence" gene set in Genebuild 1.0 (AaegL1.1) of the genome annotation as well as 15 396 "supplementary" gene models available at VectorBase (Nene et al. 2007). All amino acid translations are available as Supplementary material online. Note that although these genes and proteins are formally named as GPRgrs in Ensembl and VectorBase, and this convention was used in Hill et al. (2002), we will use the abbreviation Gr with the species suffices Aa and Ag for simplicity and in keeping with conventions in the insect chemoreceptor community (e.g., Dahanukar et al. 2001; Dunipace et al. 2001; Dobritsa et al. 2003; Robertson et al. 2003; Robertson and Wanner 2006).

The phylogenetic analysis employed a final CLUSTALX alignment with a few pseudogenic sequences removed due to their relative incompleteness (see Figure 1). In addition, extreme N- and C-termini were excluded, as a consequence of their highly divergent lengths and sequences. Three short internal regions of alignment gaps found in most sequences were excluded from the data set as well. Phylogenetic analysis of this large set of 200 often highly divergent proteins was performed using corrected distance analysis in TREE-PUZZLE v5.0 (Schmidt et al. 2002) and PAUP*v4.0b10 (Swofford 2002) (see Hill et al. 2002; Robertson et al. 2003; Robertson and Wanner 2006). Bootstrap analysis was performed using 10 000 replications of uncorrected distance analysis in PAUP*. Subtree analysis was performed using custom alignments and phylogenetics described in the Results.

Results

AaGr and AgGr gene models

We identified 79 genes in the AaGr family. We interpret 9 of these genes to be alternatively spliced yielding a potential 114 encoded Gr proteins; however, 23 of these genes or alternatively spliced exons are pseudogenes in the sequenced genome, leaving 91 putatively functional Grs (Table 1, Figures 1 and 2). In *A. gambiae* (Hill et al. 2002) and in *D. melanogaster* (Robertson et al. 2003), some of the chemoreceptor pseudogenes in the sequenced strain are intact in other strains and sometimes even in alternative haplotypes within the sequenced strain. This is also true for some of the AaOrs (Bohbot et al. 2007). We, however, detected no examples of pseudogenes in the available *A. aegypti* assembly that were intact in alternative alleles, and we have not examined other strains of *A. aegypti*. We retained and named the pseudogenes in the AaGr gene set if they encoded more than 200 amino acids, which is roughly 50% of a typical Gr. The relevance of these retained pseudogenes lies in our expectation that some of them will be "flatliners" (Stewart et al. 2005) that are intact in other strains of *A. aegypti*, whereas the remainder provide evidence about how this gene family has evolved. We ignored 6 fragments of genes that encoded less than 200 amino acids.

Unique characteristics of AaGr annotations

The AaGr gene models were often difficult to annotate in part because they commonly encode extraordinarily divergent proteins and many do not have simple AgGr orthologs. Most of the relatively conserved genes and some quite divergent genes were at least partially annotated in the AaegL1.1 Genebuild of the genome-wide automated annotation

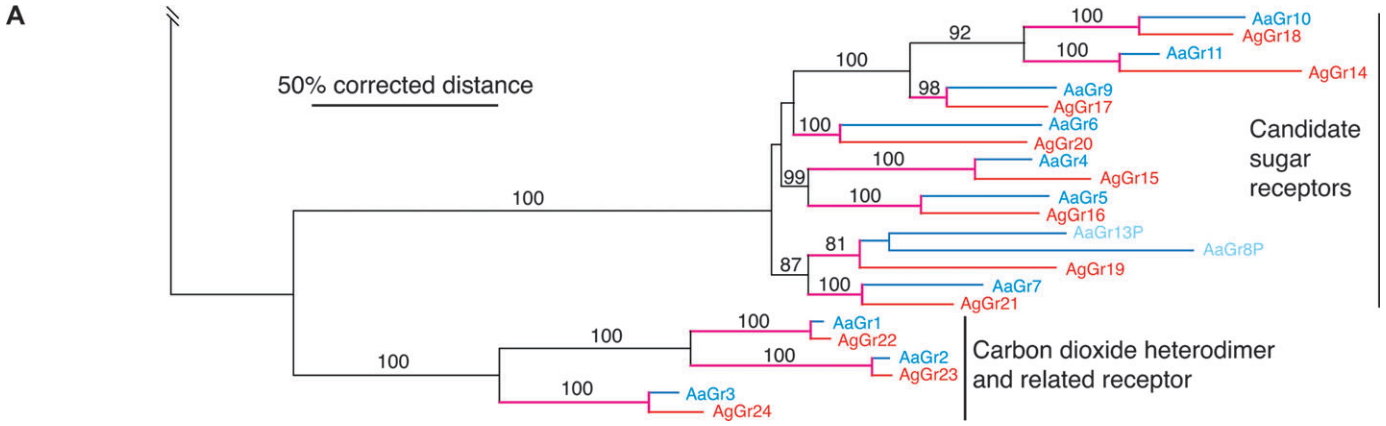

Figure 1 (A–C) Phylogenetic relationships of the *Aedes aegypti* and *Anopheles gambiae* Grs. The tree is rooted at the midpoint in the absence of a clear outgroup. Bootstrap support from 10 000 replications is shown on the relevant branch points, except that bootstrap values are not shown within the 3 largest alternatively spliced loci, as explained in the text. Lineages of particular interest are highlighted on the right. AaGr names and branches are in blue, AgGr in red, and shared branches with bootstrap support are in purple. AaGr12P, 24P, and 30P were not included in the phylogenetic analysis as they are missing their C-terminal regions, and the absence of this conserved region leads to artifactually long branches in the tree. AaGr12P is a close pseudogenic relative of AaGr11, AaGr24P is most similar to AaGr25, and AaGr30P is the apparent ortholog of AgGr47. Protein names followed by a "P" are pseudogenic and shown in lighter blue.

reported in the main genome paper (Nene et al. 2007) and deposited in VectorBase and Ensembl, which consists of 15 419 high-confidence gene models. Forty-one AaGr proteins are at least partially present in AaegL1.1, of which 11 agree with our models. The other 30 require at least some modification (see Table 1). In addition, another 36 gene models are partially represented in the 15 396 gene supplementary gene set available from VectorBase (Table 1). Gr predictions in AaegL1.1 and the supplementary gene set—particularly the alternatively spliced loci—are commonly concatenated to encode single large proteins, unlike the unique splice variants found in our annotations. Many of the genes contain long introns that include retrotransposons, which is a general feature of this genome (Nene et al. 2007). These were considered to be intact genes if the retrotransposons did not appear to affect the intron splice sites. The 23 pseudogenes are denoted by the suffix P in Table 1. All our gene models have been communicated to VectorBase for inclusion in the next version of the *A. aegypti* genome annotation, which will also be deposited in Ensembl.

