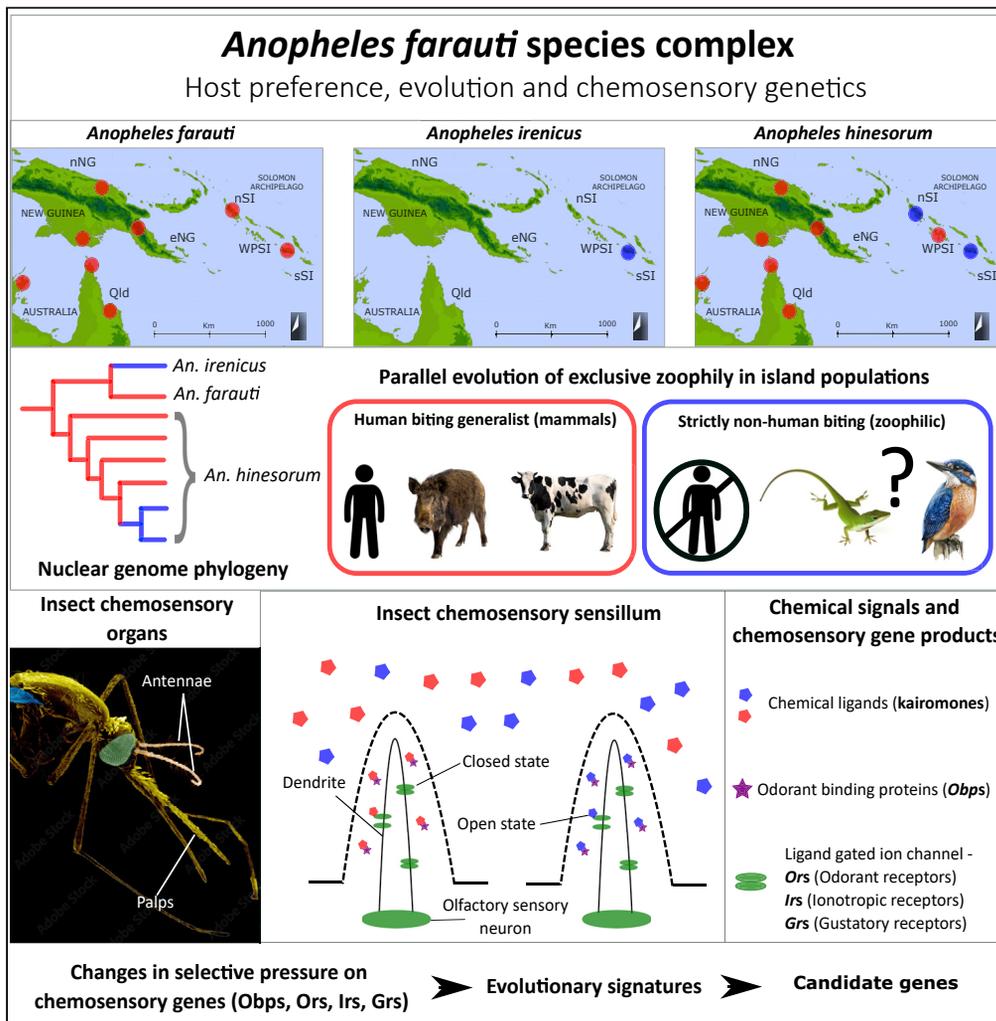


Article

Comparisons of chemosensory gene repertoires in human and non-human feeding *Anopheles* mosquitoes link olfactory genes to anthropophily



Luke Ambrose, Iva Popovic, James Hereward, Daniel Ortiz-Barrientos, Nigel W. Beebe

lukeambrose3@gmail.com

Highlights

Genomes of *Anopheles* mosquitoes with differing host preferences were sequenced

Evolutionary comparisons were performed on >200 insect chemosensory genes

These comparisons revealed candidate genes involved in human feeding

Two of the main candidates identified were co-receptor *Ir8a* and *Or75*

Ambrose et al., iScience 25, 104521
July 15, 2022 © 2022 The Author(s).
<https://doi.org/10.1016/j.isci.2022.104521>



Article

Comparisons of chemosensory gene repertoires in human and non-human feeding *Anopheles* mosquitoes link olfactory genes to anthropophilyLuke Ambrose,^{1,3,*} Iva Popovic,¹ James Hereward,¹ Daniel Ortiz-Barrientos,¹ and Nigel W. Beebe^{1,2}

SUMMARY

We investigate the genetic basis of anthropophily (human host use) in a non-model mosquito species group, the *Anopheles farauti* complex from the southwest Pacific. This complex has experienced multiple transitions from anthropophily to zoophily, contrasting with well-studied systems (the global species *Aedes aegypti* and the African *Anopheles gambiae* complex) that have evolved to be specialist anthropophiles. By performing tests of selection and assessing evolutionary patterns for >200 olfactory genes from nine genomes, we identify several candidate genes associated with differences in anthropophily in this complex. Based on evolutionary patterns (phylogenetic relationships, fixed amino acid differences, and structural differences) as well as results from selection analyses, we identify numerous genes that are likely to play an important role in mosquitoes' ability to detect humans as hosts. Our findings contribute to the understanding of the evolution of insect olfactory gene families and mosquito host preference as well as having potential applied outcomes.

INTRODUCTION

Many insect behaviors are informed by chemosensory signals, including the host preferences of blood-feeding insects (Takken, 1991; Zwiebel and Takken, 2004; Haverkamp et al., 2018). Some of these behaviors have significant economic and medical implications. In the case of mosquitoes, the most important arthropod vectors of human-disease-causing pathogens (Robert and Debboun, 2020), understanding how olfaction drives host preference will provide additional tools for vector control and disease prevention.

Mosquito host preference has been shown to be driven largely by olfaction (Takken, 1991) and to have a genetic basis (Main et al., 2016). Several major gene families are involved in this behavior including receptor proteins—olfactory receptors (Ors) (Missbach et al., 2014; Brand et al., 2018; Yan et al., 2020), gustatory receptors (Grs) (Scott, 2018), ionotropic receptors (Irs) (Benton et al., 2009)—and soluble globular proteins (odorant-binding proteins (Obps) (Pelosi et al., 2005). Owing to their central role in olfaction, genes in these families are often involved in the evolution of changes in host preference, including changes in the range of hosts on which insects feed (Matsuo et al., 2007). This host range is highly variable across different mosquito species and populations. Most mosquitoes species are generalists whose host use is largely determined by availability (Takken and Verhulst, 2013), but some species of mosquitoes are highly specialized and can only feed on hosts of one or a few closely related species (Borkent and Belton, 2006; Bartlett-Healy et al., 2008; Reeves et al., 2018). Because of their strong preference for feeding on humans, *Anopheles gambiae* and *Aedes aegypti* are two of the most efficient vectors of human diseases (Besansky et al., 2004; Ritchie, 2014). As such, they have been the focus of previous research investigating the genetic basis of human host preference in mosquitoes. Both *A. gambiae* and *Ae. aegypti* have recently evolved from host generalists into human feeding specialists as a result of human-induced changes in the environment such as land clearing, agriculture, and urbanization (Costantini et al., 1999; Powell and Tabachnick, 2013; Rose et al., 2020). This is in contrast to the *A. farauti* complex, where in the same isolated geographical area, species and populations have evolved repeatedly and independently from anthropophilic generalists into zoophilic specialists (Ambrose et al., 2012).

Research in the two major study systems mentioned above has been focused on identifying the molecular and genetic basis of differences in host preference between closely related zoophilic and anthropophilic mosquito species (Rinker et al., 2013; McBride et al., 2014; Athrey et al., 2017). This research has been

¹School of Biological Sciences, University of Queensland, Brisbane, Australia

²CSIRO, Brisbane, QLD, Australia

³Lead contact

*Correspondence:

lukeambrose3@gmail.com

<https://doi.org/10.1016/j.isci.2022.104521>



successful in identifying human kairomones that are attractive to mosquitoes (Geier et al., 1999; Meijerink et al., 2000; Braks et al., 2001; Costantini et al., 2001; Dekker et al., 2002; Dekker et al., 2005; Leal, 2010; Lacey et al., 2014; Frei et al., 2017), as well as gene products expressed in mosquito olfactory systems that are involved in detecting these kairomones (Carey et al., 2010; Wang et al., 2010). Disruption of these molecular pathways and mechanisms would disable mosquitoes' ability to sense humans as hosts, potentially providing a control target to reduce the spread of mosquito-borne diseases.

We introduce a system that is well-suited for studying the genetic basis of anthropophily in mosquitoes: the *Anopheles farauti* complex. Species of this complex are endemic to the southwest Pacific (Beebe et al., 2015) and have undergone repeated evolution of differences in host preference (Ambrose et al., 2012). Although most members of the group are anthropophilic generalists, there are distinct species and populations in the group that have evolved to be exclusively zoophilic (Ambrose et al., 2012). The strictly zoophilic behavior in *A. hinesorum* and *A. irenicus* from the Solomon Archipelago, is evidenced by numerous human landing catches in which these species have not been collected, despite being performed near highly productive larval habitats (Foley et al., 1994; Beebe et al., 2000; Cooper and Frances, 2002). These experiments were performed in both Bougainville (Foley et al., 1994; Cooper and Frances, 2002) and Guadalcanal (Beebe et al., 2000, Ambrose unpublished data). Phylogenetic and population genetic relationships suggest that zoophily has evolved more than once in the *A. farauti* complex in the Solomon Archipelago. This shift in behavior has evolved in a cryptic species of the *A. farauti* complex—*An. irenicus* found only on Guadalcanal Island, and independently at least once in *A. hinesorum* for which there are two mitochondrially distinct zoophilic populations found in different parts of the Solomon Archipelago (Ambrose et al., 2012). The independent evolution of zoophily in this system provides natural evolutionary replication of the loss of the ability or preference to use humans as hosts. However it should be noted that these species have not been successfully colonized, meaning that controlled experiments on their host preferences have not been performed.

Although most *A. hinesorum* populations in the Solomon Archipelago are strictly zoophilic, recent human landing collections in the Western Islands of the Solomon Archipelago (nearest New Guinea) collected *A. hinesorum* seeking humans (Burkot et al., 2018). Some individuals collected in this population were found to carry a mtDNA COI genotype not previously found in the Solomon Archipelago. This genotype is also present in human-feeding mainland New Guinean populations, suggesting a recent introduction of females from New Guinea to the Solomon Archipelago. Microsatellite analysis suggested that this human feeding island population has a nuclear genomic background very similar to other Solomon Archipelago populations of the species (Ambrose et al., 2021). Thus, *A. hinesorum* populations in the Solomon Archipelago allow intraspecific comparison of populations with very similar nuclear genomes but differences in behavior. Altogether, the relationships observed between species and populations of the *A. farauti* complex with differences in host preference make it a particularly useful system for identifying the genetic basis of human host detection by mosquitoes. By comparing the genomes and coding sequences of olfactory gene repertoires in the *A. farauti* complex, we hope to gain an understanding of the genetic basis of this difference in behavior.

In this study, we investigate how selection has operated on genes belonging to the major olfactory gene families—*Ors*, *Grs*, *Irs*, and *Obps*—in the *A. farauti* complex, during host shifts from mammalian generalism to exclusive zoophily. The evolution of this phenotype constitutes a loss of the ability to detect humans (and possibly mammals more broadly) as hosts. Previous literature shows that the evolution of loss of function phenotypes (including behavioral phenotypes) often involves either relaxed selection on genes associated with that function (Lahti et al., 2009; Wertheim et al., 2015; Calderoni et al., 2016; Lu et al., 2019) and/or adaptive (positive) selection on similar genes (McBride, 2007; McBride and Arguello, 2007; Harpur et al., 2014). We, therefore, hypothesize that a subset of olfactory genes may have experienced relaxed or positive selection during the evolution of zoophily in the *A. farauti* complex. Furthermore, the same or different genes may be involved in the independent evolution of zoophily in the species complex. Our aims are to assess evolutionary patterns in olfactory genes and gene families and to identify candidate genes involved in differences in human host preference in the species complex. We sequence and assemble 11 genomes from individuals in the *A. farauti* complex with different host preferences, manually extract >200 olfactory genes from nine genomes and perform comparative evolutionary tests including phylogeny-informed, hypothesis-based tests of selection. Our findings corroborate previous studies on the genetic underpinnings of host preference in mosquitoes and provide insight into previously undiscovered genes that may be involved in the ability of *Anopheles* mosquitoes to detect humans as hosts.

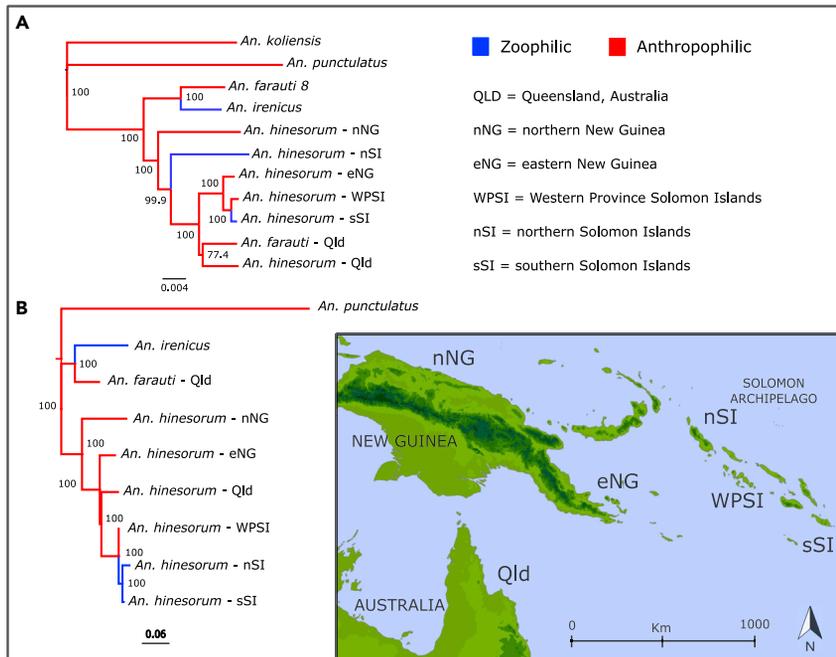


Figure 1. Mitogenome and consensus nuclear genome phylogenies

(A) Neighbor-joining phylogeny for samples used in this study based on whole mitogenome data. Support values are based on 1000 bootstrap replicates.

(B) Consensus neighbor-joining phylogeny for samples used in this study based on 164 041 SNPs from whole-genome sequence data. Support values are based on 1000 bootstrap replicates. Host preference is indicated by branch color as indicated in the key and the top-right panel is a map of the region, providing geographical context.

RESULTS

Whole-genome phylogenies and species relationships

Initially, we assessed phylogenetic relationships in the *A. farauti* species complex based on whole-genome nuclear variants and whole mitogenomes (Figure 1). We sequenced the whole genomes of 11 individuals from six species at high coverage (40-80x) (Table 1) and assembled these by mapping short reads to the *A. farauti* reference genome available on VectorBase (Neafsey et al., 2015; Vectorbase: Bioinformatic resources for invertebrate vectors of human pathogens, 2021). *Anopheles farauti* and *A. irenicus* form a monophyletic pair for the neighbor-joining nuclear SNP dendrogram (Figure 1B), supporting their previously asserted sister species relationship (Ambrose et al., 2012). However, *A. farauti* is most closely related to *A. hinesorum* based on the mitogenome phylogeny (Figure 1A). This result was expected given that *A. farauti* populations throughout northern Australia and southern New Guinea carry mitochondrial DNA that has introgressed from *A. hinesorum* (Ambrose et al., 2012), and that the *A. farauti* sample sequenced originated from QLD, Australia (Figure 1). The mitogenome phylogeny also verifies that the zoophilic Solomon Archipelago *A. hinesorum* individuals sequenced represents the two divergent northern and southern lineages previously identified (Ambrose et al., 2012). The close relationship between the nuclear genomes of *A. hinesorum* populations from the Solomon Archipelago, previously observed in microsatellite data (Ambrose et al., 2021), is further supported by the nuclear phylogeny as well as phylogenetic relationships of most olfactory genes. In the nuclear phylogeny, *A. hinesorum* lineages form a well-supported monophyletic clade, with the northern New Guinea individual being the most divergent from the rest of the species.