There were several instances of highly divergent AaGrs and AgGrs that did not appear to have obvious orthologs in the other species. We searched intensively for possible orthologs using TBLASTN queries and also undertook extensive PSI-BLASTP searches at GenBank. The latter were executed starting with each AaGr and AgGr in turn as a query and reiterated until the only new proteins from *A. aegypti* or *A. gambiae* were the distantly related Ors (Robertson et al. 2003). This search method depends on the genes already being at least partially annotated in the automated AaegL1.1 Genebuild for *A. aegypti* or the Ensembl annotations for *A. gambiae*. Together these searches using the divergent AaGrs led to the discovery of 8 new AgGr genes, AgGr53–60, encoding 13 new AgGrs through alternative splicing of the AgGr56 locus (Figure 2; Table 2). These proteins are so highly divergent from other AgGrs that in TBLASTN searches they find no other matches in the genome, explaining why they were not discovered by Hill et al. (2002). Initially, gene models for only 2 of these new AgGrs were available from Ensembl. However, neither was complete for the C-terminus; hence, they could not be found by PSI-BLASTP searches at NCBI because the only conserved motif was absent. These raise our previously reported *A. gambiae* Gr repertoire of 76 Grs encoded by 52 genes (Hill et al. 2002) to 90 Grs encoded by 60 genes and suggest that there might be additional undetected highly divergent Grs in either or both species. In addition, comparisons of AaGr and AgGr gene models allowed refinement of 15 AgGr gene models. The 8 newly recognized and 15 refined AgGr gene models have been communicated to VectorBase for inclusion in the next release of the *A. gambiae* genome annotation, which will also be deposited in Ensembl.

Orthologous Grs in the 2 mosquitoes

Phylogenetic analysis comparing the AaGrs with the entire set of AgGrs reveals several instances of highly conserved apparent orthologs, most prominently the AaGr1-3 and AgGr22-24 subfamily lineages which share 72–89% identity (Figure 1A). The orthologs of AaGr1:AgGr22 and AaGr3:AgGr24 in *D. melanogaster* are DmGr21a and DmGr63a, respectively (Hill et al. 2002), with an *Aedes*: *Drosophila* amino acid identity of 70% and 60%, respectively. DmGr21a and DmGr63a have recently been shown to function as a heterodimeric receptor for carbon dioxide (Jones et al. 2007; Kwon et al. 2007), which is a major chemical cue used by mosquitoes to find their vertebrate hosts.

Another prominent conserved subfamily is related to the trehalose receptor of *D. melanogaster*, DmGr5a (Chyb

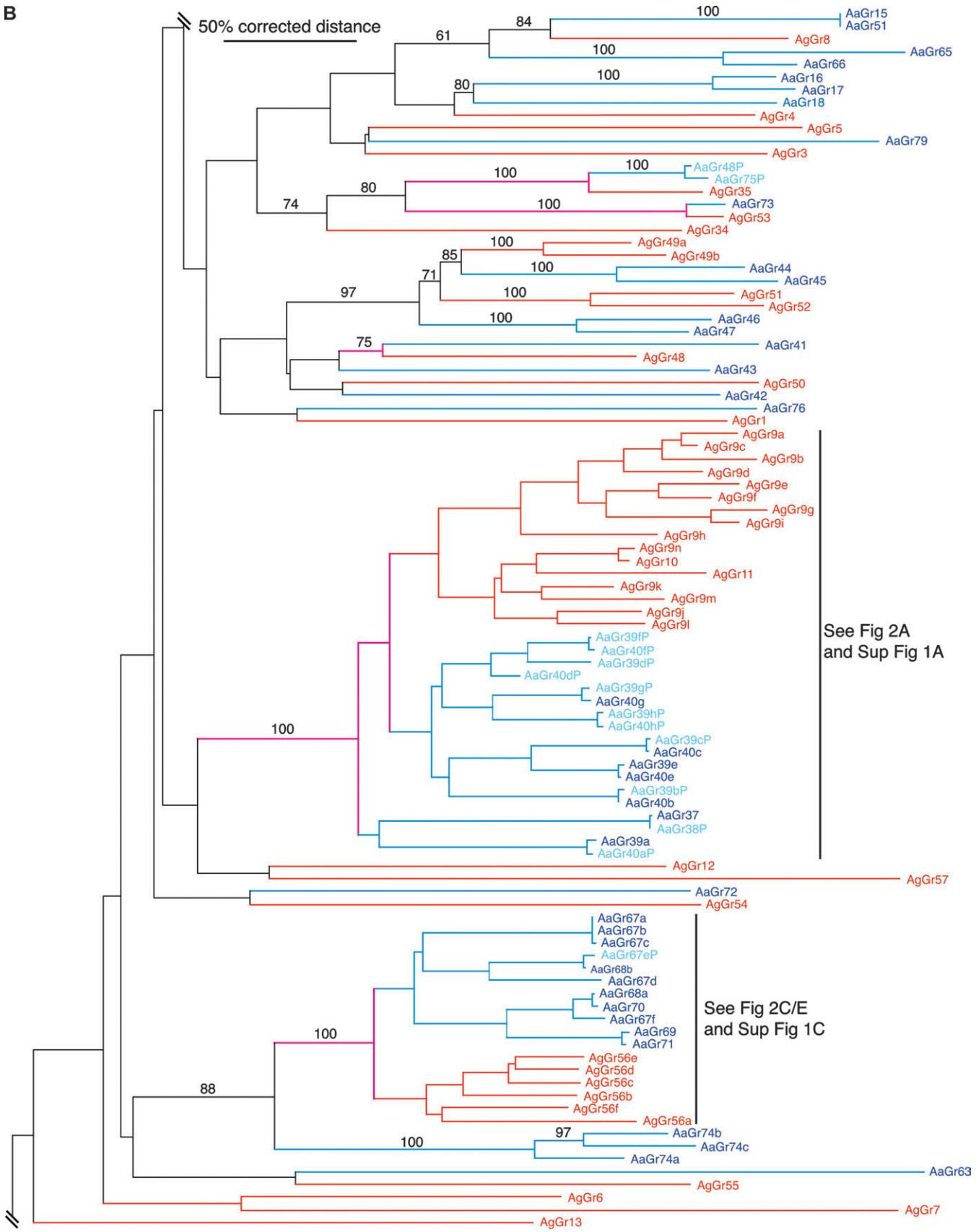

Figure 1 Continued.