Sequence divergence and genomic location of olfactory genes in the *A. farauti* complex

Using gene prediction methods (tBLASTn and Gene-Wise) (Birney et al., 2004; Gerts et al., 2006), we isolated and analyzed high-quality sequence data from 54 *Or* genes, 68 *Obp* genes, 37 *Ir* genes, and 50 *Gr* genes, from nine of the genomes assembled. This represents most of the olfactory gene repertoire of *A. gambiae*, as classified by Rinker et al., 2013 (Rinker et al., 2013); however, several olfactory genes present in *A. gambiae* appear to have been lost in *A. farauti*. For some genes, missing orthologs resulted in alignments with a reduced set of individuals (Tables S1–S4).

Table 1. Sample information on the individuals sequenced in this study

Sample ID	Species	Location	Host preference	Collection	Coverage	Lib type
QLD_far	<i>A. farauti</i> s.s.	Queensland (Aus)	Opportunist (A)	Adult (HLC)	60	250bp PE
sSI_iren	<i>A. irenicus</i>	Guadalcanal (SI)	Animal (Z)	larval	64	250bp PE
sSI_hin	<i>A. hinesorum</i>	Guadalcanal (SI)	Animal (Z)	larval	62	250bp PE
nSI_hin	<i>A. hinesorum</i>	Bougainville (SI)	Animal (Z)	larval	49	250bp PE
WPSI_hin	<i>A. hinesorum</i>	Western Province (SI)	Opportunist (A)	Adult (HLC)	30	100bp PE
QLD_hin	<i>A. hinesorum</i>	Queensland (Aus)	Opportunist (A)	Adult (CDC)	60	250bp PE
nNG_hin	<i>A. hinesorum</i>	Northern New Guinea	Opportunist (A)	Adult (HLC)	39	100bp PE
eNG_hin	<i>A. hinesorum</i>	Eastern New Guinea	Opportunist (A)	Adult (HLC)	41	100bp PE
NG_Pun	<i>A. punctulatus</i>	New Guinea	Opportunist (A)	Adult (HLC)	87	250bp PE

Sample ID = name given to the sample sequenced; **Species** = species that the sample belongs to; **Location** = geographic location that the sample was collected; **Host preference** = host preference of the population from which the sample was taken: **Opportunist (A)** = will readily feed on humans and other mammals), **Animal (Z)** = only feeds on animals other than humans (unknown hosts); **Collection** = indicates whether the sample was collected as an adult or larva and whether adults were collected in human landing catches (HLCs) or with CDC traps; **Coverage** = estimated average genome coverage of mapped reads; **Lib type** = Type of paired-end library used to generate sequence data (either 250bp paired-end reads or 100bp paired-end reads).

When looking at overall gene family evolution, we found no significant differences in gene conservation among the four gene families. Gene conservation was measured as the percentage of identical sites in both nucleotide (ANOVA, $df = 3, 209, F = 0.66, p = 0.58$) and amino acid alignments (ANOVA, $df = 3, 209, F = 0.4889, p = 0.69$). However, we did observe greater variance in *Obps* than in other gene families, including some extremely conserved genes (78.5–98.7 percent nucleotide identity; 73.8–100 percent amino acid identity). As a point of reference, the highly conserved insect olfactory co-receptor gene (*Orco*) has 98.1% identical sites for the nucleotide sequence and 99.4% identical sites for the amino acid sequence. The mean values for percentage nucleotide identity across all gene families range from 92.66 to 93.44 and median values range from 92.5 to 93.7. The mean values for percentage amino acid identity range from 92.21 to 93.18 and median values range from 92.2 to 94. See [Figure S1](#) for a visual representation of these results as well as summary statistics for each gene family and [Tables S1–S4](#) for sequence identity data for each individual gene in each family.

We assessed chromosome level synteny between *A. gambiae* and *A. farauti* by examining whether genes found on the largest *A. farauti* scaffolds are predominantly found on the same *A. gambiae* chromosome ([Figure S2](#)). Both *A. gambiae* and *A. farauti* have three sets of chromosomes (one set of sex chromosomes—X/Y, and two sets of autosomes—chromosomes two and 3). We found that chromosomal synteny between *A. gambiae* and *A. farauti* is largely conserved and found in five scaffolds from the *A. farauti* reference genome that may be located on the X chromosome. We found nine olfactory genes on *A. farauti* scaffold KI915047 (five *Obps*, two *Ors*, and two *Irs*), all of which are found on the *A. gambiae* X chromosome, apart from *Obp30*, which currently has an unknown location in *A. gambiae*. Other *A. farauti* scaffolds which are likely to be located on the X chromosome include KI915047, KI915065, KI915074, and KI915078.

Scaffolds that are likely to be located on the right arm of the *A. farauti* autosomal chromosome two (2R) include scaffolds KI915040 (30.17 Mbp), KI915046 (7.418 Mbp), KI915048 (6.651), and KI915049 (6.084 Mbp). We found 37 olfactory genes on KI915040 (13 *Ors*, nine *Obps*, nine *Irs*, and six *Grs*), with 36 of these located on chromosome 2R in *A. gambiae* and two (*Obps* 60 and 79) of unknown genomic location. For genes where genomic location is known in *A. gambiae*, all olfactory gene orthologs on *A. farauti* scaffold KI915046 (7.418 Mbp)—one *Or*, two *Obps*, two *Grs*, one *Ir*—are found on *A. gambiae* 2R. For the 19 olfactory genes found on scaffold KI915048, 18 are found on 2R with one of the unknown location in *A. gambiae*. All four olfactory genes found on KI915049 are also found on *A. gambiae* chromosome 2R.

Scaffolds associated with the left arm of chromosome two (2L) include KI915041 (22.738 Mbp) and KI915044 (12.895 Mbp). A total of 24 olfactory genes were identified on scaffold KI915041—eight *Obps*, four *Ors*, six *Grs* and seven *Irs*—21 of these are found on chromosome 2L in *A. gambiae*, with *Obp67* found on 3L and *Obp78*, *Ir137* and *Ir138* having unknown genomic locations. On KI915044 we found 20 olfactory genes (12 *Obps*, three *Ors*, two *Grs*, and one *Ir*) with 17 being found on chromosome 2L in *A. gambiae* and unknown locations for *Obp77*, 80, and 81.

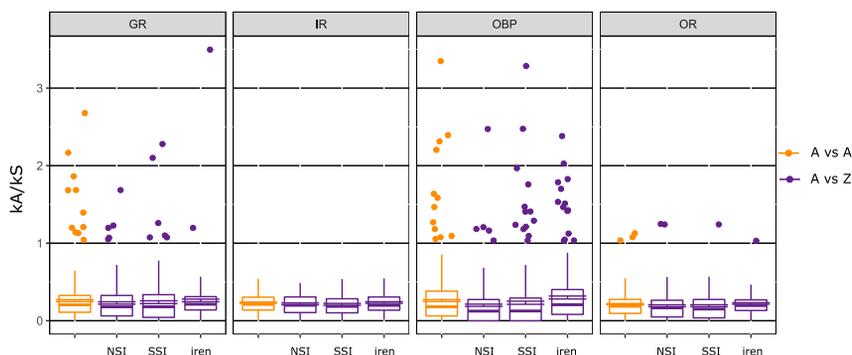


Figure 2. kA/kS by gene family, comparison class, and zoophilic lineage

Boxplots are shown for each gene family by comparison class and zoophilic lineage, with standard errors around means presented; **Z** = zoophilic, **A** = anthropophilic. Individual comparisons with $kA/kS > 1$ are shown on plots. Summary statistics for each gene family/comparison class are as follows: **Gr**s (mean, median) A vs A (0.259, 0.205), A vs NSI *hinesorum* (0.229, 0.179), A vs SSI *hinesorum* (0.239, 0.180), A vs *irenicus* (0.261, 0.213); **Irs** (mean, median) A vs A (0.233, 0.211), A vs NSI *hinesorum* (0.217, 0.199), A vs SSI *hinesorum* (0.210, 0.185), A vs *irenicus* (0.235, 0.198); **Obps** (mean, median) A vs A (0.260, 0.180), A vs NSI *hinesorum* (0.199, 0.124), A vs SSI *hinesorum* (0.232, 0.126), A vs *irenicus* (0.299, 0.206); **Ors** (mean, median) A vs A (0.214, 0.186), A vs NSI *hinesorum* (0.193, 0.166), A vs SSI *hinesorum* (0.193, 0.150), A vs *irenicus* (0.223, 0.196).

Scaffolds associated with the right arm of chromosome three (3R) include KI915042 (16.089 Mbp) and KI915043 (15.719 Mbp). All 12 genes found on contig KI915042—one *Obp*, two *Ors*, three *Irs*, and six *Gr*s—are found on chromosome 3R in *A. gambiae*. Of the 24 olfactory genes (seven *Obps*, 15 *Ors*, and one *Gr*) found on *A. farauti* scaffold KI915043, 18 are located on *A. gambiae* chromosome 3R. However, *Obps* 70 and 71 are found on chromosome 2L in *A. gambiae*, *Obp69* and *Ir140.1* are found on 2R and *Obps* 74 and 76 are of unknown location in *A. gambiae*.

Scaffolds likely to be located on the left arm of chromosome three (3L) include KI915045 (12.084 Mbp) and KI915047 (6.913 Mbp). Fifteen olfactory genes were found on scaffold KI915045 (12.084 Mbp)—one *Gr*, two *Irs*, one *Or*, 11 *Obps*—of these, 13 are found on chromosome 3L, with *Obp58* and *59* of unknown location in *A. gambiae*. Nine olfactory genes were found on KI915047 (6.913 Mbp)—five *Obps* (one unknown in *A. gambiae*), two *Ors*, and two *Irs*.

Patterns of olfactory gene family evolution

To further explore patterns of selection within and between olfactory gene families we analyzed distributions of kA/kS values for entire gene families. Owing to non-normal distributions of residuals, we used Kruskal Wallis tests to assess whether differences in distributions are significant. All but two pairwise tests between gene families were significant—*Ors/Gr*s (Kruskal-Wallis chi-squared = 11.431, $df = 1$, $p = 7.22 \times 10^{-4}$), *Ors/Irs* (Kruskal-Wallis chi-squared = 17.933, $df = 1$, $p = 2.29 \times 10^{-5}$), *Obps/Gr*s (Kruskal-Wallis chi-squared = 17.968, $df = 1$, $p = 2.25 \times 10^{-5}$), *Obps/Irs* (Kruskal-Wallis chi-squared = 22.332, $df = 1$, $p = 2.29 \times 10^{-6}$). The two non-significant comparisons were between *Ors* and *Obps* (Kruskal-Wallis chi-squared = 1.66, $df = 1$, $p = 0.19$) and *Gr*s and *Irs* (Kruskal-Wallis chi-squared = 0.439, $df = 1$, $p = 0.51$). Although the overall mean kA/kS values are similar between gene families (*Ors* = 0.216, *Gr*s = 0.252, *Irs* = 0.221, *Obps* = 0.245), the median of *Obps* is much lower than for the three other gene families (*Obps* = 0.158, *Ors* = 0.182, *Gr*s = 0.194, *Irs* = 0.199), despite *Obps* having the second highest mean kA/kS .

To compare overall rates of positive selection operating on gene families between phenotype comparison classes, we also compared kA/kS by gene family and phenotype comparison, with anthropophilic versus anthropophilic comparisons providing a baseline for comparison to anthropophilic versus zoophilic comparisons (Figure 2). We separated anthropophilic versus zoophilic comparisons into the three separate lineages/individuals sequenced to assess species/lineage-specific patterns, though it should be noted based on nuclear phylogenetic patterns observed, that zoophagy may not have evolved independently in the two zoophilic *A. hinesorum* lineages.

We found that mean pairwise kA/kS ratios for gene families overall were consistently highest in comparisons between *A. irenicus* and anthropophilic individuals (Figure 2), as were median kA/kS ratios in all

gene families apart from *Irs* where anthropophilic vs anthropophilic comparisons showed the highest median value. The distributions of *kA/kS* values in comparisons between *A. irenicus* and anthropophilic individuals were also found to be significantly different from comparisons between zoophilic *A. hinesorum* from the both the northern and southern Solomon Archipelago and anthropophilic individuals for *Obps* (Wilcoxon test: $p = 5.3 \times 10^{-7}$, 3.2×10^{-6}), *Ors* (Wilcoxon test: 5.9×10^{-4} , 2.2×10^{-4}) and *Grs* (Wilcoxon test: $p = 0.0018$, 0.0031). The only significant difference in *kA/kS* distributions for *Irs* was found between the anthropophilic vs anthropophilic comparison and the southern Solomon hinesorum vs anthropophilic comparison (Wilcoxon test: $p = 0.017$). Although mean *kA/kS* values were also consistently higher in comparisons between *A. irenicus* and anthropophilic individuals than in anthropophilic vs anthropophilic comparisons, distributions of *kA/kS* were not found to be significantly different for any gene family between these comparison classes; however, some were close to significant, assuming a type I error rate of 5% (*OBPs*, $p = 0.054$). Average *kA/kS* values for all gene families were higher for anthropophilic vs anthropophilic comparisons compared with comparisons between anthropophilic individuals and both zoophilic *A. hinesorum* lineages (Figure 2). Distributions for these comparisons were all significantly different except for the anthropophilic vs anthropophilic comparison against the anthropophilic vs southern Solomon *A. hinesorum* lineage for *Irs*.

We found that *Obps* proportionately have the most *kA/kS* values above one in all comparison classes, including in zoophilic vs anthropophilic comparisons. We found that this excess of *kA/kS* values over one in zoophilic/anthropophilic (32/976) comparisons versus anthropophilic/anthropophilic comparisons (12/654), was close to significant when tested with a two-sample test for equality of proportions (Chi-squared = 2.58, $df = 1$, $p = 0.054$). To visually summarize this pattern, we plotted the overall mean and median *kA/kS*, as well as outlier *kA/kS* values for each gene family by comparison type (Figure 2). See Figure 2 for summary statistics for each comparison class by the gene family.

Candidate genes involved in human host detection – evolutionary patterns predicted for zoophilic host shifts

We applied four criteria indicating a potential association of a gene with the observed host shift to zoophily: (i) fixed amino acid differences between anthropophilic and zoophilic species and populations; (ii) specific phylogenetic relationships between species and populations (Figure 3); (iii) higher *kA/kS* ratios in comparisons of different (anthropophilic vs zoophilic) versus same (anthropophilic vs anthropophilic) behavioral classes; and (iv) differences in gene structure (putative insertions or deletions in coding sequences) (see Table 2).