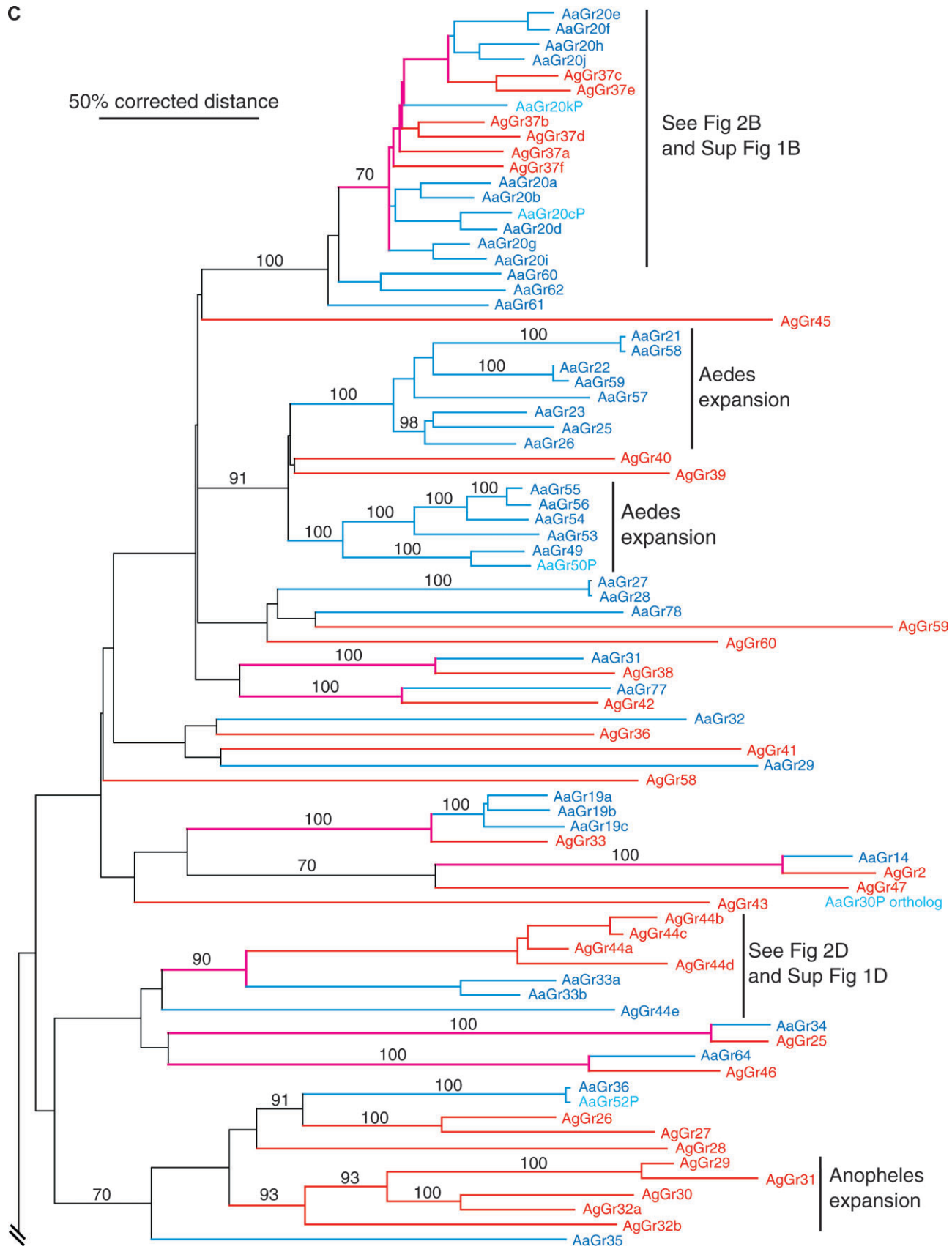

Figure 1 Continued.

Table 1 Details of the 79 AaGrs including genomic location, current annotations, and comments on the gene models

Gene name	Putative ortholog	GenBank contig	Supercontig	Base pair range on supercontig	AaegL1.1 VectorBase accession number	AaegL1.1 GenBank accession number	Comments
AaGr1	AgGr22	1.3634	1.55	922180–894114	AAEL002380	EAT46439.1	
AaGr2	AgGr23	1.3358	1.50	1749145–1747646	AAEL002167	EAT46689.1	
AaGr3	AgGr24	1.17909	1.450	643060–644583	AAEL010058	EAT38010.1	
AaGr4	AgGr15	1.39	1.1	2297653–2299068	AAEL000048	EAT48931.1	AaegL1.1 includes extra 16 aa at N-terminus.
AaGr5	AgGr16	1.40	1.1	2320762–2323613	AAEL000043	EAT48932.1	AaegL1.1 and our version differ in the first 32 aa.
AaGr6	AgGr20	1.46	1.1	2681591–2683193	AAEL000012	EAT48938.1	
AaGr7	AgGr21	1.45	1.1	2593701–2614090	AAEL000060	EAT48937.1	AaegL1.1 missing first 40 aa.
AaGr8P	AgGr19	1.41	1.1	2396677–2384265	AAEL000069	EAT48936.1	Pseudogene: stop codon in first exon. AaegL1.1 missing first 57 aa.
AaGr9	AgGr17	1.40	1.1	2348850–2350981	AAEL000075	EAT48934.1	
AaGr10	AgGr18	1.40	1.1	2334445–2348734	AAEL000082	EAT48933.1	AaegL1.1 missing first 6 aa.
AaGr11	AgGr14	1.20181-1.20180	1.548	103691–93118	AAEL011174	EAT36763.1	AaegL1.1 missing first 91 aa.
AaGr12P	AgGr14	1.31680	1.1446	94601–95745	AAEL015071	EAT32706.1	Pseudogene; interrupted by transposon. AaegL1.1 missing first 91 aa.
AaGr13P	AgGr19	1.14978	1.339	1222635–1224116	SUPP_AEDES008446 and SUPP_AEDES008454		Pseudogene; stop codon in first exon. AaegL1.1 supplementary peptides match aa 1–65 and 213–385, respectively.
AaGr14	AgGr2	1.21018	1.590	30124–37038	AAEL011571	EAT36329.1	AaegL1.1 adds 88 aa to N-terminus.
AaGr15		1.14987	1.340	299396–298037			
AaGr16		1.14986	1.340	232802–234345			
AaGr17		1.14986	1.340	244814–247018	SUPP_AEDES008465		AaegL1.1 supplementary peptides match aa 56–169.
AaGr18	AgGr4	1.14986	1.340	267063–268305			
AaGr19a		1.20006	1.539	496893–534546	SUPP_AEDES010755		AaegL1.1 supplementary peptides match aa 1–179.
AaGr19b		1.20007	1.539	510124–534546	AAEL011077	EAT36879.1	AaegL1.1 missing first 99 aa and last 96 aa.
AaGr19c		1.20008-1.20009	1.539	522771–534546	AAEL011073 and SUPP_AEDES014926	EAT36880.1	AaegL1.1 supplementary peptides match aa 193–354.
AaGr20a		1.13511	1.290	1031309–920885			
AaGr20b		1.13511	1.290	1029522–920885			
AaGr20cP		1.13511	1.290	1014213–920885			Pseudogene; frameshift in first exon.
AaGr20d		1.13511	1.290	1010330–920885			
AaGr20e		1.13511	1.290	999491–920885	AAEL007937	EAT40328.1	AaegL1.1 differs/missing last 146 aa.
AaGr20f		1.13511	1.290	985517–920885	AAEL007935	EAT40327.1	AaegL1.1 appears to translate through parts of introns.