Only one gene, *Ir8a*, contains a fixed amino acid substitution unique to all zoophilic individuals. This substitution (Glutamine to Histidine) has resulted from a different nucleotide substitution in *A. irenicus* and zoophilic *A. hinesorum*. Within *Ir8a* there are three additional fixed amino acid differences at sites 47 (Aspartic acid to Glutamic acid), 75 (Leucine to Valine), and 234 (Threonine to Alanine) between both zoophilic *A. hinesorum* individuals sequenced and the other the species/populations (Figure 4). An additional 12 genes contained at least one fixed amino acid substitution unique to zoophilic *A. hinesorum* lineages (one *Obp* – *Obp* 22; three *Irs*; *Ir7i*, *7u*, and *8a*; four *Grs* – *Gr39*, *40*, *57*, and *60*; and four *Ors* – *Or11*, *24*, *66*, and *75*). For further details of amino acid substitutions in these genes see Table 2.

Based on prior knowledge of genetic relationships between populations in the *A. farauti* complex, we expect genes involved in differences in host preference to show specific phylogenetic patterns (Figure 3). The strongest phylogenetic signal suggesting that a gene is correlated with host preference would be anthropophilic and zoophilic lineages (populations and species) forming monophyletic groups. We found no genes showing this relationship. Another pattern suggestive of a gene is a strong candidate is genes for which anthropophilic and zoophilic *A. hinesorum* forming monophyletic clades (Figure 3B). We found 12 genes showing this phylogenetic pattern with strong bootstrap support (>80) and a further 14 genes showing this pattern with weaker support, including *Ir8a*. Eleven genes showed evidence of insertions or deletions in the coding sequences of at least one zoophilic lineage. Furthermore, most genes showed the Solomon Archipelago samples to be monophyletic (34/38 *Irs*, 46/50 *Grs*, 48/57 *Ors*, 44/69 *Obps*).

For 17 genes we found overall higher *kA/kS* ratios in comparisons of anthropophilic vs zoophilic, than anthropophilic vs anthropophilic comparisons based on non-overlapping SE bars (Figure S3). These include two *Grs*, three *Irs*, six *Obps*, and six *Ors* (Figure S3). Three of these genes (*Ir8a*, *Gr57*, and *Or75*) also contained a fixed amino acid substitution between anthropophilic and zoophilic populations. As there

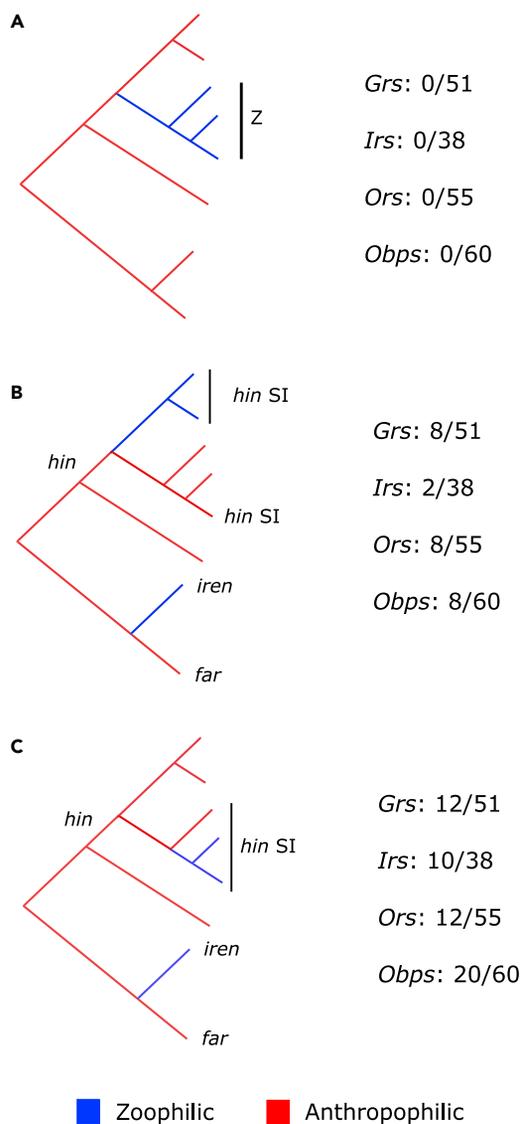


Figure 3. Expected phylogenetic relationship in candidate genes

Blue branches represent zoophilic lineages while red branches represent anthropophilic lineages. *hin* = *A. hinesorum*; *hin SI* = *A. hinesorum* from the Solomon Archipelago; *far* = *A. farauti*; *iren* = *A. irenicus*. *Grs* = Gustatory receptors; *Irs* = Ionotropic receptors; *Ors* = Olfactory receptors; *Obps* = Olfactory binding proteins. Numbers to the right of the figure show the proportion of genes in which each relationship was observed.

(A) The strongest hypothetical phylogenetic signal of a gene being involved in differences in host preference. Observing this relationship would suggest that a gene has introgressed from one zoophilic lineage to the others.

(B) The strongest phylogenetic relationship observed in this study suggestive of a gene being a potential candidate. For genes showing this relationship, zoophilic (Z) *A. hinesorum* from the Solomon Archipelago form an exclusive clade and anthropophilic (A) *A. hinesorum* fall within a clade containing other anthropophilic *A. hinesorum* (from Queensland and/or New Guinea).

(C) A weaker phylogenetic relationship potentially suggestive of a gene being a candidate. For genes showing this relationship, all *A. hinesorum* samples from the Solomon Islands form a monophyletic clade with the two zoophilic individuals being most closely related.

may also be lineage and species-specific differences in selection and substitution rate on olfactory genes, we also present pairwise kA/kS ratios between the three zoophilic lineages and anthropophilic individuals for each gene separately (Figures S4–S7). This revealed only two genes (*Ir8a* and *Obp71*) showing consistently higher kA/kS ratios in different vs same phenotype comparisons (Figures S6 and S7). Although many genes showed higher average kA/kS ratios uniquely in comparisons between *A. irenicus* and anthropophilic individuals relative to anthropophilic vs anthropophilic comparisons ($n = 28$), fewer genes showed a similar lineage-specific signal for comparisons involving zoophilic *A. hinesorum* lineages (*nSI hinesorum*: $n = 4$; *sSI hinesorum*: $n = 5$; Figures S4–S7 and Table 2). An additional five genes showed higher kA/kS ratios only in comparisons involving both zoophilic *A. hinesorum* lineages; however, this may be owing to gene flow or shared inheritance of these loci rather than selection acting independently in these lineages, especially as four of these genes show phylogenetic signals, suggesting that zoophilic *A. hinesorum* lineages are most closely related to each other. Other genes show higher kA/kS ratios in *A. irenicus* and only one of the zoophilic *A. hinesorum* lineages. Additionally, we observed conflicting signals of selection for many genes with regards to kA/kS ratios for anthropophilic vs zoophilic comparisons relative to anthropophilic vs anthropophilic comparisons (Table 2). Altogether, these results suggest that some genes are involved in detecting olfactory signals not associated with the detection of hosts and that different combinations of olfactory genes may be involved in host detection in the different zoophilic lineages.

Table 2. Evolutionary patterns for candidate genes

GeneID	Fixed aa	Sub/site	Phyl	> <i>kA/kS</i>	< <i>kA/kS</i>	Exons
<i>Obp3</i>	-	-	B	> iren, sSI_hin	-	-
<i>Obp4</i>	-	-	C	> sSI_hin	< nSI_hin	-
<i>Obp5</i>	-	-	-	-	-	Y Exons – Z_hin
<i>Obp7</i>	-	-	C	>> Z_hin	-	-
<i>Obp8</i>	-	-	-	> nSI_hin	< sSI_hin	-
<i>Obp10</i>	-	-	C	>> iren	<< nSI_hin	-
<i>Obp13</i>	-	-	-	>> iren, nSI_hin	-	-
<i>Obp14</i>	-	-	C	-	-	-
<i>Obp15</i>	-	-	-	> iren	< nSI_hin	-
<i>Obp20</i>	-	-	C	> iren	-	-
<i>Obp22</i>	Z_hin	T to I, 139	B	>> iren	< Z_hin	Y Exons sSI_iren
<i>Obp24</i>	-	-	-	> iren	<< Z_hin	-
<i>Obp25</i>	-	-	C	> nSI_hin	-	-
<i>Obp27</i>	-	-	C	-	< Z_hin	-
<i>Obp28</i>	-	-	C	-	-	-
<i>Obp29</i>	-	-	C	-	-	-
<i>Obp30</i>	-	-	C	-	-	-
<i>Obp31</i>	-	-	B	> iren	-	-
<i>Obp34</i>	-	-	-	>> iren, > nSI_hin	-	-
<i>Obp35</i>	-	-	-	>> iren	<< Z_hin	-
<i>Obp40</i>	-	-	-	>> iren	-	-
<i>Obp41</i>	-	-	C	>> iren	-	-
<i>Obp43</i>	-	-	-	> sSI_hin	-	-
<i>Obp44</i>	-	-	C	-	<< iren	-
<i>Obp45</i>	-	-	B	-	-	Y Exons
<i>Obp46</i>	-	-	C	> iren	< Z_hin	-
<i>Obp47</i>	-	-	B	-	-	-
<i>Obp48</i>	-	-	C	-	<< iren, < Z_hin	Y Exons – nSI_hin
<i>Obp52</i>	-	-	-	>> iren, nSI_hin	< sSI_hin	-
<i>Obp53</i>	-	-	-	>> iren, nSI_hin	< sSI_hin	-
<i>Obp54</i>	-	-	-	> iren	-	-
<i>Obp55</i>	-	-	-	> iren	< nSI_hin	-
<i>Obp56</i>	-	-	-	> sSI_hin,	< iren	-
<i>Obp58</i>	-	-	C	-	< iren	-
<i>Obp59</i>	-	-	C	-	-	-
<i>Obp60</i>	-	-	B	> sSI_hin	-	Y Exons
<i>Obp63</i>	-	-	C	-	< A v Z	-
<i>Obp64</i>	-	-	-	>> iren	-	-
<i>Obp66</i>	-	-	C	-	<< iren	-
<i>Obp67</i>	-	-	C	-	< iren, nSI_hin	-
<i>Obp70</i>	-	-	B	-	< nSI_hin	-
<i>Obp71</i>	-	-	B	>> A v Z	-	-
<i>Obp74</i>	-	-	-	-	-	Exons sSI_iren
<i>Obp77</i>	-	-	-	>> sSI_hin	-	-
<i>Obp80</i>	-	-	C	-	<< Z_hin, < iren	-

(Continued on next page)

Table 2. Continued

GeneID	Fixed aa	Sub/site	Phyl	> kA/kS	< kA/kS	Exons
Or1	-	-	C	-	-	-
Or2	-	-	-	>> iren, > sSI_hin	-	-
Or5	-	-	-	> iren	< sSI_hin	-
Or6	-	-	C	-	<< sSI_hin, < iren, nSI_hin	Exons – nSI_hin
Or11	Z_hin	A to V,276	C	-	-	-
Or13	-	-	-	> iren	-	-
Or14	-	-	-	> iren	-	-
Or16	-	-	-	>> iren	-	-
Or23	-	-	B	-	-	-
Or24	Z_hin	I to M, 342	C	-	<< iren	-
Or25	-	-	B	-	< sSI_hin	-
Or26	-	-	-	>> iren	-	-
Or28	-	-	C	> iren, nSI_hin	-	-
Or31	-	-	C	>> iren	-	-
Or33	-	-	C	-	< sSI_hin	-
Or34	-	-	-	> iren	-	-
Or37	-	-	-	> Z_hin	-	-
Or38	-	-	C	-	< Z_hin	Exons – sSI_iren
Or40	-	-	-	>> iren, > sSI_hin	<< nSI_hin	-
Or41	-	-	C	> iren	-	-
Or42	-	-	-	>> sSI_hin	< iren	-
Or43	-	-	B	-	< A v Z	-
Or45	-	-	B	-	-	-
Or46	-	-	-	> sSI_hin	<< iren, nSI_hin	-
Or52	-	-	-	> iren	-	-
Or53	-	-	-	>> nSI_hin	-	Exons – nSI_hin
Or55	-	-	-	> iren	-	-
Or58	-	-	B	-	< nSI_hin	-
Or64	-	-	C	-	< iren	-
Or66	Z_hin	A to V, 236	C	-	-	-
Or75	Z_hin	E to K, 138	B	>> iren, > sSI_hin	-	-
Or77	-	-	B	-	-	-
Or80	-	-	C	-	-	-
Gr2	-	-	-	> iren	-	-
Gr4	-	-	C	>> iren	-	-
Gr5	-	-	B	-	-	-
Gr8	-	-	-	> iren	<< nSI_hin	-
Gr9	-	-	-	>> iren	-	-
Gr12	-	-	C	-	-	-
Gr14	-	-	C	> iren	-	-
Gr15	-	-	C	-	< Z_hin	-
Gr22	-	-	B	-	<< Z_hin	-
Gr23	-	-	B	>> iren	< nSI_hin	-
Gr24	-	-	C	> iren	-	-
Gr28	-	-	-	> iren	-	-

(Continued on next page)

Table 2. Continued

GeneID	Fixed aa	Sub/site	Phyl	> <i>kA/kS</i>	< <i>kA/kS</i>	Exons
Gr29	-	-	C	>> iren, > nSI_hin	-	-
Gr30	-	-	B	-	-	-
Gr35	-	-	-	>> iren	< sSI_hin	-
Gr36	-	-	B	-	<< iren, sSI_hin	-
Gr39	Z_hin	K to N, 345	B	> iren	-	-
Gr40	Z_hin	S to T, 371	C	-	-	-
Gr45	-	-	-	> nSI_hin	-	-
Gr46	-	-	C	-	<< iren	-
Gr48	-	-	-	> sSI_hin	-	-
Gr50	-	-	-	>> iren	< nSI_hin	-
Gr51	-	-	-	> iren	-	-
Gr52	-	-	-	> iren	-	-
Gr54	-	-	C	-	-	-
Gr55	-	-	C	-	< iren	-
Gr57	Z_hin	A/T to G, 288	B	>> Z_hin	-	-
Gr58	-	-	C	>> nSI_hin	-	-
Gr59	-	-	C	-	< iren	-
Gr60	Z_hin	A/- to G; D to T, 42; 286	B	-	<< iren	-
lr7h.1	-	-	C	> iren	-	-
lr7i	Z_hin	K to N; R to H, 2; 492	C	-	<< iren	-
lr7s	-	-	C	-	< iren	-
lr7u	Z_hin	S to A, 109	C	>> iren	-	-
lr7w	-	-	-	> sSI_hin	-	-
lr8a	Z (1), Z_hin (3)	Figure 4	B	>> A v Z	-	-
lr21a	-	-	B	-	<< sSI_hin, < iren	Exons – nSI_hin
lr25a	-	-	-	>> iren, nSI_hin	-	-
lr31a	-	-	C	-	-	-
lr40a	-	-	-	>> iren,	<< nSI_hin, < sSI_hin	-
lr40c	-	-	C	-	-	Exons – Z
lr41a	-	-	-	> nSI_hin	<< iren, sSI_hin	-
lr41b	-	-	C	>> Z_hin	-	-
lr41c	-	-	-	>> iren	<< nSI_hin, < sSI_hin	-
lr68a	-	-	C	-	-	-
lr101	-	-	C	> Z_hin	-	-
lr134	-	-	C	> iren	<< Z_hin	-
lr136	-	-	-	> iren	-	-
lr137	-	-	-	>> iren	-	-
lr139	-	-	B	-	-	-

iren = *A. irenicus*; Z_hin = zoophilic *A. hinesorum*; nSI_hin = *A. hinesorum* from northern Solomon Archipelago; sSI_hin = *A. hinesorum* from southern Solomon Archipelago. GeneID = Gene orthologue from *Anopheles gambiae*; Fixed aa = genes showing evidence of fixed amino acid differences between anthropophilic and zoophilic lineages; Sub/site = amino acid substitution (anthropophilic to zoophilic) and position in amino acid alignment when substitution has occurred; Phyl = candidates showing phylogenetic patterns B or C, as shown in Figure 3; > *kA/kS* = genes showing higher *kA/kS* ratios in zoophilic/anthropophilic (Z/A) comparisons relative to anthropophilic/anthropophilic (A/A), based on differences in SE or IQR (>) or SE and IQR (>>); < *kA/kS* = genes showing lower *kA/kS* ratios in zoophilic/anthropophilic (Z/A) comparisons relative to anthropophilic/anthropophilic (A/A), based on differences in SE or IQR (>) or SE and IQR (>>); Exons = genes with evidence of insertions or deletions in coding regions. See also Figures S4–S7.