Table 1 Continued

Gene name	Putative ortholog	GenBank contig	Supercontig	Base pair range on supercontig	AaegL1.1 VectorBase accession number	AaegL1.1 GenBank accession number	Comments
AaGr20g		1.13511	1.290	984173–920885	AAEL007935	EAT40327.1	AaegL1.1 appears to translate through parts of introns and appends 160 extra aa at N-terminus.
AaGr20h		1.13511	1.290	971628–920885			
AaGr20i		1.13511	1.290	941989–920885			
AaGr20j		1.13511	1.290	940741–920885			
AaGr20kP		1.13510	1.290	931477–920885			Pseudogene; stop codon in first exon.
AaGr21		1.24951	1.804	137675–136386	AAEL013192	EAT34575.1	AaegL1.1 missing much of the last 77 aa.
AaGr22		1.24951	1.804	135737–134357	AAEL013190	EAT34574.1	
AaGr23		1.24950	1.804	100268–101638	AAEL013196	EAT34571.1	
AaGr24P		1.24950	1.804	106866–107730	SUPP_AEDES012681		Pseudogene; 2 stop codons in first exon. AaegL1.1 supplementary peptides match aa 1–132.
AaGr25		1.24950	1.804	111266–112640	AAEL013193	EAT34572.1	AaegL1.1 missing last 28 aa, and differs in its first intron boundaries.
AaGr26		1.24950	1.804	123735–122387	AAEL013195	EAT34573.1	AaegL1.1 differs in its second intron boundaries.
AaGr27		1.34685	1.3237	1978–3327	AAEL015556	EAT32317.1	
AaGr28		1.18321	1.467	681026–679677	AAEL010277	EAT37763.1	
AaGr29		1.18319	1.467	602167–603447	AAEL010275	EAT37762.1	AaegL1.1 missing first 30 aa.
AaGr30P		1.26154	1.889	214955–216699	AAEL013646	EAT34079.1	Pseudogene; C-terminus deleted. AaegL1.1 missing first 21 aa and last 67 aa.
AaGr31C	AgGr38	1.13506-1.13505	1.290	539562–532983	AAEL007933	EAT40324.1	Partially annotated; unable to resolve C-terminus. AaegL1.1 missing last 48 aa.
AaGr32		1.18322	1.467	726756–742750	SUPP_AEDES010098 and SUPP_AEDES010086		AaegL1.1 supplementary peptides match aa 1–272 and 1–125, respectively.
AaGr33a		1.9018	1.168	1143715–1152756	SUPP_AEDES005411		AaegL1.1 supplementary peptides match aa 322–398.
AaGr33b		1.9018	1.168	1148504–1152756	SUPP_AEDES005411		AaegL1.1 supplementary peptides match aa 143–266, 327–403.
AaGr34	AgGr25	1.244	1.3	1696832–1653660	AAEL000162	EAT48790.1	AaegL1.1 missing seventh exon and last 87 aa.
AaGr35		1.3605	1.54	3104490–3103063	SUPP_AEDES002361		AaegL1.1 supplementary peptides match aa 22–345.
AaGr36		1.701	1.8	4564006–4562590			
AaGr37		1.11997	1.246	1522840–1524097			
AaGr38P		1.27227	1.965	55071–58797			Pseudogene; interrupted by transposon.
AaGr39a		1.11997-1.12000	1.246	1561309–1601062			
AaGr39bP		1.11998-1.12000	1.246	1566147–1601062			Pseudogene; frameshift in first exon.

Table 1 Continued

Gene name	Putative ortholog	GenBank contig	Supercontig	Base pair range on supercontig	AaegL1.1 VectorBase accession number	AaegL1.1 GenBank accession number	Comments
AaGr39cP		1.11998-1.12000	1.246	1572168–1601062			Pseudogene; early frameshift in first exon.
AaGr39dP		1.11998-1.12000	1.246	1573912–1601062			Pseudogene; stop codon in first exon.
AaGr39e		1.11998-1.12000	1.246	1584353–1601062			
AaGr39fP		1.11998-1.12000	1.246	1586772–1601062			Pseudogene; stop codon in first exon.
AaGr39gP		1.11998-1.12000	1.246	1588388–1601062			Pseudogene; early frameshift in first exon.
AaGr39hP		1.11999-1.12000	1.246	1598834–1601062			Pseudogene; stop codon in first exon.
AaGr40aP		1.27228-?	1.965	103063–139050			Pseudogene; early frameshift and stop codon in first exon.
AaGr40b		1.27229-?	1.965	107323–139050			
AaGr40c		1.27229-?	1.965	117663–139050			
AaGr40dP		1.27229-?	1.965	119415–139050			Pseudogene; stop codon in first exon.
AaGr40e		1.27230-?	1.965	127067–139050			
AaGr40fP		—	—	—			Pseudogene; found in trace files only.
AaGr40g		—	—	—			Found in trace files only.
AaGr40hP		1.27231-?	1.965	138763–139050			Pseudogene; stop codon in first exon.
AaGr41	AgGr48	1.10557	1.208	1602623–1601167	AAEL006500	EAT41925.1	AaegL1.1 missing last 162 aa.
AaGr42		1.10557	1.208	1602697–1604191			
AaGr43		1.10557	1.208	1605186–1606886			
AaGr44		1.10556	1.208	1521955–1520529	SUPP_AEDES006276		AaegL1.1 matches aa 144–273 and 324–347.
AaGr45		1.10556	1.208	1535088–1523536	AAEL006494	EAT41923.1	AaegL1.1 matches ours only from third through fourth exons.
AaGr46		1.10556	1.208	1593009–1591578	SUPP_AEDES006283		AaegL1.1 matches aa 368–417.
AaGr47		1.10554	1.208	1318760–1320644			
AaGr48P		1.10889	1.217	1150040–1148630			Pseudogene; stop codon in first exon.
AaGr49		1.18322	1.467	742842–744310	SUPP_AEDES010089 and SUPP_AEDES010100		AaegL1.1 matches aa 28–211 and 232–398, respectively.
AaGr50P		1.18322	1.467	777066–778434	SUPP_AEDES010090		Pseudogene; frameshift in first exon. AaegL1.1 matches aa 49–97 and 293–371.
AaGr51		1.14987	1.340	302609–303968			
AaGr52P		1.3607	1.54	3172207–3164557			Pseudogene; interrupted by transposon.
AaGr53		1.18323	1.467	802925–801556	AAEL010279	EAT37764.1	
AaGr54		1.18323	1.467	805546–804176	SUPP_AEDES010092 and SUPP_AEDES010093		AaegL1.1 matches aa 1–120 and (174–312 + 366–393), respectively.
AaGr55		1.18324	1.467	829268–827893	AAEL010274	EAT37765.1	AaegL1.1 missing first and last 81 aa.

Table 1 Continued

Gene name	Putative ortholog	GenBank contig	Supercontig	Base pair range on supercontig	AaegL1.1 VectorBase accession number	AaegL1.1 GenBank accession number	Comments
AaGr56		1.18324	1.467	845688–847078	AAEL010278	EAT37766.1	AaegL1.1 missing first intron.
AaGr57		1.18324	1.467	854747–853359	AAEL010272	EAT37767.1	AaegL1.1 missing last 196 aa.
AaGr58		1.24952	1.804	139951–141240	AAEL013200	EAT34576.1	AaegL1.1 is a combination of GPRgr58 and GPRgr59; it matches our version only through the first 306 aa.
AaGr59		1.24952	1.804	141457–142836	AAEL013200	EAT34576.1	AaegL1.1 is a combination of GPRgr58 and GPRgr59; it matches our version only through the last 353 aa.
AaGr60		1.13510-1.13509	1.290	908743–888437	SUPP_AEDES007704 and SUPP_AEDES007695		AaegL1.1 matches aa 1–204 and (204–257 + 323–407), respectively.
AaGr61		1.13508-1.13509	1.290	830477–875333	SUPP_AEDES007702		AaegL1.1 matches aa 315–338 and 368–399.
AaGr62		1.13508	1.290	805173–812399	SUPP_AEDES007702		AaegL1.1 matches aa 1–76 and 119–257.
AaGr63	AgGr45	1.11712	1.238	814303–801691	SUPP_AEDES006853		AaegL1.1 matches aa 72–240 and 241–294.
AaGr64l	AgGr46	1.11712-1.11715	1.238	836965–918890	AAEL007142 and SUPP_AEDES006859	EAT41198.1	Partially annotated; internal exons missing from annotation. AaegL1.1 matches ours only through first 253 aa. AaegL1.1 supplementary peptides match aa 360–417.
AaGr65		1.14987	1.340	292334–290974			
AaGr66		1.14986	1.340	280911–282245			
AaGr67a		1.17432	1.430	959198–915385	SUPP_AEDES009701 and SUPP_AEDES009702		AaegL1.1 matches aa 289–365 and (1–288 + 34–288; 2 different regions), respectively.
AaGr67b		1.17432	1.430	952565–915385	Same as above		same as above
AaGr67c		1.17432	1.430	943287–915385	SUPP_AEDES009702		AaegL1.1 matches aa 1–288.
AaGr67d		1.17432	1.430	938093–915385	SUPP_AEDES009686		AaegL1.1 matches aa 261–305.
AaGr67eP		1.17430	1.430	920003–915385	SUPP_AEDES009701		Pseudogene; frameshift in first exon. AaegL1.1 supplementary peptides match aa 117–377.
AaGr67f		1.17430	1.430	916686–915385	SUPP_AEDES009701		AaegL1.1 matches aa 289–366.
AaGr68a		1.17429	1.430	886266–884970	SUPP_AEDES009704		AaegL1.1 matches aa 264–366.
AaGr68b		1.17428	1.430	905705–884970	SUPP_AEDES009701 and SUPP_AEDES009704		AaegL1.1 matches aa 100–301 and 301–378, respectively.
AaGr69		1.17428	1.430	884290–876803			
AaGr70		1.28444	1.1070	249414–248120	SUPP_AEDES009704		AaegL1.1 matches aa 264–363.
AaGr71		1.28444-1.28442	1.1070	246143–229686	SUPP_AEDES013919		AaegL1.1 matches aa 142–290.
AaGr72l		1.21065	1.593	574451–584358			Partially annotated; internal exons missing from annotation.

Table 1 Continued

Gene name	Putative ortholog	GenBank contig	Supercontig	Base pair range on supercontig	AaegL1.1 VectorBase accession number	AaegL1.1 GenBank accession number	Comments
AaGr73I	AgGr53	1.19698	1.526	483899–529565	AAEL010962, SUPP_AEDES010636, and SUPP_AEDES010633	EAT36998.1	Partially annotated; internal exons missing from annotation. AaegL1.1 missing first 72 aa. AaegL1.1 supplementary peptides match 1–72 and 73–132, respectively.
AaGr74a		1.26594	1.918	229714–287726	SUPP_AEDES013275, SUPP_AEDES009690, and SUPP_AEDES013279		AaegL1.1 matches aa 1–274, 1–219, and 297–328 respectively.
AaGr74b		1.26596	1.918	269060–287726	SUPP_AEDES013282 and SUPP_AEDES013279		AaegL1.1 matches aa 23–295 and 329–360, respectively.
AaGr74c		1.26596	1.918	270575–287726	SUPP_AEDES013282 and SUPP_AEDES013279		AaegL1.1 matches aa 145–284 and 307–338, respectively.
AaGr75P		1.10891	1.217	1159284–1157910	SUPP_AEDES006470		Pseudogene; stop codon in first exon. AaegL1.1 supplementary peptides match aa 329–394.
AaGr76		1.16782	1.405	885272–860883	AAEL009545	EAT38580.1	
AaGr77	AgGr42	1.13507	1.290	667711–610378	AAEL007940 and SUPP_AEDES007693	EAT40325.1	Partial match; AaegL1.1 protein is only 125 aa long; matches our annotation from aa 135–229. AaegL1.1 supplementary peptides match aa 274–430.
AaGr78		1.21573	1.616	563816–565425	SUPP_AEDES011407		AaegL1.1 matches aa 1–126.
AaGr79		1.14986	1.340	261762–260371			

aa, amino acid.