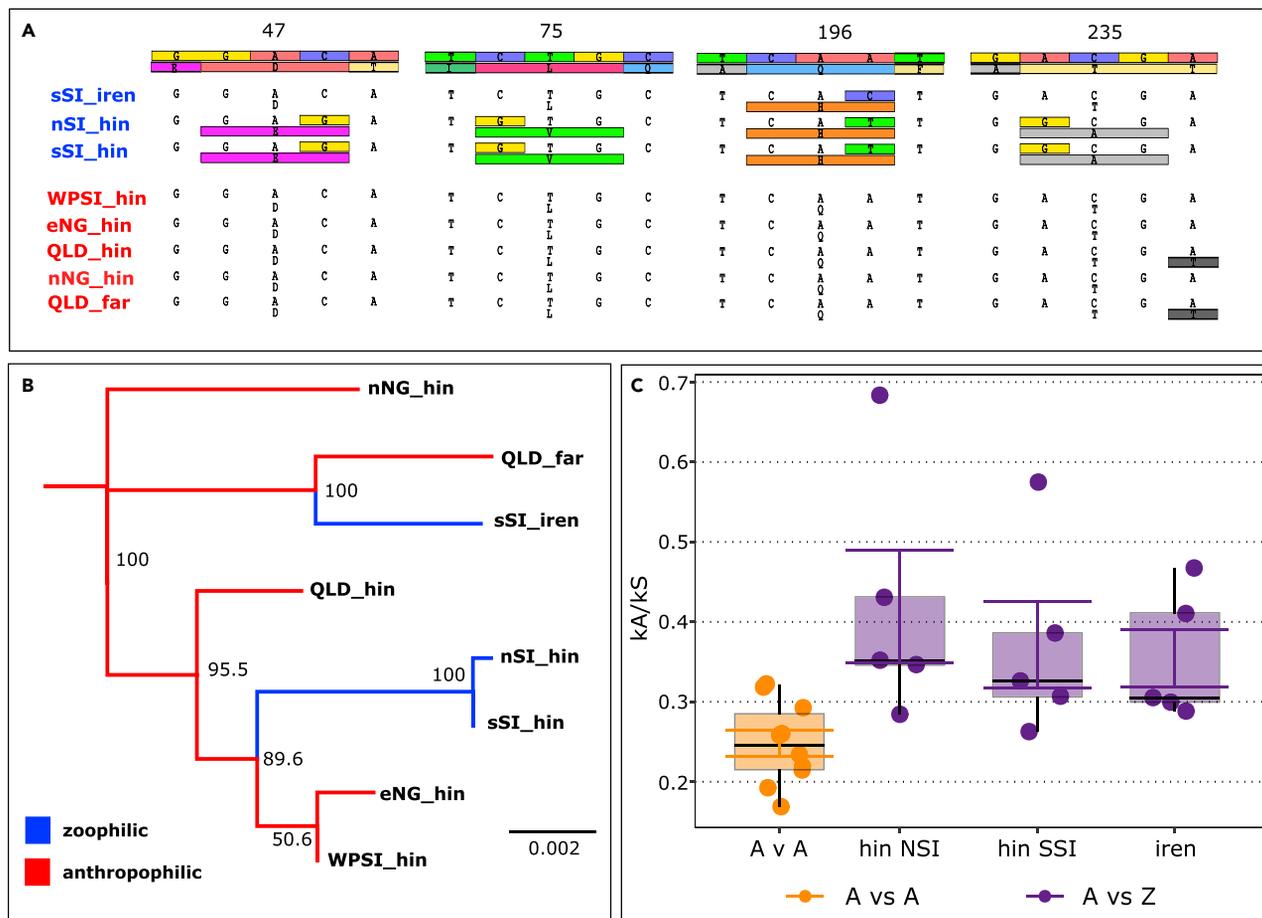


Figure 4. *Ir8a* – patterns of molecular evolution

(A) Codons in *Ir8a* in with fixed amino acid differences between anthropophilic and zoophilic populations/species. Numbers above the alignment represent the amino acid position in the protein alignment.

(B) Neighbor-joining phylogeny (Jukes Cantor) for *Ir8a*. Support values are based on 1000 bootstrap replicates.

(C) Box-pots comparing pairwise kA/kS ratios between same (A vs A) and different (A vs Z) phenotype comparisons for *Ir8a*, including A vs Z comparisons for each separate zoophilic individual sequenced. A = anthropophilic; Z = zoophilic. Median for each group is shown by black line, standard errors are shown by colored lines and interquartile ranges are shown by shaded areas.

See Table 2 for a summary of candidate genes involved in differences in host preference based on evolutionary patterns.

Candidate genes involved in human host detection – tests of selection

To examine the intensity of selection that may have operated on genes belonging to the four major olfactory gene families (*Ors*, *Gr*s, *Irs*, and *Obps*) in the *A. farauti* complex, we use the program RELAX (Wertheim et al., 2015). We used this program to identify genes for which zoophilic lineages are likely to have experienced either relaxed or intensified selection. We initially screened genes for intragenic recombination and found no evidence of significant recombination breakpoints within gene alignments using the GARD method in Hyphy ($p > 0.01$). Using RELAX, we specified zoophilic lineages as foreground (test) branches and anthropophilic lineages as background branches. Using a significance threshold of $p < 0.05$, we found that *Or46*, *Or48*, *Gr2*, *Ir100a*, and *Obp15* are likely to have experienced relaxed selection in all three zoophilic lineages. At $p < 0.1$, *Or80*, may also have experienced relaxed selection in zoophilic lineages. We found strong evidence ($p < 0.05$) of intensified selection in all zoophilic lineages for *Obp2*, *Obp10*, *Obp48*, *Or6*, *Gr13*, *Gr44* and *Ir135*, and at $p < 0.1$ for *Or28*, *Or29*, *Or75*, *Gr7*, *Gr22*, *Gr40*, *Gr41* and *Ir8a* (Table 3). *Ir8a* is a gene that has been previously implicated as a candidate locus for the preference of humans by *Ae. aegypti* (Raji et al., 2019).

Table 3. Results from tests of selection performed using HyPhy (Kosakovsky Pond et al., 2020)

GeneID	aBSREL, p	ω , %sites	BUSTED, p	RELAX, p, LR, K
<i>Obp2</i>	-	-	-	Int, 0.022, 5.21, 50
<i>Obp3</i>	sSI_hin, 0.0078	6840, 0.83	Y, 0.01	-
<i>Obp5</i>	sSI_hin, 0.0187	228, 0.71	Y, 0.026	-
<i>Obp10</i>	sSI_iren, 0.0293	554, 3.7	Y, 0.033	Int, 0.044, 4.07, 3.13
<i>Obp15</i>	-	-	-	Rel, 0.044, 4.05, 0.23
<i>Obp48</i>	-	-	-	Int, 0.001, 10.93, 50
<i>Obp55</i>	sSI_hin, 0.0184	6350, 0.86	Y ^a , 0.095	-
<i>Obp71</i>	sSI_iren ^a , 0.075	82.5, 2.7	-	-
<i>Or6</i>	-	-	-	Int, 0.034, 4.52, 2.20
<i>Or28</i>	-	-	-	Int ^a , 0.090, 2.88, 7.61
<i>Or40</i>	sSI_hin ^a , 0.0627	120, 0.59	-	-
<i>Or46</i>	-	-	-	Rel, 0.008, 7.09, 0.05
<i>Or48</i>	-	-	-	Rel, 0.029, 4.76, 0.48
<i>Or75</i>	-	-	-	Int ^a , 0.081, 3.04, 4.79
<i>Or80</i>	-	-	-	Rel ^a , 0.094, 2.18, 0.17
<i>Gr1</i>	nSI_hin ^a , 0.0584	132, 0.76	Y ^a , 0.091	-
<i>Gr2</i>	-	-	-	Rel, 0.013, 6.21, 0.38
<i>Gr7</i>	-	-	-	Int ^a , 0.088, 2.91, 3.60
<i>Gr13</i>	sSI_hin, 0.0093	5010, 0.26	Y, 0.021	Int, 0.018, 5.56, 6.23
<i>Gr22</i>	-	-	-	Int ^a , 0.091, 2.85, 43.46
<i>Gr40</i>	-	-	-	Int ^a , 0.059, 3.58, 50
<i>Gr41</i>	-	-	-	Int ^a , 0.071, 3.26, 24.69
<i>Gr44</i>	-	-	-	Int, 0.022, 5.28, 4.96
<i>Gr58</i>	-	-	Y, 0.017	-
<i>Ir8a</i>	NHB_hin ^a , 0.0849	54.9, 1.1	-	Int ^a , 0.066, 3.37, 3.51
<i>Ir40c</i>	-	-	Y ^a , 0.075	-
<i>Ir100a</i>	-	-	-	Rel ^a , 0.055, 3.68, 0
<i>Ir135</i>	-	-	-	Int, 0.013, 6.14, 2.39
<i>Ir142</i>	sSI_iren, 0.0316	32, 0.8	Y ^a , 0.089	Int sSI_iren ^a , 0.055, 3.67, 2.12

GeneID = Gene orthologue from *Anopheles gambiae*; ω , %sites = omega values on branches under selection for aBSREL (Smith et al., 2015) analyses and percentage of sites under selection on those branches; BUSTED, p = evidence of selection in BUSTED (Murrell et al., 2015) analysis (Y/-) and associated p value; RELAX, p, LR, K = genes under selection based on RELAX (Wertheim et al., 2015) analyses: Int = evidence of intensifying selection, Rel = evidence of relaxed selection, p = p value associated with test; LR = likelihood ratio; K = K value associated with test;

^aIndicates tests with p values >0.05 but <0.1.

We also used the program aBSREL (Smith et al., 2015) to find genes for which zoophilic lineages may have experienced episodic diversifying selection and the program BUSTED (Murrell et al., 2015) to identify genes for which zoophilic lineages may have experienced gene-wide episodic selection. The aBSREL method found evidence of episodic diversifying selection in at least one zoophilic lineage for the following genes: *Obp3*, *Obp5*, *Obp10*, *Obp55*, *Obp71**, *Or40**, *Gr1**, *Gr13*, *Ir8a**, *Ir142*. BUSTED found evidence of diversifying selection in at least one zoophilic lineage for *Obp3*, *Obp5*, *Obp10*, *Obp55*, *Gr13*, and *Gr58*. Genes showing marginally significant evidence for selection (p < 0.1) in at least one zoophilic lineage include *Gr1*, *Ir40c*, and *Ir142*.

Overall, we found five genes that are potential candidates based on meeting at least two expected evolutionary patterns as well as showing at least one significant test of positive selection. We found an additional two genes that show three evolutionary patterns but show no evidence of selection, ten genes that show

one evolutionary pattern and positive selection and five genes that show evidence of relaxed selection alone (Table 4). Some of these genes have been previously identified for potential roles in mosquito blood-feeding behavior, as well as differences in human host preference. We have narrowed these to a subset of four genes (*Ir8a*, *Or75*, *Obp22*, and *Gr57*) that, based on the combined evidence, are the strongest candidates for involvement in differences in human host preference (Table 4). These genes show three of the four evolutionary criteria expected for candidate genes and two of them (*Ir8a* and *Or75*) also show evidence of positive or intensified selection in zoophilic lineages.

DISCUSSION

Overview

To identify genes associated with anthropophily in *Anopheles* mosquitoes, we compared the evolution of known olfactory genes in an *Anopheles* species complex showing differences in human host preference. We found that synteny on a chromosomal level is largely conserved between the *Anopheles farauti* species complex and *Anopheles gambiae*. Furthermore, we found that most olfactory genes (*Ors*, *Grs*, *Irs*, and *Obps*) have been subject to purifying selection signified by mean *kA/kS* values much less than one. However, we identified a subset of genes that have been subject to either positive (ten genes aBSREL, nine genes BUSTED – seven genes in an agreement between these methods), relaxed (six genes), or intensified selection (15 genes) in some zoophilic lineages (Table 3). Eleven genes showed fixed amino acid differences in intraspecific zoophilic lineages (within *A. hinesorum*), and one gene (*Ir8a*) contains a fixed amino acid difference at the same location on the sequence in all zoophilic lineages (*A. hinesorum* and *A. irenicus*) (Table 2). Phylogenetic relationships strongly suggestive of olfactory genes involved in anthropophily were also found in a further 28 genes (Table 2, Figure 3); however, we find no evidence for introgression of olfactory genes between the two zoophilic species. Based on this combined evidence, we find 22 genes that may be involved in differences in host preference, with a small subset of four genes that show promise as the strongest candidates for being involved in these mosquitoes' capacity for anthropophily (Table 4). Some of the genes identified have been previously identified in other mosquito study systems, but others have not previously been implicated in mosquito anthropophily.

Olfactory genes in insects: function, evolution, and role in host preference

Insect olfactory genes consist of four major families—odorant receptors (*Ors*), Gustatory receptors (*Grs*), Ionotropic receptors (*Irs*), and odorant-binding proteins (*Obps*). Three of these families (*Ors*, *Grs*, and *Irs*) produce proteins that are ligand-gated ion channels that span membranes of olfactory neurons, located in insect olfactory organs of insects (Clyne et al., 1999; Benton et al., 2009; Isono and Morita, 2010). When activated by chemical compounds, olfactory receptors can generate either excitatory or inhibitory signals (Xu et al., 2019), which are interpreted in the olfactory ganglia of the insect brain (Li and Liberman, 2015). Different receptors also vary in their sensitivity and specificity to odors (Andersson et al., 2015) and operate in a combinatorial fashion which is still not fully understood (Andersson et al., 2015; Haverkamp et al., 2018). The other major olfactory gene family, odorant-binding proteins, are soluble globular proteins found in the hemolymph of insect olfactory organs (antennae and palps) (Steinbrecht, 1998). There, they bind hydrophilic chemicals at the surface of the olfactory organs, and transport them through the hemolymph, to enable their contact with the receptors outlined above (Pelosi and Maida, 1995; Steinbrecht, 1998).

The oldest of the four major insect olfactory gene families are the *Grs*, followed by the *Irs*. Both these gene families evolved prior to invertebrates colonizing land at approximately 400 Mya (Robertson et al., 2003) and are also present in other invertebrate groups (Croset et al., 2010; Eyun et al., 2017). The other two gene families, *Ors* and *Obps*, are only present in insects and are thought to have evolved as an adaptation to the colonization of the terrestrial environment by early insects (Sánchez-Gracia et al., 2009; Eyun et al., 2017; Brand et al., 2018). Olfactory genes in insects evolve by birth-death evolution (Sánchez-Gracia et al., 2009), a process that involves gene duplication. In some genes, these duplication events can result in a release of constraint from purifying selection and potential evolution of novel functions (birth), or a loss of function via pseudogenization (death) (Nei and Rooney, 2005). This means that there are frequent losses and gains of olfactory genes between insect taxa resulting in highly variable numbers of olfactory genes in different insect species. Overall, insect olfactory gene families have been found to be under purifying selection; however, there is evidence that they may be evolving faster on average than other gene families in some taxa, including *Anopheles* mosquitoes (Neafsey et al., 2015).

Table 4. Candidate genes based on combined evidence

Gene	Evidence	Function	Expression	Study system	Previous ID
3 patterns + selection					
Or75 (AGAP002045)	int sel (RELAX); <i>kA/kS</i> , aa (hin Z), phyl	Detecting terpenes	No Rinker; Y Athrey & Pitts	<i>Anopheles</i>	Expressed antennae downregulated after bloodmeal (Omondi et al., 2019)
Ir8a (AGAP010411)	+ sel (aBSREL hin Z); <i>kA/kS</i> , aa (all Z), phyl	Detecting lactic acid	Y Rinker, Pitts, Athrey	<i>Aedes and Anopheles</i>	Combined – see discussion (Jason Pitts et al., 2017; Raji et al., 2019; Athrey et al., 2020)
3 patterns					
Obp22 (AGAP010409)	aa (hin Z), phyl, indels (hin Z)	Diel cycle	Y Rinker, Pitts, Athrey	<i>Anopheles</i>	Upregulated in dark (Bivand and Rundel, 2013)
Gr57 (AGAP004716)	<i>kA/kS</i> , aa (hin Z), phyl	–	Y Rinker, Pitts, Athrey	<i>Anopheles</i>	<i>kA/kS</i> (Rinker et al., 2013)
2 patterns + selection					
Obp3 (AGAP001409)	+ sel (aBSREL hin sSI, BUSTED); <i>kA/kS</i> , phyl	–	Y Rinker, Athrey, Pitts	<i>Anopheles</i>	Expression (Rund et al., 2011; Rinker et al., 2013; Athrey et al., 2017)
Or28 (AGAP002722)	int sel (RELAX); <i>kA/kS</i> , phyl	Detecting sulcatone (Suh et al., 2016)	Palps, f & m Y Rinker, Athrey, Pitts	<i>Anopheles</i>	–
Ir40c	+ sel (BUSTED); indels (all Z)	–	–	–	–
1 pattern + selection					
Obp5 (AGAP009629)	+ sel (aBSREL hin sSI), BUSTED; indels (hin Z)	–	Y Rinker, Pitts, Athrey	<i>Anopheles</i>	Expression (Athrey et al., 2020), Fst (Main et al., 2016)
Obp10 (AGAP001189)	+ sel (aBSREL iren), BUSTED, int (RELAX); phyl	–	Y Rinker, Pitts, Athrey	<i>Anopheles</i>	Male biased expression (Athrey et al., 2020)
Obp71 (AGAP006074)	+ sel (aBSREL iren); <i>kA/kS</i>	–	Y Rinker, Pitts, Athrey	–	–

(Continued on next page)

Table 4. Continued

Gene	Evidence	Function	Expression	Study system	Previous ID
<i>Or6</i> (AGAP006167)	int sel (RELAX); indels (hin nSI)	–	Y Rinker, Pitts, Athrey	–	-
<i>Or40</i> (AGAP002558)	+ sel (aBSREL hin sSI); <i>kA/kS</i>	–	N Rinker, Y Pitts & Athrey	–	-
<i>Gr1</i> (AGAP004114)	+ sel (aBSREL hin nSI, BUSTED); phyl	–	Y Rinker, Pitts, Athrey	–	-
<i>Gr22</i> (AGAP009999)	int sel (RELAX); phyl	Detecting CO2	Y Rinker, Pitts, Athrey	<i>Anopheles</i>	Expression (Athrey et al., 2021)
<i>Gr40</i> (AGAP001120)	int sel (RELAX); aa (hin Z)	–	N Rinker, Y Athrey	–	-
<i>Gr41</i> (AGAP001122)	int sel (RELAX); phyl	–	N Rinker, Y Athrey	–	-
Relaxed selection					
<i>Obp15</i>	–	–	–	<i>Anopheles</i>	<i>kA/kS</i> (Rinker et al., 2013)
<i>Or46</i> (AGAP009392)	–	–	Y Rinker, Pitts, Athrey	<i>Anopheles</i>	Expression (Athrey et al., 2017)
<i>Or48</i> (AGAP006666)	–	Detecting ketones and alcohol (larvae) (Xia et al., 2008)	Y Rinker, Pitts, Athrey	<i>Anopheles</i>	–
<i>Or80</i> (AGAP005495)	Phyl	–	Y Rinker, Pitts, Athrey	<i>Anopheles</i>	Expression (Athrey et al., 2020)
<i>Gr2</i>	–	–	–	–	–
<i>Ir100a</i>	–	–	Antennae	<i>Anopheles</i>	Expression (Rinker et al., 2013; Athrey et al., 2020)

GeneID and evidence = Gene ortholog from *Anopheles gambiae*; **Function/Expression** = information on gene function and/or expression (if known); **Previously identified** = has the gene been previously identified as a potential candidate in human blood-feeding behavior? (Yes/No); **Study system** = the study system in which the gene was identified as a candidate; **Evidence** = evidence for the involvement of gene in anthropophily.

Because of the central role of olfaction in insect host preference, it has been hypothesized that host shifts are likely to be associated with changes in olfactory genes (Matsuo et al., 2007; McBride and Arguello, 2007; Vieira et al., 2007). Consistent with previous studies (Vieira et al., 2007; Sánchez-Gracia et al., 2009), we find that olfactory genes are largely conserved, with purifying selection being the dominant evolutionary force that is operating on these gene families in the *A. farauti* complex. Unlike in some *Drosophila* systems (Vieira et al., 2007), we find no evidence for pseudogenization playing a role in host shifts in the species complex. Confirming our initial hypothesis, we find evidence that both positive and relaxed purifying selection have acted on a subset of olfactory genes that may be involved in differences in host preference in the species complex. This is consistent with previous *Drosophila* studies investigating olfactory gene evolution in host specialist species (McBride, 2007; McBride and Arguello, 2007; Nozawa and Nei, 2007; Yassin et al., 2016), as well as other studies that have shown that the relaxation of selection is a common precursor to behavioral changes involving a loss of function (Coss, 1999; Lahti et al., 2009; Wund et al., 2015; Calderoni et al., 2016; Lu et al., 2019; Tiwary, 2020).

The *Anopheles farauti* complex – A model system for studying host shifts and the genetic basis of anthropophily in mosquitoes

The zoophilic species and populations of the *A. farauti* complex are only found within the Solomon Archipelago, having evolved from widespread species (*A. farauti* and *A. hinesorum*) that exist in New Guinea, northern Australia, and into the Solomon Archipelago (Beebe et al., 2002, 2015). The ancestral host preference in the *Anopheles farauti* complex is generalist mammal-biting (Ambrose et al., 2012) but the host preferences and host ranges are unknown for zoophilic populations. At the time of the colonization of the Solomon Archipelago approximately two Mya (Ambrose et al., 2012) there would have been and a limited range of hosts on these volcanically emerging islands. The archipelago's limited biodiversity could have resulted in host range (host specialization) and possibly for a host shift to divergent host group/s such as birds, amphibians, or reptiles. Humans arrived in the Archipelago approximately 30 Kya (Wickler and Spriggs, 1988), long after the initial colonization of the Archipelago by species of the *Anopheles farauti* complex.

In *A. gambiae* and *Ae. aegypti*, the two main established systems for studying the genetic basis of anthropophily in mosquitoes, highly anthropophilic populations/species have evolved from generalist ancestors (White et al., 2011; Rose et al., 2020). This is in contrast to the *A. farauti* complex where specialization has occurred in the opposite direction, with exclusively zoophilic populations/species evolving from anthropophilic (generalist) ancestors (Ambrose et al., 2012). Population genetic and phylogenetic relationships (including convergent evolution of zoophily) within the *A. farauti* complex, as well as differences in host preference also make them an ideal system to study the evolution of changes in host preference in mosquitoes (Ambrose et al., 2012). As olfaction is central to host preference in mosquitoes, this behavioral shift is likely linked to changes in the olfactory gene repertoire of zoophilic lineages.

Microsatellite data, as well as nuclear sequence data, have shown that the Solomon Archipelago *A. hinesorum* populations are more closely related to each other for most of their nuclear genomes than to other populations of the species from other locations (Ambrose et al., 2021). This is despite these populations belonging to highly divergent mitochondrial lineages that exist in different parts of the Archipelago (the north and the south). Based on their mitochondrial divergence from the rest of the species which occur through northern Australia and New Guinea, these two lineages are thought to have colonized the Solomon Archipelago at different times in the past. One lineage is estimated to have arrived at approximately 0.5 Mya and the other over two Mya (Ambrose et al., 2012). In this study, we also observe that Solomon Archipelago populations are most closely related to most olfactory genes. This is consistent with the hypothesis that gene flow may have contributed to the spread of zoophily from the older lineage into the more recently arriving lineage; however, we cannot rule out incomplete lineage sorting as an alternative explanation.

It has already been established, via shared mitochondrial haplotypes, that there has likely been recent movement of females from a New Guinean population into the recently discovered anthropophilic island population (Ambrose et al., 2021). Parts of the nuclear genome associated with anthropophily may have been retained following the recent movement of females from New Guinea into this population. This would be reflected by close phylogenetic relationships between anthropophilic *A. hinesorum* populations (found in New Guinea and Australia) and the newly discovered anthropophilic *A. hinesorum* population in the

Solomon Archipelago, a pattern we observe in 28 olfactory genes. Owing to the presence of humans in the Archipelago, anthropophily may now present a selective advantage in this population.

Candidate genes associated with mosquito anthropophily – insights from a new model system

Mosquitoes use many cues to detect suitable hosts. In mosquitoes that feed on humans, odor is considered the most important (McBride, 2016), but at closer range thermal and visual cues are also used (Van Breugel et al., 2015). Previous work on the two primary mosquito model systems, *A. gambiae* (*An. coluzzii*) and *Ae. aegypti*, have identified chemicals (kairomones) that may be involved in detecting humans as hosts, as well as receptors and genes which respond to these kairomones. Both of these species have evolved to be anthropophilic from zoophilic species making them useful systems for studying the molecular basis of anthropophily. Studies on these model systems have included behavioral and electrophysiological assays, revealing the attractiveness of various kairomones and combinations of chemical cues. Some of the most important kairomones identified in human host detection by mosquitoes include lactic acid (Acree et al., 1968; Steib et al., 2001; Dekker et al., 2002), sulcatone (McBride et al., 2014), indole (Biessmann et al., 2010) and ammonia (Geier et al., 1999; Braks et al., 2001; Smallegange et al., 2005). It is well established that carbon dioxide is important in the host-seeking response in many mosquitoes (Gillies, 1980), acting synergistically to sensitize mosquitoes to other kairomones and host cues (Dekker et al., 2005; McMeniman et al., 2014). The genes involved in the detection of carbon dioxide have been identified and are gustatory receptors expressed on the maxillary palps of mosquitoes (Erdelyan et al., 2012).

In this study, we identified several genes that may be involved in differences in human host detection in *Anopheles* mosquitoes. These candidates were identified based on a combination of tests of selection as well as four evolutionary criteria. Many of these genes have been shown to be expressed in the olfactory organs of other *Anopheles* mosquitoes (Rinker et al., 2013; Athrey et al., 2017) (Table 4). The olfactory co-receptor, *Ir8a* was one of the strongest candidate genes, meeting three of four evolutionary criteria (phylogenetic pattern, *kA/kS*, fixed amino acid substitution—see Table 2, Figure 4), as well as showing evidence for intensified selection (Table 3). It was the only gene that showed more than two fixed amino acid substitutions between anthropophilic and zoophilic *A. hinesorum* populations (with three additional fixed amino acids in zoophilic *A. hinesorum*). *Ir8a* was also unique in that we found a single common fixed amino acid substitution in all zoophilic lineages (*A. hinesorum* and *A. irenicus*). Even more intriguing is that this substitution resulted from a different nucleotide mutation at the third position of codon 196 in the two different species (A to C in *A. irenicus* and A to T in *A. hinesorum*). These results together provide evidence that variation in *Ir8a* may be central to governing differences in human host preference in mosquitoes.

Consistent with our findings, Raji et al., (2019) performed gene knockdown of the ionotropic co-receptor *Ir8a* in *Ae. aegypti* and showed that mosquitoes lacking a functional copy of this gene did not respond to human odor (Raji et al., 2019). *Ir8a* has been shown to be important in the detection of carboxylic acids in *Anopheles* and Raji et al. showed that in *Ae. aegypti*, it responds to lactic and other acids which are components of human sweat. They also showed that knocking this gene down was sufficient to remove the host-seeking response of *Ae. aegypti* despite the retention of functional *Orco* and carbon dioxide receptor genes. *Ir8a* has also been shown to form functional complexes with *Ir64a* and *Ir75k* which are involved in detecting acids (Jason Pitts et al., 2017). This gene has also been shown to be involved in the detection of carboxylic acids in *Anopheles* (Jason Pitts et al., 2017). Our findings add to the growing body of evidence supporting *Ir8a* as an important functional candidate for human host preference and provide the first direct genetic evidence to our knowledge that this gene may be important in host preference in *Anopheles* mosquitoes.

The finding of an olfactory co-receptor gene (which is usually highly conserved) undergoing positive selection is unexpected. One previous study has reported possible evidence for positive selection acting on the highly conserved insect olfactory co-receptor, *Orco*, in some broad (order level) insect lineages (Soffman et al., 2018). The authors of this study suggest that this may imply functional flexibility of this highly conserved co-receptor but do not discuss details of this functional flexibility. Positive selection acting on a co-receptor may facilitate the rapid evolution of differences in sensitivity to entire classes of compounds (such as acids), which may, in turn, allow a species to quickly adapt to divergent hosts which emit a very different array of kairomones. It is also possible that the amino acid changes that have occurred only affect interactions with specific tuning proteins, meaning that possible pleiotropic effects may have more narrow

implications on chemical compound detection. Our study is the only published work, to our knowledge, reporting rapid adaptive evolution in an olfactory ionotropic co-receptor. This novel result is, therefore, worth investigating in other insect and mosquito species that have undergone host shifts.

As well to *Ir8a*, we find six other genes that we identify as strong candidates based on combined criteria and six genes based on evidence for relaxed selection in zoophilic lineages (Table 4). One of these, *Obp22*, contains a fixed amino acid substitution, indels, and shows phylogenetic relationships between anthropophilic and zoophilic *A. hinesorum*, suggesting a possible role in differences in host preference phenotype. These gene neighbors (and is therefore possibly co-regulated with) *Ir8a* in both *A. gambiae* and *A. farauti* genomes. Expression of *Obp22* in the antennae of adult female *A. gambiae* has previously been shown to be downregulated following exposure to light (Das and Dimopoulos, 2008). Another candidate identified, *Or75*, is expressed in adult female antennae of *An. coluzzii*, with increased transcript abundance in older (host-seeking) females than in newly emerged females. *Or75* responds to terpenes—a component of human sweat—with human host preference by *An. coluzzii* apparently being dependent on the presence of terpenes and other human odors. Furthermore, host-seeking females are more attracted to odor blends including terpenes than blends not containing these kairomones (Omondi et al., 2019).