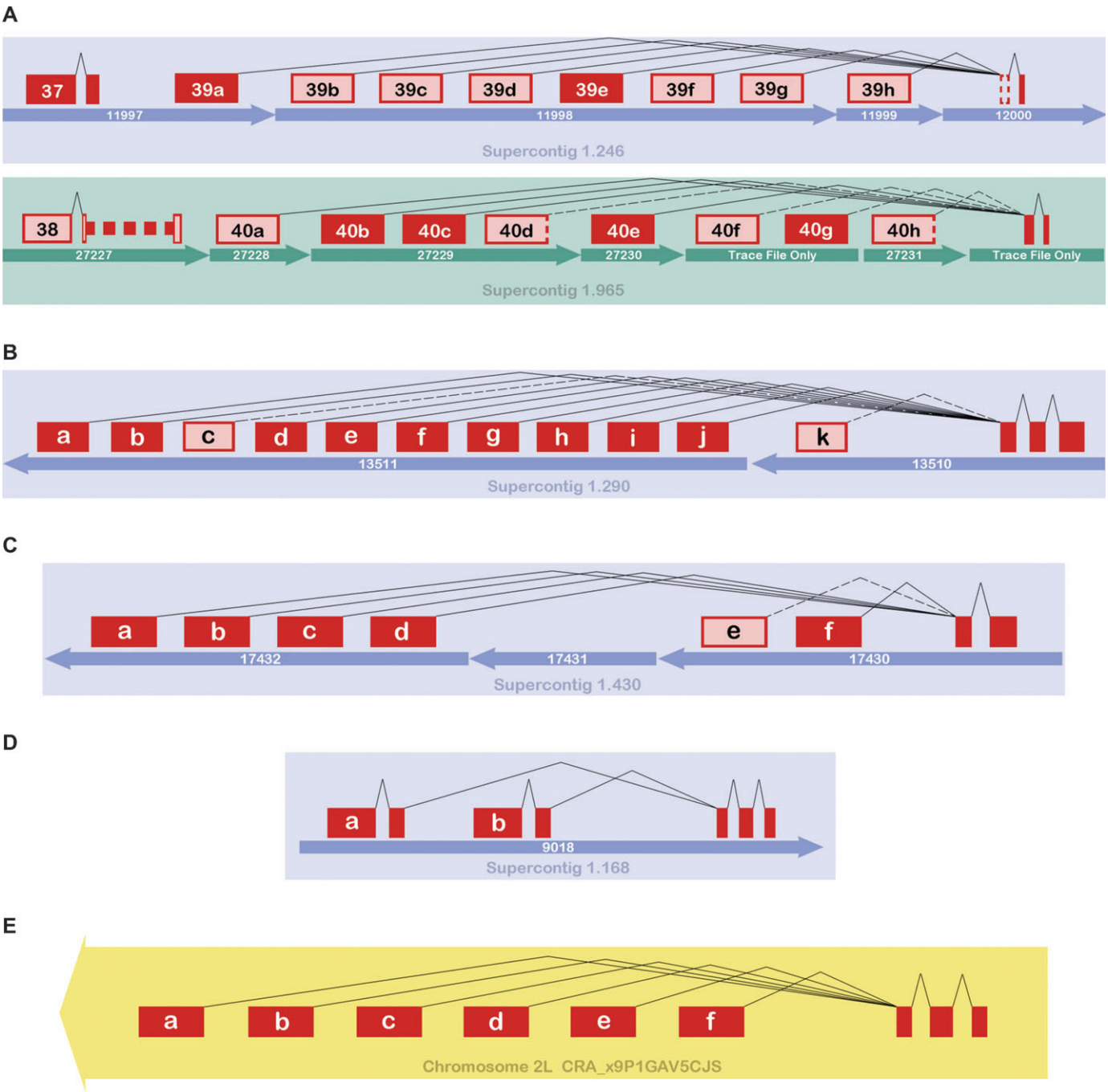

Figure 2 Schematic diagrams of 4 of the 8 *Aedes aegypti* gustatory receptor (AaGr) loci and one new *Anopheles gambiae* Gr locus inferred to be alternatively spliced (not drawn to scale). All genes are shown as 5′-3′ with the alternatively spliced N-terminal exon on the left and the shared C-terminal exons on the right, whereas the contigs that encode them might be oriented in either direction (indicated by arrows). **(A)** AaGr39 and AaGr40 are thought to result from a duplication event in *A. aegypti* resulting in splice isoforms with high similarity in both amino acid sequence and proposed splicing patterns. In some cases, splice isoforms have become pseudogenes (light boxes) in one or both genes. Isoforms with dashed 3′ ends were only partially annotated as a result of reaching the end of a contig. The first exon encoding the shared C-terminus of AaGr39 (shown on contig 12 000) was not found (dotted outline box). Like other parts of the AaGr40 locus, which crosses shorter contigs that are not as well assembled as those encoding AaGr39, the 2 C-terminal exons were only found in raw trace reads. It cannot formally be determined whether the traces correspond to AaGr40 exclusively or AaGr39 as well. Dashed splice lines indicate instances where sequence was not complete enough to find the splice donor. The upstream loci AaGr37 and AaGr38P also appear to be the result of the same duplication, but with a 2.5-kb transposon inactivating the latter through insertion in the second exon (dashed line). **(B)** The AaGr20 locus. **(C)** The AaGr67 locus. **(D)** The AaGr33 locus. **(E)** The AgGr56 locus.

Table 2 Details of the 13 newly recognized AgGrs including genomic location, current annotations, and comments on the gene models

Gene name	Chromosome	Base pair range on chromosome arm	GenBank accession number	Base pair range on GenBank scaffold	Comments
AgGr53	2R	24698605–24694665	AAAB01008859.1	6834985–6838925	GenBank version differs in last intron boundary.
AgGr54	2R	54385686–54384164	AAAB01008898.1	548989–547467	
AgGr55	2L	39995592–39994248	AAAB01008807.1	9368734–9370078	
AgGr56a	2L	27145481–27137739	AAAB01008960.1	13201059–13193317	GenBank version concatenates splice variants into one massive gene.
AgGr56b	2L	27144445–27137739	AAAB01008960.1	13200023–13193317	
AgGr56c	2L	27143405–27137739	AAAB01008960.1	13198983–13193317	
AgGr56d	2L	27142291–27137739	AAAB01008960.1	13197869–13193317	
AgGr56e	2L	27141048–27137739	AAAB01008960.1	13196626–13193317	
AgGr56f	2L	27139089–27137739	AAAB01008960.1	13194667–13193317	
AgGr57	2L	2624121–2627818	AAAB01008968.1	1406954–1403257	
AgGr58	2R	454505–453144	AAAB01008987.1	15768093–15769454	
AgGr59	2R	568742–567440	AAAB01008987.1	15653856–15655158	
AgGr60	2R	446801–445364	AAAB01008987.1	15775797–15777234	

et al. 2003) (Figure 1A). There are 7 relatives of DmGr5a in the *D. melanogaster* genome, all of which are therefore candidate sugar receptors (Robertson et al. 2003). *Anopheles gambiae* also has 8 genes in this lineage (AgGr14-21) (Hill et al. 2002), and *A. aegypti* has apparently simple orthologs for each of these (AaGr4-13) ranging in amino acid identity from 39% to 65%, except that AaGr8 and AaGr13 are pseudogenes most closely related to AgGr19 (31% and 34% identity, respectively), and AaGr12 is a truncated pseudogenic copy of AaGr11 not shown in Figure 1A. If these are indeed all sugar receptors then it might be of biological interest that *A. aegypti* apparently does not have the sensory capacity conferred upon *A. gambiae* by AgGr19. These 8 mosquito candidate sugar receptors are not simple orthologs of the 8 in *Drosophila* (Hill et al. 2002; Robertson and Wanner 2006; Kent LB, Robertson HM, unpublished data). Consequently, drawing connections to the ligand specificity of these proteins in *Drosophila*, once established, will not necessarily be straightforward.

Other apparent simple 1:1 orthologs include AaGr73:AgGr53 (71% amino acid identity), AaGr34:AgGr25 (70%), AaGr14:AgGr2 (54%), AaGr64:AgGr46 (47%), AaGr48P/75P:AgGr35 (43%), AaGr31:AgGr38 (37%), AaGr77:AgGr42 (32%), and AaGr41:AgGr48 (28%) in Figure 1B/C. In addition, AaGr30P is a pseudogene relative of AgGr47 truncated by apparent loss of the last 2 exons (37% identity in shared N-terminal region). Of the ortholog pairs listed, only AaGr34:AgGr25 and AaGr14:AgGr2 show conservation to *D. melanogaster*, specifically with DmGr43a and DmGr66a, respectively (Hill et al. 2002; Robertson and Wanner 2006), having an *Aedes*:*Drosophila* amino acid identity of 40% and 37%, respectively. Interestingly, DmGr66a was recently demonstrated to be required for caffeine perception in *Drosophila* (Moon et al. 2006), suggesting a similar role for its orthologs in mosquitoes. Despite their conservation in mosquitoes at 71% amino acid identity, second only to the carbon dioxide subfamily gene conservation, *Drosophila* species do not have an ortholog of the AaGr73/AgGr53 gene pair, and this conserved gene is absent from the silk moth, flour beetle, and honey bee genomes. These 8 "orthologous" pairs are likely to have similar ligands and roles in the 2 mosquitoes, and sometimes even *Drosophila*, and AaGr73:AgGr53 is of particular interest as an apparently mosquito-specific receptor. Another noteworthy orthologous relationship is the many-to-one association of the alternatively spliced locus AaGr19:AgGr33, which also has an alternatively spliced ortholog in *Drosophila*, DmGr28b (Hill et al. 2002), and even distant relatives in honey bee (Robertson and Wanner 2006). Other orthologous relationships of alternatively spliced loci, including AaGr20:AgGr37, AaGr33:AgGr44, AaGr39/40:AgGr9, and AaGr67/68:AgGr56a-f, do not show conservation in *Drosophila* and are detailed further below.

Mosquito-specific Grs

There are many highly divergent AaGrs and AgGrs without obvious orthologs in the other species, and it is likely that for many of them the ortholog in the other species has been lost. The remaining genes in each species represent candidates for mediating species-specific behaviors. Expanded Gr

subfamily lineages in *Aedes* and *Anopheles* are scattered throughout the phylogenetic tree and highlighted in Figure 1B,C. The largest of these in *Aedes* is a subfamily of 8 genes and in *Anopheles*, a subfamily of 5 genes (Figure 1C). These too likely mediate species-specific behaviors.

Several of the alternatively spliced genes also exhibit species-specific expansions (Figures 1 and 2), and for 3 of them (AaGr20:AgGr37, AaGr39/40:AgGr9, and AaGr67/68:AgGr56), the tree suggests that most or all the alternative splicing originated independently in each species lineage. The second locus is particularly complicated because it is duplicated in *Aedes* (AaGr39/40), and different alternatively spliced isoforms have become pseudogenes in the 2 loci (Figure 2), leaving a set of 6 intact Grs in *Aedes* compared with 14 in *Anopheles*. These alternatively spliced loci encode Gr isoforms that differ in their N-termini but share a common C-terminus (Figure 2), an architecture also shared by several DmOrs and DmGrs (Clyne et al. 1999, 2000; Robertson et al. 2003) and one AaOr locus (Bohbot et al. 2007). The extent of shared sequence varies from gene to gene, but the differing N-terminal regions are always at least one half of the protein, are usually encoded by a single long alternatively spliced 5′ exon, and are presumed to confer distinct ligand-binding properties on the resultant protein isoforms. Because the isoforms from each locus cluster with each other by species in the Gr tree (Figure 1), it appears that they originated independently in each mosquito lineage over the past ~150 million years (Myr) since the split of the culicine (*Aedes*) and anopheline (*Anopheles*) subfamily lineages (Krzywinski et al. 2006). However, this conclusion could be biased by the necessary inclusion of the shared C-terminal sequences for each isoform in the alignments on which the Gr tree is based, which would tend to cause the isoforms from each species to cluster within the species rather than with potentially orthologous isoforms from the other species. Therefore, we do not include bootstrap support values for these branches in Figure 1.