The six genes that were found to be under relaxed selection (Table 4) should also be considered potential strong candidates for playing a role in host preference differences in the species complex. Three of these genes (*Or46*, *Or80*, and *Ir100a*) show differences in expression indicative of a role in anthropophily and one (*Obp15*) has a *kA/kS* ratio over one in a comparison made between *A. gambiae* and *An. quadriannulatus*, suggesting that it has been subject to positive selection in other species (Rinker et al., 2013). The response profile of *Or46* has been assessed against a wide array of chemicals and was found to be a broadly tuned receptor that responds strongly to a wide variety of chemicals but most strongly to isobutyl-acetate in *A. gambiae* (Carey et al., 2010). The ionotropic receptor, *Ir100a* is a receptor that has been found to be expressed at significantly higher levels in the antennae and maxillary palps of the anthropophilic species *An. coluzzii* compared with the zoophilic *An. quadriannulatus* (Athrey et al., 2017), also suggests a potential role in host preference differences. Although no evidence for a role in anthropophily, *Or80* has been shown to be expressed at significantly higher levels in female versus male antennae in *An. quadriannulatus* (Athrey et al., 2020), potentially suggesting a role in host-seeking behavior that is not yet understood. Finding some of the same genes in the *A. farauti* complex as has been previously identified in other species provides evidence that these genes may be important in anthropophilic behavior in a diverse mosquito species.

Conclusions and future directions

This study contributes significantly to our understanding of the genetic basis of mosquito host preference. Using a novel *Anopheles* study system, we focus on a set of gene families with broadly known chemosensory function, identifying candidate genes involved in the human host preference, some known and others novel. The four strongest candidate genes identified (Table 4) have been previously implicated in human host preference in other mosquito taxa, validating the evolutionary genetics approaches used in this study for finding genes related to anthropophily. Future work using this powerful system will include comparative population genomics studies to further reveal the genetic basis of human host preference in *Anopheles* mosquitoes. By investigating both the evolution of genes of known function (such as by performing McDonald-Kreitman (McDonald and Kreitman, 1991) tests on the candidate genes identified by this study) as well as performing comparative whole-genome analyses (such as by using PCA based methods), a more complete picture of the molecular basis of differences in host preference in *Anopheles* mosquitoes may emerge. Additionally, gene knockdowns of candidate genes performed in colonies derived from anthropophilic populations will further validate the roles of these genes in human host preference of mosquitoes, as well as uncovering their specific molecular functions.

This work is an initial step toward the understanding of the genetic basis of anthropophily in the *A. farauti* complex. However, additional work is needed to gain a more complete understanding of the host preference of zoophilic species in the complex, as well as the genetic basis of the differences in behavior observed. Although we can be confident that exclusively zoophilic lineages do not feed on humans in their natural environment, determining the preferred hosts and the host range of zoophilic *Anopheles* species in the complex will provide an important complement to understanding the genetics underlying this phenotype. The *de novo* assembly of mosquitoes from anthropophilic and zoophilic lineages of the *A. farauti*

complex may also reveal novel chemosensory genes and structural genomic changes (such as inversions) involved in the observed differences in behavior. Differences in gene expression are also likely to explain some variation in host preference but were not investigated in this study. It would be valuable for future work to investigate the role of gene expression by performing RNA-seq experiments on olfactory organs of mosquitoes in this species complex.

Overall, our findings provide a valuable complement to the large body of work conducted on the other two major study systems (*Ae. aegypti* and the *A. gambiae* complex), focused on identifying the genetic basis of anthropophily in mosquitoes. This system provides a promising addition to advancing our understanding of this behavior, which will have important implications for eradicating malaria from the southwest Pacific, and possibly more broadly on mosquito disease transmission.

Limitations of the study

As stated above, this study is an initial step toward developing the *A. farauti* complex as a complementary system to advance and broaden our knowledge of the genetic basis of anthropophily in mosquitoes. Some specific limitations are mentioned in the [conclusions and future directions](#) section of the [discussion](#). These include a lack of knowledge of the hosts used by zoophilic members of the complex and no current data on potential differences in gene expression in species/populations with differing host preferences. Manipulative experiments involving behavioral assays and gene-knockdowns would increase our confidence in the role of the candidate olfactory genes identified in differences in host preference.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
 - Lead contact
 - Materials availability
 - Data and code availability
- [EXPERIMENTAL MODEL AND SUBJECT DETAILS](#)
- [METHOD DETAILS](#)
 - Whole genome sequencing, mapping, variant calling and phylogenies
 - Characterization and genomic location of olfactory genes in the *An. farauti* complex
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)
 - Candidate genes involved in human host detection – Evolutionary patterns
 - Candidate genes involved in human host detection – Tests of selection

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2022.104521>.

ACKNOWLEDGMENTS

This work was supported by funding from the Hermon Slade Foundation (grant number 2014000818). Samples were provided by Robert Cooper (Australian Defense Force Malaria and Infectious Disease Institute), Thomas Burkot, and Tanya Russell (James Cook University).

AUTHOR CONTRIBUTIONS

LA and NWB conceived the project. LA wrote the article, generated data for the project, and performed most analyses. IP and JH assisted with data analysis, advised on selection analyses, and provided input on interpreting results. NWB, DOB, IP, and JH also provided feedback and had significant input in writing the article. DOB assisted in the logical conception and interpretation of analyses and results and provided feedback on article drafts.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 9, 2021

Revised: May 8, 2022

Accepted: May 30, 2022

Published: July 15, 2022

REFERENCES

- Acree, F., Turner, R.B., Gouck, H.K., Beroza, M., and Smith, N. (1968). L-lactic acid: a mosquito attractant isolated from humans. *Science* **161**, 1346–1347. <https://doi.org/10.1126/science.161.3848.1346>.
- Ambrose, L., Ortiz-Barrientos, D., Cooper, R.D., Lobo, N.F., Burkot, T.R., Russell, T.L., and Beebe, N.W. (2021). Gene flow between island populations of the malaria mosquito, *Anopheles hinesorum*, may have contributed to the spread of divergent host preference phenotypes. *Evol. Appl.* **14**, 2244–2257. <https://doi.org/10.1111/evo.13288>.
- Ambrose, L., Riginos, C., Cooper, R.D., Leow, K.S., Ong, W., and Beebe, N.W. (2012). Population structure, mitochondrial polyphyly and the repeated loss of human biting ability in anopheline mosquitoes from the southwest Pacific. *Mol. Ecol.* **21**, 4327–4343. <https://doi.org/10.1111/j.1365-294X.2012.05690.x>.
- Andersson, M.N., Löfstedt, C., and Newcomb, R.D. (2015). Insect olfaction and the evolution of receptor tuning. *Front. Ecol. Evol.* **3**. <https://doi.org/10.3389/fevo.2015.00053>.
- Andrews, S. (2010). FastQC - a quality control tool for high throughput sequence data. Babraham Bioinform.
- Athrey, G., Cosme, L.V., Popkin-Hall, Z., Pathikonda, S., Takken, W., and Slotman, M.A. (2017). Chemosensory gene expression in olfactory organs of the anthropophilic *Anopheles coluzzii* and zoophilic *Anopheles quadriannulatus*. *BMC Genom.* **18**, 751. <https://doi.org/10.1186/s12864-017-4122-7>.
- Athrey, G., Popkin-Hall, Z., Cosme, L.V., Takken, W., and Slotman, M.A. (2020). Species and sex-specific chemosensory gene expression in *Anopheles coluzzii* and *An. quadriannulatus* antennae. *Parasit. Vectors* **13**, 212. <https://doi.org/10.1186/s13071-020-04085-3>.
- Athrey, G., Popkin-Hall, Z.R., Takken, W., and Slotman, M.A. (2021). The expression of chemosensory genes in male maxillary palps of *Anopheles coluzzii* (Diptera: Culicidae) and *An. Quadriannulatus*. *J. Med. Entomol.* **58**, 1012–1020. <https://doi.org/10.1093/jme/tjaa290>.
- Bartlett-Healy, K., Crans, W., and Gaugler, R. (2008). Phonotaxis to amphibian vocalizations in *Culex territans* (Diptera: Culicidae). *Ann. Entomol. Soc. Am.* **101**, 95–103. [https://doi.org/10.1603/0013-8746\(2008\)101\[95:ptavic\]2.0.co;2](https://doi.org/10.1603/0013-8746(2008)101[95:ptavic]2.0.co;2).
- Beebe, N.W., Bakote'e, B., Ellis, J.T., and Cooper, R.D. (2000). Differential ecology of *Anopheles punctulatus* and three members of the *Anopheles farauti* complex of mosquitoes on Guadalcanal, Solomon Islands, identified by PCR-RFLP analysis. *Med. Vet. Entomol.* **14**, 308–312. <https://doi.org/10.1046/j.1365-2915.2000.00248.x>.
- Beebe, N.W., Cooper, R.D., Waterson, D.G.E., Frances, S.P., Beebe, N.W., and Sweeney, A.W. (2002). Speciation and distribution of the members of the *Anopheles punctulatus* (Diptera: Culicidae) group in Papua New Guinea. *J. Med. Entomol.* **39**, 16–27. <https://doi.org/10.1603/0022-2585-39.1.16>.
- Beebe, N.W., Russell, T., Burkot, T.R., and Cooper, R.D. (2015). *Anopheles punctulatus* group: evolution, distribution, and control. *Annu. Rev. Entomol.* **60**, 335–350. <https://doi.org/10.1146/annurev-ento-010814-021206>.
- Benton, R., Vannice, K.S., Gomez-Diaz, C., and Vossball, L.B. (2009). Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell* **136**, 149–162. <https://doi.org/10.1016/j.cell.2008.12.001>.
- Besansky, N.J., Hill, C.A., and Costantini, C. (2004). No accounting for taste: host preference in malaria vectors. *Trends Parasitol.* **20**, 249–251. <https://doi.org/10.1016/j.pt.2004.03.007>.
- Biessmann, H., Andronopoulou, E., Biessmann, M.R., Douris, V., Dimitratos, S.D., Eliopoulos, E., Guerin, P.M., Iatrou, K., Justice, R.W., Kröber, T., et al. (2010). The *Anopheles gambiae* odorant binding protein 1 (AgamOBP1) mediates indole recognition in the antennae of female mosquitoes. *PLoS One* **5**, e9471. <https://doi.org/10.1371/journal.pone.0009471>.
- Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and genomewise. *Genome Res.* **14**, 988–995. <https://doi.org/10.1101/gr.1865504>.
- Bivand, R., and Rundel, C. (2013). rgeos: Interface to Geometry Engine - Open Source (GEOS). R Package Version 0.3-2. R package version 0.1-8.
- Borkent, A., and Belton, P. (2006). Attraction of female *Uranotaenia lowii* (Diptera: Culicidae) to frog calls in Costa Rica. *Can. Entomol.* **138**, 91–94. <https://doi.org/10.4039/N04-113>.
- Braks, M.A.H., Meijerink, J., and Takken, W. (2001). The response of the malaria mosquito, *Anopheles gambiae*, to two components of human sweat, ammonia and L-lactic acid, in an olfactometer. *Physiol. Entomol.* **26**, 142–148. <https://doi.org/10.1046/j.1365-3032.2001.00227.x>.
- Brand, P., Robertson, H.M., Lin, W., Pothula, R., Klingeman, W.E., Jurat-Fuentes, J.L., and Johnson, B.R. (2018). The origin of the odorant receptor gene family in insects. *Elife* **7**, e38340. <https://doi.org/10.7554/eLife.38340>.
- Van Breugel, F., Riffell, J., Fairhall, A., and Dickinson, M.H. (2015). Mosquitoes use vision to associate odor plumes with thermal targets. *Curr. Biol.* **25**, 2123–2129. <https://doi.org/10.1016/j.cub.2015.06.046>.
- Burkot, T.R., Bugoro, H., Apairamo, A., Cooper, R.D., Echeverry, D.F., Odabasi, D., Beebe, N.W., Makuru, V., Xiao, H., Davidson, J.R., et al. (2018). Spatial-temporal heterogeneity in malaria receptivity is best estimated by vector biting rates in areas nearing elimination. *Parasit. Vectors* **11**, 606. <https://doi.org/10.1186/s13071-018-3201-1>.
- Calderoni, L., Rota-Stabelli, O., Frigato, E., Panziera, A., Kirchner, S., Foulkes, N.S., Kruckenhauser, L., Bertolucci, C., Fuselli, S., and Panzieri, A. (2016). Relaxed selective constraints drove functional modifications in peripheral photoreception of the cavefish *P. andruzzii* and provide insight into the time of cave colonization. *Heredity* **117**, 383–392. <https://doi.org/10.1038/hdy.2016.59>.
- Carey, A.F., Wang, G., Su, C.Y., Zwiebel, L.J., and Carlson, J.R. (2010). Odorant reception in the malaria mosquito *Anopheles gambiae*. *Nature* **464**, 66–71. <https://doi.org/10.1038/nature08834>.
- Charif, D., Lobry, J.R., Necseulea, A., Palmeira, L., Perriere, G., and Penel, M.S. (2015). Package seqinr. R package.
- Clyne, P.J., Warr, C.G., Freeman, M.R., Lessing, D., Kim, J., and Carlson, J.R. (1999). A novel family of divergent seven-transmembrane proteins: candidate odorant receptors in *Drosophila*. *Neuron* **22**, 327–338. [https://doi.org/10.1016/S0896-6273\(00\)81093-4](https://doi.org/10.1016/S0896-6273(00)81093-4).
- Cooper, R.D., and Frances, S.P. (2002). Malaria vectors on Buka and Bougainville islands, Papua New Guinea. *J. Am. Mosq. Control Assoc.* **18**, 100–106.
- Coss, R.G. (1999). Effects of relaxed natural selection on the evolution of behavior. In *Geographic Variation in Behavior*, S.A. Foster and J.A. Endler, eds. (Oxford Scholarship Online). <https://doi.org/10.1093/oso/9780195082951.003.0013>.
- Costantini, C., Birkett, M.A., Gibson, G., Ziesmann, J., Sagnon, N., Mohammed, H.A., Coluzzi, M., and Pickett, J.A. (2001). Electroantennogram and behavioural responses of the malaria vector *Anopheles gambiae* to human-specific sweat components. *Med. Vet. Entomol.* **15**, 259–266. <https://doi.org/10.1046/j.0269-283X.2001.00297.x>.
- Costantini, C., Sagnon, N.F., Della Torre, A., Coluzzi, M., and della Torre, A. (1999). Mosquito behavioural aspects of vector-human interactions in the *Anopheles gambiae* complex. *Parassitologia* **41**, 209–217.
- Croset, V., Rytz, R., Cummins, S.F., Budd, A., Brawand, D., Kaessmann, H., Gibson, T.J., and Benton, R. (2010). Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. *PLoS Genet.* **6**, 1001064. <https://doi.org/10.1371/journal.pgen.1001064>.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter,

- G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>.
- Das, S., and Dimopoulos, G. (2008). Molecular analysis of photic inhibition of blood-feeding in *Anopheles gambiae*. *BMC Physiol.* 8, 23. <https://doi.org/10.1186/1472-6793-8-23>.
- Dekker, T., Geier, M., and Cardé, R.T. (2005). Carbon dioxide instantly sensitizes female yellow fever mosquitoes to human skin odours. *J. Exp. Biol.* 208, 2963–2972. <https://doi.org/10.1242/jeb.01736>.
- Dekker, T., Steib, B., Cardé, R.T., and Geier, M. (2002). L-lactic acid: a human-signifying host cue for the anthropophilic mosquito *Anopheles gambiae*. *Med. Vet. Entomol.* 16, 91–98. <https://doi.org/10.1046/j.0269-283x.2002.00345.x>.
- Erdelyan, C.N.G., Mahood, T.H., Bader, T.S.Y., and Whyard, S. (2012). Functional validation of the carbon dioxide receptor genes in *Aedes aegypti* mosquitoes using RNA interference. *Insect Mol. Biol.* 21, 119–127. <https://doi.org/10.1111/j.1365-2583.2011.01120.x>.
- Eyun, S.I., Soh, H.Y., Posavi, M., Munro, J.B., Hughes, D.S.T., Murali, S.C., Qu, J., Dugan, S., Lee, S.L., Chao, H., et al. (2017). Evolutionary history of chemosensory-related gene families across the Arthropoda. *Mol. Biol. Evol.* 34, 1838–1862. <https://doi.org/10.1093/molbev/msx147>.
- Foley, D.H., Meek, S.R., and Bryan, J.H. (1994). The *Anopheles punctulatus* group of mosquitoes in the Solomon Islands and Vanuatu surveyed by allozyme electrophoresis. *Med. Vet. Entomol.* 8, 340–350. <https://doi.org/10.1111/j.1365-2915.1994.tb00098.x>.
- Frei, J., Kröber, T., Troccaz, M., Starckenmann, C., and Guerin, P.M. (2017). Behavioral response of the malaria mosquito, *Anopheles gambiae*, to human sweat inoculated with axilla bacteria and to volatiles composing human axillary odor. *Chem. Senses* 42, 121–131. <https://doi.org/10.1093/chemse/bjw106>.
- Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing - free bayes - variant Calling - Longranger. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1207.3907>.
- Geier, M., Bosch, O.J., and Boeckh, J. (1999). Ammonia as an attractive component of host odour for the yellow fever mosquito, *Aedes aegypti*. *Chem. Senses* 24, 647–653. <https://doi.org/10.1093/chemse/24.6.647>.
- Gerts, E.M., Yu, Y.K., Agarwala, R., Schäffer, A.A., and Altschul, S.F. (2006). Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biol.* 4, 41. <https://doi.org/10.1186/1741-7007-4-41>.
- Gillies, M.T. (1980). The role of carbon dioxide in host-finding by mosquitoes (Diptera: Culicidae): a review. *Bull. Entomol. Res.* 525–532. <https://doi.org/10.1017/S0007485300007811>.
- Harpur, B.A., Kent, C.F., Molodtsova, D., Lebon, J.M.D., Alqarni, A.S., Owayss, A.A., and Zayed, A. (2014). Population genomics of the honey bee reveals strong signatures of positive selection on worker traits. *Proc. Natl. Acad. Sci. U. S. A.* 111, 2614–2619. <https://doi.org/10.1073/pnas.1315506111>.
- Haverkamp, A., Hansson, B.S., and Knaden, M. (2018). Combinatorial codes and labeled lines: how insects use olfactory cues to find and judge food, mates, and oviposition sites in complex environments. *Front. Physiol.* 9, 49. <https://doi.org/10.3389/fphys.2018.00049>.
- Isono, K., and Morita, H. (2010). Molecular and cellular designs of insect taste receptor system. *Front. Cell. Neurosci.* 4, 20. <https://doi.org/10.3389/fncel.2010.00020>.
- Jason Pitts, R., Derryberry, S.L., Zhang, Z., and Zwiebel, L.J. (2017). Variant ionotropic receptors in the malaria vector mosquito *Anopheles gambiae* tuned to amines and carboxylic acids. *Sci. Rep.* 7, 40297. <https://doi.org/10.1038/srep40297>.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. <https://doi.org/10.1093/bioinformatics/bts199>.
- Kosakovsky Pond, S.L., Frost, S.D.W., and Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679. <https://doi.org/10.1093/bioinformatics/bti079>.
- Kosakovsky Pond, S.L., Poon, A.F.Y., Velazquez, R., Weaver, S., Hepler, N.L., Murrell, B., Shank, S.D., Magalis, B.R., Bouvier, D., Nekrutenko, A., et al. (2020). HyPhy 2.5 - a customizable platform for evolutionary hypothesis testing using phylogenies. *Mol. Biol. Evol.* 37, 295–299. <https://doi.org/10.1093/molbev/msz197>.
- Lacey, E.S., Ray, A., and Cardé, R.T. (2014). Close encounters: contributions of carbon dioxide and human skin odour to finding and landing on a host in *Aedes aegypti*. *Physiol. Entomol.* 39, 60–68. <https://doi.org/10.1111/phen.12048>.
- Lahti, D.C., Johnson, N.A., Ajie, B.C., Otto, S.P., Hendry, A.P., Blumstein, D.T., Coss, R.G., Donohue, K., and Foster, S.A. (2009). Relaxed selection in the wild. *Trends Ecol. Evol.* 24, 487–496. <https://doi.org/10.1016/j.tree.2009.03.010>.
- Leal, W.S. (2010). Behavioural neurobiology: the treacherous scent of a human. *Nature*, 37–38. <https://doi.org/10.1038/464037a>.
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. <https://doi.org/10.48550/arXiv.1303.3997>.
- Li, Q., and Liberles, S.D. (2015). Aversion and attraction through olfaction. *Curr. Biol.* R120–R129. <https://doi.org/10.1016/j.cub.2014.11.044>.
- Lu, Y., Xing, H., and Zhang, D. (2019). Evidence for relaxed selection of mitogenome in rapid-flow cyprinids. *Genes Genomics* 41, 863–869. <https://doi.org/10.1007/s13258-019-00817-7>.
- Main, B.J., Lee, Y., Ferguson, H.M., Kreppel, K.S., Kihonda, A., Govella, N.J., Collier, T.C., Cornel, A.J., Eskin, E., Kang, E.Y., et al. (2016). The genetic basis of host preference and resting behavior in the major African malaria vector, *Anopheles arabiensis*. *PLoS Genet.* 12, e1006303. <https://doi.org/10.1371/journal.pgen.1006303>.
- Matsuo, T., Sugaya, S., Yasukawa, J., Aigaki, T., and Fuyama, Y. (2007). Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in *Drosophila sechellia*. *PLoS Biol.* 5, e118. <https://doi.org/10.1371/journal.pbio.0050118>.
- McBride, C.S. (2007). Rapid evolution of smell and taste receptor genes during host specialization in *Drosophila sechellia*. *Proc. Natl. Acad. Sci. U. S. A.* 104, 4996–5001. <https://doi.org/10.1073/pnas.0608424104>.
- McBride, C.S. (2016). Genes and odors underlying the recent evolution of mosquito preference for humans. *Curr. Biol.* R41–R46. <https://doi.org/10.1016/j.cub.2015.11.032>.
- McBride, C.S., and Arguello, J.R. (2007). Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* 177, 1395–1416. <https://doi.org/10.1534/genetics.107.078683>.
- McBride, C.S., Baier, F., Omondi, A.B., Spitzer, S.A., Lutomia, J., Sang, R., Ignell, R., and Vosshall, L.B. (2014). Evolution of mosquito preference for humans linked to an odorant receptor. *Nature* 515, 222–227. <https://doi.org/10.1038/nature13964>.
- McDonald, J.H., and Kreitman, M. (1991). Adaptive protein evolution at the Adh locus in *Drosophila*. *Nature*, 652–654. <https://doi.org/10.1038/351652a0>.
- McMeniman, C.J., Corfas, R.A., Matthews, B.J., Ritchie, S.A., and Vosshall, L.B. (2014). Multimodal integration of carbon dioxide and other sensory cues drives mosquito attraction to humans. *Cell* 156, 1060–1071. <https://doi.org/10.1016/j.cell.2013.12.044>.
- Meijerink, J., Braks, M.A.H., Brack, A.A., Adam, W., Dekker, T., Posthumus, M.A., Van Beek, T.A., and Van Loon, J.J.A. (2000). Identification of olfactory stimulants for *Anopheles gambiae* from human sweat samples. *J. Chem. Ecol.* 26, 1367–1382. <https://doi.org/10.1023/A:1005475422978>.
- Missbach, C., Dweck, H.K.M., Vogel, H., Vilcinskas, A., Stensmyr, M.C., Hansson, B.S., and Grosse-Wilde, E. (2014). Evolution of insect olfactory receptors. *Elife* 2014, e02115. <https://doi.org/10.7554/eLife.02115>.
- Murrell, B., Weaver, S., Smith, M.D., Wertheim, J.O., Murrell, S., Aylward, A., Eren, K., Pollner, T., Martin, D.P., Smith, D.M., et al. (2015). Gene-wide identification of episodic selection. *Mol. Biol. Evol.* 32, 1365–1371. <https://doi.org/10.1093/molbev/msv035>.
- Neafsey, D.E., Waterhouse, R.M., Abai, M.R., Aganezov, S.S., Alekseyev, M.A., Allen, J.E., Amon, J., Arcà, B., Arensburg, P., Artemov, G., et al. (2015). Highly evolvable malaria vectors: the genomes of 16 *Anopheles* mosquitoes. *Science*

- 347, 1258522. <https://doi.org/10.1126/science.1258522>.
- Nei, M., and Rooney, A.P. (2005). Concerted and birth-and-death evolution of multigene families. *Annu. Rev. Genet.* 121–152. <https://doi.org/10.1146/annurev.genet.39.073003.112240>.
- Nozawa, M., and Nei, M. (2007). Evolutionary dynamics of olfactory receptor genes in *Drosophila* species. *Proc. Natl. Acad. Sci. U. S. A.* 104, 7122–7127. <https://doi.org/10.1073/pnas.0702133104>.
- Omondi, A.B., Ghaninia, M., Dawit, M., Svensson, T., and Ignell, R. (2019). Age-dependent regulation of host seeking in *Anopheles coluzzii*. *Sci. Rep.* 9, 9699. <https://doi.org/10.1038/s41598-019-46220-w>.
- Pelosi, P., Calvello, M., and Ban, L. (2005). Diversity of odorant-binding proteins and chemosensory proteins in insects. *Chem. Senses* 30, i291-292. <https://doi.org/10.1093/chemse/bjh229>.
- Pelosi, P., and Maida, R. (1995). Odorant-binding proteins in insects. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 111, 503–514. [https://doi.org/10.1016/0305-0491\(95\)00019-5](https://doi.org/10.1016/0305-0491(95)00019-5).
- Powell, J.R., and Tabachnick, W.J. (2013). History of domestication and spread of *Aedes aegypti*—a review. *Mem. Inst. Oswaldo Cruz*, 11–17. <https://doi.org/10.1590/0074-0276130395>.
- R Core Team (2018). R: A Language and Environment for Statistical Computing (Vienna, Austria: R Foundation for Statistical Computing). <http://www.r-project.org/>.
- Raji, J.I., Melo, N., Castillo, J.S., Gonzalez, S., Saldana, V., Stensmyr, M.C., and DeGennaro, M. (2019). *Aedes aegypti* mosquitoes detect acidic volatiles found in human odor using the IR8a pathway. *Curr. Biol.* 29, 1253–1262.e7. <https://doi.org/10.1016/j.cub.2019.02.045>.
- Reeves, L.E., Holderman, C.J., Blosser, E.M., Gillett-Kaufman, J.L., Kawahara, A.Y., Kaufman, P.E., and Burkett-Cadena, N.D. (2018). Identification of *Uranotaenia sapphirina* as a specialist of annelids broadens known mosquito host use patterns. *Commun. Biol.* 1, 92. <https://doi.org/10.1038/s42003-018-0096-5>.
- Rinker, D.C., Zhou, X., Pitts, R.J., Rokas, A., Zwiebel, L.J., Neafsey, D.E., Berlin, A.M., Christophides, G.K., Collins, F.H., Emrich, S.J., et al. (2013). Antennal transcriptome profiles of anopheline mosquitoes reveal human host olfactory specialization in *Anopheles gambiae*. *BMC Genom.* 14, 749. <https://doi.org/10.1186/1471-2164-14-749>.
- Ritchie, S.A. (2014). Dengue vector bionomics: why *Aedes aegypti* is such a good vector. In *Dengue and Dengue Hemorrhagic Fever*, Second Edition, pp. 455–480. <https://doi.org/10.1079/9781845939649.0455>.
- Robert, L.L., and Debboun, M. (2020). Arthropods of public health importance. In *Hunter's Tropical Medicine and Emerging Infectious Diseases*, pp. 1055–1062. <https://doi.org/10.1016/b978-0-323-55512-8.00146-0>.
- Robertson, H.M., Warr, C.G., and Carlson, J.R. (2003). Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 100, 14537–14542. <https://doi.org/10.1073/pnas.2335847100>.
- Rose, N.H., Sylla, M., Badolo, A., Lutomiah, J., Ayala, D., Aribodor, O.B., Ibe, N., Akorli, J., Otoo, S., Mutebi, J.P., et al. (2020). Climate and urbanization drive mosquito preference for humans. *Curr. Biol.* 30, 3570–3579.e6. <https://doi.org/10.1016/j.cub.2020.06.092>.
- Rund, S.S.C., Hou, T.Y., Ward, S.M., Collins, F.H., and Duffield, G.E. (2011). Genome-wide profiling of diel and circadian gene expression in the malaria vector *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U. S. A.* 108, E421–E430. <https://doi.org/10.1073/pnas.1100584108>.
- Sánchez-Gracia, A., Vieira, F.G., and Rozas, J. (2009). Molecular evolution of the major chemosensory gene families in insects. *Heredity*, 208–216. <https://doi.org/10.1038/hdy.2009.55>.
- Scott, K. (2018). Gustatory processing in *Drosophila melanogaster*. *Annu. Rev. Entomol.* 15–30. <https://doi.org/10.1146/annurev-ento-020117-043331>.
- Smallegange, R.C., Qiu, Y.T., van Loon, J.A., and 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. <https://doi.org/10.1093/chemse/bji010>.
- Smith, M.D., Wertheim, J.O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S.L. (2015). Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol. Biol. Evol.* 32, 1342–1353. <https://doi.org/10.1093/molbev/msv022>.
- Soffman, A., Subandiyah, S., Makino, H., Watanabe, T., and Horiike, T. (2018). Evolutionary analysis of the highly conserved insect odorant coreceptor (Orco) revealed a positive selection mode, implying functional flexibility. *J. Insect Sci.* 18, 1–8.
- Steib, B.M., Geier, M., and Boeckh, J. (2001). The effect of lactic acid on odour-related host preference of yellow fever mosquitoes. *Chem. Senses* 26, 523–528. <https://doi.org/10.1093/chemse/26.5.523>.
- Steinbrecht, R.A. (1998). Odorant-binding proteins: expression and function. *Ann. N. Y. Acad. Sci.* 323–332. <https://doi.org/10.1111/j.1749-6632.1998.tb10591.x>.
- Suh, E., Choe, D.H., Saveer, A.M., and Zwiebel, L.J. (2016). Suboptimal larval habitats modulate oviposition of the malaria vector mosquito *Anopheles coluzzii*. *PLoS One* 11, e0149800. <https://doi.org/10.1371/journal.pone.0149800>.
- Takken, W. (1991). The role of olfaction in host-seeking of mosquitoes: a review. *Int. J. Trop. Insect Sci.* 12, 287–295. <https://doi.org/10.1017/s1742758400020816>.
- Takken, W., and Verhulst, N.O. (2013). Host preferences of blood-feeding mosquitoes. *Annu. Rev. Entomol.* 433–453. <https://doi.org/10.1146/annurev-ento-120811-153618>.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673–4680. <https://doi.org/10.1093/nar/22.22.4673>.
- Tiway, B.K. (2020). The cognitive and speech genes are jointly shaped by both positive and relaxed selection in the human lineage. *Genomics* 112, 2922–2927. <https://doi.org/10.1016/j.ygeno.2020.05.006>.
- The VEuPathDB Project Team (2021). Vectorbase: Bioinformatic resources for invertebrate vectors of human pathogens. https://vectorbase.org/vectorbase/app/record/dataset/DS_eeaaed0a32.
- Vieira, F.G., Sánchez-Gracia, A., and Rozas, J. (2007). Comparative genomic analysis of the odorant-binding protein family in 12 *Drosophila* genomes: purifying selection and birth-and-death evolution. *Genome Biol.* 8, R235. <https://doi.org/10.1186/gb-2007-8-11-r235>.
- Wang, G., Carey, A.F., Carlson, J.R., and Zwiebel, L.J. (2010). Molecular basis of odor coding in the malaria vector mosquito *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U. S. A.* 107, 4418–4423. <https://doi.org/10.1073/pnas.0913392107>.
- Weaver, S., Shank, S.D., Spielman, S.J., Li, M., Muse, S.V., and Kosakovsky Pond, S.L. (2018). Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol. Biol. Evol.* 35, 773–777. <https://doi.org/10.1093/molbev/msx335>.
- Wertheim, J.O., Murrell, B., Smith, M.D., Kosakovsky Pond, S.L., and Scheffler, K. (2015). RELAX: detecting relaxed selection in a phylogenetic framework. *Mol. Biol. Evol.* 32, 820–832. <https://doi.org/10.1093/molbev/msu400>.
- White, B.J., Collins, F.H., and Besansky, N.J. (2011). Evolution of *Anopheles gambiae* in relation to humans and malaria. *Annu. Rev. Ecol. Syst.* 42, 111–132. <https://doi.org/10.1146/annurev-ecolsys-102710-145028>.
- Wickler, S., and Spriggs, M. (1988). Pleistocene human occupation of the Solomon Islands, Melanesia. *Antiquity* 62, 703–706. <https://doi.org/10.1017/S0003598X00075104>.
- Wund, M.A., Baker, J.A., Golub, J.L., and Foster, S.A. (2015). The evolution of antipredator behaviour following relaxed and reversed selection in Alaskan threespine stickleback fish. *Anim. Behav.* 106, 181–189. <https://doi.org/10.1016/j.anbehav.2015.05.009>.
- Xia, Y., Wang, G., Buscariollo, D., Pitts, R.J., Wenger, H., and Zwiebel, L.J. (2008). The molecular and cellular basis of olfactory-driven behavior in *Anopheles gambiae* larvae. *Proc. Natl. Acad. Sci. U. S. A.* 105, 6433–6438. <https://doi.org/10.1073/pnas.0801007105>.