To address this issue further, we undertook focused phylogenetic analysis of each set of alternatively spliced loci and after alignment of their encoded isoforms removed the shared C-terminal regions so that the phylogenetic analyses were based only on their differing N-terminal regions encoded by the long alternatively spliced exons. Ideally, these analyses should be rooted with the next closest Gr protein sequence in the tree declared as the outgroup. However, in the absence of the conserved C-terminal regions, these relatives are usually so highly divergent as to be ineffective as outgroups; hence, these subtrees are rooted at the midpoint, which still reveals how they evolved. These subtrees in Supplementary Figure 1 largely support the inference that at each locus most or all the alternatively spliced forms originated independently in each mosquito lineage. The only clear-cut exception is AaGr20:AgGr37 where 2 alternatively spliced N-terminal exons appear to have been present before these mosquito lineages split (purple branches in Supplementary Figure 1B), and they both were subsequently duplicated independently and repeatedly in each mosquito.

The argument for independent origin of splice forms is further supported by comparison of the AaGr39/40 and AgGr9 loci. In the former case, there is an upstream solo copy of a paralogous gene (AaGr37/38, see Figure 2A), whereas in the latter, there are 2 downstream paralogous solo loci (AgGr10 and 11, see Hill et al. 2002). It nevertheless seems more than coincidental that all 3 of these orthologous loci should have undergone such extensive duplication of their alternatively spliced 5′ exons independently in each lineage, and we propose that in fact they were alternatively spliced in their common mosquito ancestor, but that each lineage lost all but one (or 2 for AaGr20 and AgGr37) of these alternative 5′ exons independently and ended up with nonorthologous 5′ exons which were subsequently duplicated again independently in each mosquito lineage. Such extensive loss of Gr-coding sequences is to be expected. Indeed as noted above, orthologs of many of the Grs in each species have clearly been lost from the other (Figure 1) and similar losses and duplications are seen even in close comparisons of 1- to 50-Myr-old *Drosophila* species (Guo and Kim 2007; Nozawa and Nei 2007; Robertson HM, unpublished data).

Finally, there are several AaGr pairs that seem to represent extremely recent duplication events within the *Aedes* lineage; however, an alternative explanation is that these are alternative assemblies of haplotypes or alleles. For example, AaGr27 is almost identical to AaGr28 but is encoded by a 7-kb contig that is >90% identical along >90% length to the equivalent region of the AaGr28 contig. Similarly, AaGr21/22 are very similar to AaGr58/59, and the 7-kb contig that encodes them is 95% identical to the start of the contig that encodes AaGr58/59. We excluded 4 other examples of gene fragments in short contigs entirely represented elsewhere within much larger contigs from the gene set as representing alternative haplotypes or alleles. The almost identical AaGr15/51 duplication appears to be real, however, in an inverted orientation in the same contig. Additionally, the AaGr48P/75P duplication copies are in tandem in neighboring contigs and 90% amino acid identical but with different stop codons inactivating them. Finally, the AaGr36/52P pair are clearly separate loci with equal read depths of about 8× each, yet AaGr52P has an insertion of a copy of the 6.3-kb Loner_Ele2 retrotransposon documented in the TEfam database (see Nene et al. 2007, supplementary information).

Discussion

Using the AaegL1 assembly of the *A. aegypti* genome sequence, we report our annotation of this mosquito's 114 identifiable Grs and their phylogenetic relationships with previously annotated members of this gene family in other insects, primarily the African malaria vector mosquito, *A. gambiae*. We have chosen to number these AaGrs sequentially and independently of the AgGr numbering system,

despite the fact that several of them are apparent orthologs of AgGrs (Figure 1 and Table 1). Despite the similarities between the genes, we believe that coordinating the naming schemes would introduce a level of ambiguity with respect to the inferred significance of the names. Confident orthology cannot be garnered at the exploratory level, and by prematurely extending putative orthology to a permanent name, we would in some cases lose clarity in defining where the orthology actually ends.

Our analysis reveals remarkable similarities between these distantly related mosquitoes. If we were to tally our final AaGr number in the same manner as done with the AaOrs (Bohbot et al. 2007), 6 pseudogene fragments and 4 excluded alternative haplotypes would bring the total to 124. However, when we take into account pseudogenes and possible alternate haplotype loci and focus on functional genes, there are 91 AaGrs. This is essentially the same as the AgGrs, now totaling 90. Remarkably, there are no apparent pseudogenes in *Anopheles*, although a few unannotated fragments may have indicated that some lost genes had already evolved beyond the scope of our identification. This pattern of additional pseudogenes in *Aedes* reflects the accumulation of transposons in more than 50% of the genome sequence (Nene et al. 2007).

One of our more noteworthy findings is that AaGr1:AgGr22 and AaGr3:AgGr24 are among the most highly conserved putative orthologs found in the mosquito Gr repertoire. These 2 proteins form a heterodimeric receptor for carbon dioxide perception (Jones et al. 2007; Kwon et al. 2007). The high amino acid conservation in these 2 orthologous pairs, extending to the *D. melanogaster* orthologs at 70% and 68% identity, respectively, presumably reflects stringent requirements for their structures both for heterodimerization and binding of carbon dioxide or a derivative, such as HCO_3. These 2 proteins are also encoded by the silk moth, *Bombyx mori*, and flour beetle, *Tribolium castaneum*, genomes at 69/67% and 61/60% identity, respectively (unpublished results); however, the entire gene lineage is absent from the more basal honey bee *A. mellifera* genome (Robertson and Wanner 2006). *Drosophila melanogaster* has lost the third even more highly conserved gene in this subfamily, AaGr2:AgGr23 (89% identity, and 74% to *B. mori* and 63% to *T. castaneum*), and this gene loss occurred over 50 Myr ago because it is not present in the other 12 *Drosophila* species whose genome sequences are now available (see Robertson 2005). Ammonia is another gas detected by mosquitoes in their perception of mammalian hosts (Geier et al. 1999; Meijerink et al. 2001; Smallegange et al. 2005; Qiu et al. 2006); however, *Drosophila* flies can perceive this gas (Yao et al. 2005), so AaGr2:AgGr23 seem unlikely to be involved in its perception. Nevertheless the extraordinary conservation of AaGr2:AgGr23 argues for an important and widespread role in insect chemoperception.

At the other extreme, these 2 mosquito lineages exhibit several lineage-specific Grs, and in particular, almost all the alternatively spliced isoforms appear to have evolved subsequent to the culicine–anopheline split ~150 Myr ago. Each of the Grs we have reported must ultimately be involved in the mosquito's ability to perceive chemicals in its environment, thus impacting the animal's behavior and ecology. Some of these behaviors are particularly relevant to vector and disease control. The obvious next step will be to define the ligands of these receptors (e.g., Chyb et al. 2003; Wanner et al. 2007). With this information in hand, we can begin to work on developing novel methods of disrupting mosquito ability to perceive its human host, improving our ability to control disease transmission.