Xu, P., Choo, Y.M., Chen, Z., Zeng, F., Tan, K., Chen, T.Y., Cornel, A.J., Liu, N., and Leal, W.S. (2019). Odorant inhibition in mosquito olfaction. *iScience* 19, 25–38. <https://doi.org/10.1016/j.isci.2019.07.008>.

Yan, H., Jafari, S., Pask, G., Zhou, X., Reinberg, D., and Desplan, C. (2020). Evolution, developmental

expression and function of odorant receptors in insects. *J. Exp. Biol.* 223, jeb208215. <https://doi.org/10.1242/jeb.208215>.

Yassin, A., Debat, V., Bastide, H., Gidaszewski, N., David, J.R., and Pool, J.E. (2016). Recurrent specialization on a toxic fruit in an island *Drosophila* population. *Proc. Natl. Acad.*

Sci. U. S. A. 113, 4771–4776. <https://doi.org/10.1073/pnas.1522559113>.

Zwiebel, L.J., and Takken, W. (2004). Olfactory regulation of mosquito-host interactions. In *Insect Biochemistry and Molecular Biology*, pp. 645–652. <https://doi.org/10.1016/j.ibmb.2004.03.017>.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, peptides, and recombinant proteins		
Illumina TruSeq DNA Nano kit	Illumina	20015964
Deposited data		
<i>Anopheles farauti</i> reference genome	https://vectorbase.org/vectorbase/app/record/dataset/TMPTX_afarFAR1	AfarF2.7
Data and code used in this paper	(https://data.mendeley.com/datasets/4krvx9z86/draft?a=a5c5bfed-d5e5-4ee0-87ed-dc93efd6ad1a)	n/a
Software and algorithms		
FastQC	Babraham Bioinformatics	https://www.bioinformatics.babraham.ac.uk/projects/fastqc/
Geneious	Kearse et al., 2012	https://www.geneious.com/
BWA mem v0.7	Li, 2013	https://arxiv.org/abs/1303.3997
CLUSTALW	Thompson et al., 1994	https://academic.oup.com/nar/article-abstract/22/22/4673/2400290?login=false
FreeBayes	Garrison and Marth 2012	https://arxiv.org/abs/1207.3907v2
VCFtools	Danecek et al., 2011	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3137218/
tBLASTn	Gerts et al., 2006	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1779365/
GeneWise	Birney et al., 2004	https://www.ebi.ac.uk/Tools/psa/genewise/
R	R Core Team, 2018	https://www.R-project.org/
Seqinr (R package)	Charif et al., 2015	https://cran.r-project.org/web/packages/seqinr/index.html
HyPhy package	Kosakovsky Pond et al., 2020	https://pubmed.ncbi.nlm.nih.gov/31504749/
DataMonkey Webserver	Weaver et al., 2018	https://pubmed.ncbi.nlm.nih.gov/29301006/
aBSREL	Smith et al., 2015	https://pubmed.ncbi.nlm.nih.gov/25697341/
BUSTED	Murrell et al., 2015	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4408417/
RELAX	Wertheim et al., 2015	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4327161/

RESOURCE AVAILABILITY

Lead contact

Enquiries, requests, and comments related to this paper should be directed to lead contact, Luke Ambrose (l.ambrose@uq.edu.au or lukeambrose3@gmail.com).

Materials availability

This study did not generate new unique reagents or materials.

Data and code availability

Raw data and code used in this paper can be downloaded from Mendeley Data:

[Anopheles farauti complex olfactory gene alignments and whole genome SNP data - Mendeley Data \(doi: 10.17632/4krvx9z86.1\)](https://doi.org/10.17632/4krvx9z86.1).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All samples used in this study were wild caught organisms. See [introduction](#) and [method details](#), as well as [Table 1](#), for information on the study system and individuals used.

METHOD DETAILS

Whole genome sequencing, mapping, variant calling and phylogenies

Nine individuals from closely related species and populations of the *An. farauti* complex were selected for genomic sequencing, as well as two individuals as outgroups, *An. punctulatus* and *An. koliensis*. These species and populations were chosen based on a combination of two criteria: their host preferences concerning feeding on humans (or that of their species/population), and their phylogenetic and/or population genetic relationships. These traits and relationships are outlined in [Table 1](#), and population relationships within *An. hinesorum* are outlined in more detail in [Ambrose et al., 2021 \(Ambrose et al., 2021\)](#). Genomic DNA of the nine individuals selected was extracted from whole mosquito samples and sent to Macrogen Inc. (South Korea), who generated paired-end libraries using an Illumina TruSeq DNA Nano kit. Sequencing was performed on Illumina HiSeq 2500 and 2000 machines, with 250bp paired-end (PE) reads. This system provides a unique contrast to well-studied systems (*An. gambiae* and *Ae. aegypti*), where species have evolved from zoophilic generalists into anthropophilic specialists ([White et al., 2011](#); [Rose et al., 2020](#)). Like *An. gambiae*, mosquitoes in this system have compact genomes, being 180-220 Mbp in size ([Neafsey et al., 2015](#)). Finally, a reference genome is available for one of the species in the group, *An. farauti* ([Neafsey et al., 2015](#); [The VEuPathDB Project Team, 2021](#)).

The assembled *An. farauti* genome Afar2 (with repetitive elements masked) was downloaded from VectorBase in August 2016. Raw read quality was assessed using FastQC ([Andrews, 2010](#)). Details of the samples sequenced, including species and population, host preference, library type and coverage can be found in [Table 1](#). We mapped all reads to the *An. farauti* reference genome, assembly Afar2, using BWA mem v0.7 ([Li, 2013](#)).

To assess the overall phylogenetic relationships between the samples sequenced, we compared phylogenies for both nuclear and whole mitochondrial genomes ([Figure 1](#)). Two additional individuals belonging to different species were also included in this analysis (*An. koliensis* (another outgroup) and *An. farauti* 8). These additional samples were included to reconfirm previously established species relationships in the complex, and to provide an additional sample that is closely related to *An. irenicus* for mtDNA, as *An. farauti* from Queensland was previously found to contain introgressed mitochondria from *An. hinesorum*. The eleven mitochondrial genomes were constructed by mapping raw sequencing data to a whole mitochondrial genome sequence from *An. gambiae* in Geneious v.8.1.8 ([Kearse et al., 2012](#)). Mitochondrial genomes were aligned in Geneious v.8.1 using CLUSTALW ([Thompson et al., 1994](#)) and this alignment was used to construct a neighbor-joining phylogeny using the HKY substitution model with 1000 bootstrap replicates ([Figure 1](#)), also performed in Geneious v.8.1. *Anopheles koliensis* was used as an outgroup in this analysis, based on previous knowledge of phylogenetic relationships between species in the group.

We identified variants using FreeBayes haplotype-variant detector ([Garrison and Marth, 2012](#)) and filtered single nucleotide polymorphisms (SNPs) to mapping quality Q30 using VCFtools ([Danecek et al., 2011](#)), then to a minimum depth of 3 and maximum depth of 30. Following this, we removed any marker that was missing more than 30% of the data, and then removed any locus with a minor allele count below three (e.g., one homozygote and one heterozygote). All indels and non-bi-allelic SNPs were removed, leaving only bi-allelic SNPs and thinned SNPs to a maximum of one every 1000 bp. The remaining 164 041 high-quality variants were used to construct a neighbor-joining tree using Geneious.

Characterization and genomic location of olfactory genes in the *An. farauti* complex

We manually isolated and characterized gene orthologues from four olfactory gene families (*Ors*, *Grs*, *Irs* and *Obps*), in each of the nine genomes sequenced. A local database of the *An. farauti* reference genome was created in Geneious v.8.1, which was then queried with local tBLASTn ([Gerts et al., 2006](#)) searches of known amino acid sequences from *An. gambiae*. For these searches, a cut-off value of 1e-10 was used. This allowed us to locate most genes queried in the *An. farauti* reference genome.

Once gene orthologues were isolated from the *An. farauti* reference genome, the online gene prediction tool, GeneWise ([Birney et al., 2004](#)), was used to predict the coding sequences of these genes. We used the *An. gambiae* amino acid sequence of each gene to predict the coding sequence in the *An. farauti* reference genome. This was done for each individual by using the top basic local alignment search tool (BLAST) hit, and the surrounding region as a nucleotide sequence to query against each reference amino acid

sequence. Default prediction settings were used in GeneWise searches, apart from using modelled splice sites as opposed to the simpler GC/AT based splicing mode in most cases.

In cases where an orthologous sequence could be identified in the *An. farauti* genome, the location of the gene on the reference scaffolds was identified and the alignments of mapped reads for these regions were extracted from each genome. The consensus sequence was then used in a second GeneWise query with the *An. farauti* amino acid sequence used to predict the orthologous coding sequences in each of the other nine genomes. The genomic location of each olfactory gene was identified in both the *An. gambiae* and *An. farauti* genomes. From this, chromosomal synteny of olfactory genes was assessed between *An. gambiae* and *An. farauti*, for the 10 largest *An. farauti* scaffolds (Tables S1–S4 and Figure S2). Some BLAST queries resulted in top BLAST hits in the same location of the *An. farauti* genome, possibly due to gene loss or gain between *An. farauti* and *An. gambiae*. In these cases, the queried gene with the best BLAST score was used to predict the corresponding *An. farauti* gene orthologue. Coding sequences of each olfactory gene were aligned in Geneious v.8.1, initially using CLUSTALW, with manual adjustment if necessary. These alignments were translated into amino acid sequences and checked for the presence of stop codons. To assess levels of gene conservation, the percentage of identical sites was calculated in Geneious v.8.1, for both nucleotide and amino acid alignments (Figure S1, Tables S1–S4). Differences between gene families for these metrics was assessed in R (R Core Team, 2018) using ANOVA. Both nucleotide and amino acid alignments were exported for further analyses.

QUANTIFICATION AND STATISTICAL ANALYSIS

Candidate genes involved in human host detection – Evolutionary patterns

To identify candidate genes that may be involved in human feeding, we initially explored qualitative evolutionary patterns, including 1) fixed amino acid differences, 2) different intron-exon structure between anthropophilic and zoophilic lineages, 3) higher rates of non-synonymous to synonymous nucleotide substitutions (kA/kS) in different versus same phenotype comparisons, and 4) phylogenetic relationships (see Figure 3). Any olfactory genes where phylogeny reflects host-preference phenotype as opposed to known species relationships would be the strongest candidates for being involved in human host preference (Figure 3A). This is followed by olfactory genes for which the two zoophilic *An. hinesorum* individuals are sister and the anthropophilic *An. hinesorum* individual from the Solomon Archipelago is closely related to other anthropophilic *An. hinesorum* lineages (Figure 3B). We expect that for most non-candidate olfactory genes, the three individuals of *An. hinesorum* from the Solomon Archipelago will form a monophyletic clade in phylogenies (see Figures 3C and 3D), reflecting known ‘average’ nuclear genetic relationships.

Substitution rate variation for each gene was assessed by taking the ratio of the rate of non-synonymous (kA) to synonymous (kS) mutations (kA/kS). This was calculated pairwise for each gene between all individuals using the *kaks* function in the R package *seqinr* (Charif et al., 2015). Each pairwise comparison was also categorized into one of three phenotype comparison classes: anthropophilic/zoophilic, anthropophilic/anthropophilic and zoophilic/zoophilic which were used in analyses outlined below. We calculated overall mean and median kA/kS for each gene family. Because the data is not normally distributed, we performed non-parametric (Kruskal-Wallis) tests to assess significant differences between gene families. We performed further tests to assess whether there are significant differences in kA/kS ratios between same and different phenotype comparisons within gene family. Again, we used non-parametric Kruskal-Wallis tests to assess the significance of differences in these tests. We also produced box plots for kA/kS for each olfactory gene family by same and different phenotype comparison. These provide a useful way to visualize the mean, median, interquartile range, and points with values over one. We used a one-sided two-sample test for equality of proportions implemented in R, to test whether there are significantly more kA/kS values over one in each gene family for anthropophilic/zoophilic comparisons than in anthropophilic/anthropophilic comparisons.

Candidate genes involved in human host detection – Tests of selection

We used the HyPhy package (Kosakovsky Pond, Frost and Muse, 2005) implemented in the web server, Data Monkey (Weaver et al., 2018), to assess evidence of selection on branches of interest (those leading to zoophilic lineages). Prior to running selection analyses, we checked each gene alignment for the effects of intragenic recombination using the single break point method implemented in HyPhy. The GARDprocessor.bf module was used to assess whether significant breakpoints were due to topological incongruence (rather than rate variation), using the Shimodaira-Hasegawa test ($p \leq 0.01$). We used the branch

site REL method (aBSREL) ([Smith et al., 2015](#)) to detect episodic diversifying selection on branches of interest; BUSTED ([Murrell et al., 2015](#)) to detect evidence of gene-wide selection on zoophilic branches; and RELAX ([Wertheim et al., 2015](#)) to detect genes for which zoophilic lineages may have experienced relaxed or intensified purifying or positive selection in relation to other branches. For BUSTED, aBSREL, all zoophilic individuals were selected as test branches for each analysis.