Supplementary material

Supplementary material can be found at http://www.chemse.oxfordjournals.org.

Funding

National Institutes of Health (AI56081).

Acknowledgements

We thank Catherine Hill, David Hogenkamp, and Larry Zwiebel for assistance in early stages of this project.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, ZhaAng Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Bargmann CI. 2006. Comparative chemosensation from receptors to ecology. Nature. 444:295–301.

Benton R, Sachse S, Michnick SW, Vosshall LB. 2006. Atypical membrane topology and heteromeric function of *Drosophila* odorant receptors in vivo. PLoS Biol. 4:e20.

Bohbot J, Pitts RJ, Kwon H-W, Rützler M, Robertson HM, Zwiebel LJ. 2007. Molecular characterization of the *Aedes aegypti* odorant receptor gene family. Insect Mol Biol. 16:525–537.

Chyb S, Dahanukar A, Wickens A, Carlson JR. 2003. *Drosophila Gr5a* encodes a taste receptor tuned to trehalose. Proc Natl Acad Sci USA. 100:14526–14530.

Clyne PJ, Warr CG, Carlson JR. 2000. Candidate taste receptors in *Drosophila*. Science. 287:1830–1834.

Clyne PJ, Warr CG, Freeman MR, Lessing D, Kim J, Carlson JR. 1999. A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. Neuron. 22:327–338.

Dahanukar A, Foster K, van der Goes van Naters WM, Carlson JR. 2001. A *Gr* receptor is required for response to the sugar trehalose in taste neurons of *Drosophila*. Nat Neurosci. 4:1182–1186.

Davis EE. 1975. Temperature responses of antennal receptors of the mosquito, *Aedes aegypti*. J Comp Physiol. 96:223–236.

Dobritsa AA, van der Goes van Naters WG, Warr CG, Steinbrecht RA, Carlson JR. 2003. Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. Neuron. 37:827–841.

Dunipace L, Meister S, McNealy C, Amrein H. 2001. Spatially restricted expression of candidate taste receptors in the *Drosophila* gustatory system. Curr Biol. 11:822–835.

Fishilevich E, Vosshall LB. 2005. Genetic and functional subdivision of the *Drosophila* antennal lobe. Curr Biol. 15:1548–1553.

Geier M, Bosch OJ, Boeckh J. 1999. Ammonia as an attractive component of host odour for the yellow fever mosquito, *Aedes aegypti.* Chem Senses. 24:647–653.

Guo S, Kim J. 2007. Molecular evolution of *Drosophila* odorant receptor genes. Mol Biol Evol. 24:1198–1207.

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

Jones WD, Cayirlioglu P, Kadow IG, Vosshall LB. 2007. Two chemosensory receptors together mediate carbon dioxide detection in *Drosophila.* Nature. 445:86–90.

Krzywinski J, Grushko OG, Besansky NJ. 2006. Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. Mol Phylogenet Evol. 19:417–427.

Kwon JY, Dahanukar A, Weiss LA, Carlson JR. 2007. The molecular basis of CO2 reception in *Drosophila.* Proc Natl Acad Sci USA. 104:3574–3578.

Lee R. 1974. Structure and function of the fascicular stylets, and the labral and cibarial sense organs of male and female *Aedes aegypti* (L.) (Diptera, Culicidae). Quaest Entomol. 10:187–215.

Mackenzie JS, Gubler DJ, Petersen LR. 2004. Emerging flaviviruses: the spread and resurgence of Japanese encephalitis, West Nile and dengue viruses. Nat Med. 10:S98–S109.

Meijerink J, Braks MA, Van Loon JJ. 2001. Olfactory receptors on the antennae of the malaria mosquito *Anopheles gambiae* are sensitive to ammonia and other sweat-borne components. J Insect Physiol. 47:455–464.

Moon SJ, Köttgen M, Jiao Y, Xu H, Montell C. 2006. A taste receptor required for caffeine response in vivo. Curr Biol. 16:1812–1817.

Morlais I, Mori A, Schneider JR, Severson DW. 2003. A targeted approach to the identification of candidate genes determining susceptibility to *Plasmodium gallinaceum* in *Aedes aegypti.* Mol Genet Genomics. 269:753–764.

Nene V, Wortman JR, Lawson D, Haas B, Kodira C, Tu ZJ, Loftus B, Xi Z, Megy K, Grabherr M, et al. 2007. Genome sequence of *Aedes aegypti*, a major arbovirus vector. Science. 316:1718–1723.

Nozawa M, Nei M. 2007. Evolutionary dynamics of olfactory receptor genes in *Drosophila* species. Proc Natl Acad Sci USA. 104:7122–7127.

Pialoux G, Gaüzère BA, Jauréguiberry S, Strobel M. 2007. Chikungunya, an epidemic arbovirosis. Lancet Infect Dis. 7:319–327.

Qiu YT, van Loon JJ, Takken W, Meijerink J, Smid HM. 2006. Olfactory coding in antennal neurons of the malaria mosquito, *Anopheles gambiae.* Chem Senses. 31:845–863.

Robertson HM. 2005. Insect genomes. Am Entomol. 51:166–171.

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

Robertson HM, Warr CG, Carlson JR. 2003. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster.* Proc Natl Acad Sci USA. 100(Suppl 2):14537–14542.

Salama HS. 1966. The function of mosquito taste receptors. J Insect Physiol. 12:1051–1060.

Schmidt HA, Strimmer K, Vingron M, von Haeseler A. 2002. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics. 18:502–504.

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

Smallegange RC, Qiu YT, van Loon JJ, Takken W. 2005. Synergism between ammonia, lactic acid and carboxylic acids as kairomones in the host-seeking behaviour of the malaria mosquito *Anopheles gambiae* sensu stricto (Diptera: Culicidae). Chem Senses. 30:145–152.

Stewart MK, Clark NL, Merrihew G, Galloway EM, Thomas JH. 2005. High genetic diversity in the chemoreceptor superfamily of *Caenorhabditis elegans.* Genetics. 69:1985–1996.

Suh GSB, Wong AM, Hergarden AC, Wang JW, Simon AF, Benzer S, Axel R, Anderson DJ. 2004. A single population of olfactory neurons mediates an innate avoidance behaviour in *Drosophila.* Nature. 431: 854–859.

Swofford DL. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland (MA): Sinauer Associates.

Thathy V, Severson DW, Christensen BM. 1994. Reinterpretation of the genetics of susceptibility of *Aedes aegypti* to *Plasmodium gallinaceum.* J Parasitol. 80:705–712.

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

Tomori O. 2004. Yellow fever: the recurring plague. Crit Rev Clin Lab Sci. 41:391–427.

Wanner KW, Nichols AS, Walden KKO, Brockmann A, Luetje CW, Robertson HM. 2007. A honeybee odorant receptor for the queen substance 9-oxo-2-decenoic acid. Proc Natl Acad Sci USA. 104:14383–14388.

Werner-Reiss U, Galun, R, Crnjar R, Liscia A. 1999. Sensitivity of the mosquito *Aedes aegypti* (Culicidae) labral apical chemoreceptors to phagostimulants. J Insect Physiol. 45:629–636.

Yao CA, Ignell R, Carlson JR. 2005. Chemosensory coding by neurons in the coeloconic sensilla of the *Drosophila* antenna. J Neurosci. 25:8359–8367.

Accepted August 31, 2007