HHS Public Access

Author manuscript

Nature. Author manuscript; available in PMC 2015 May 13.

Published in final edited form as:

Nature. 2014 November 13; 515(7526): 222–227. doi:10.1038/nature13964.

Evolution of mosquito preference for humans linked to an odorant receptor

Carolyn S. McBride[1,2,†], **Felix Baier**[1,†], **Aman B. Omondi**[3,†], **Sarabeth A. Spitzer**[1,†], **Joel Lutomiah**[4], **Rosemary Sang**[4], **Rickard Ignell**[3], and **Leslie B. Vosshall**[1,2]

[1]Laboratory of Neurogenetics and Behavior, The Rockefeller University, New York, NY 10065 USA

[2]Howard Hughes Medical Institute

[3]Unit of Chemical Ecology, Department of Plant Protection Biology, Swedish University of Agricultural Sciences, Box 102, Sundsvägen 14, 230 53, Alnarp, Sweden

[4]Center for Virus Research, Kenya Medical Research Institute, P.O. Box 54840 - 00200, Off Mbagathi Way, Nairobi, Kenya

Abstract

Female mosquitoes are major vectors of human disease and the most dangerous are those that preferentially bite humans. A ‘domestic’ form of the mosquito *Aedes aegypti* has evolved to specialize in biting humans and is the major worldwide vector of dengue, yellow fever, and Chikungunya viruses. The domestic form coexists with an ancestral, animal-biting ‘forest’ form along the coast of Kenya. We collected the two forms, established laboratory colonies, and document striking divergence in preference for human *versus* animal odour. We further show that the evolution of preference for human odour in domestic mosquitoes is tightly linked to increases in the expression and ligand-sensitivity of the odorant receptor *AaegOr4*, which we found recognises a compound present at high levels in human odour. Our results provide a rare example

Correspondence and requests for materials should be addressed to L.B.V. (Leslie.Vosshall@rockefeller.edu).

†Present address: Princeton Neuroscience Institute and Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544 USA (C.S.M.); Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138 USA (F.B.); Plant Protection Research Institute, Agricultural Research Council, Private Bag X134, Queenswood 0121, South Africa (A.B.O.); Harvard College, Harvard University, Cambridge, MA 02138 (S.A.S.).

Author Contributions C.S.M. and L.B.V. conceived the study. C.S.M. participated in the execution and analysis of all aspects of the study. J.L. helped coordinate mosquito collection in Rabai, Kenya under the supervision of R.S. S.A.S. helped design and carry out the morphological assays presented in Fig. 1e-i. F.B. helped clone, analyse, and genotype mosquitoes for the *Or4* alleles presented in Fig. 5a-d, and construct transgenic *Drosophila* lines for use in single sensillum recordings. A.B.O. and R.I. designed, conducted, and analysed the GC-SSR and GC-MS experiments presented in Fig. 4 and carried out pilot experiments comprising dose-response curves and spontaneous activity analysis of alleles A and E similar to those presented in Fig. 5e-g. C.S.M. and L.B.V. designed all other experiments, interpreted the results, designed the figures, and wrote the paper.

Online Content Methods, Extended Data Figures 1-5, and Supplementary Table 1 are available in the online version of the paper; references unique to these sections appear only in the online paper.

Supplementary Information is available in the online version of the paper.

The authors declare no competing financial interests.

Raw RNAseq data are available for download at the NCBI Sequence Read Archive (accession number SRP035216). Coding sequences of *AaegOr4* alleles are at GenBank (accession numbers KF801614-15 and KF801617-21).

of a gene contributing to behavioural evolution and provide insight into how disease-vectoring mosquitoes came to specialise on humans.

INTRODUCTION

Blood-feeding as a behavioural adaptation is exceedingly rare in insects. Of the 1-10 million insect species on earth, only ~10,000 feed on the blood of live animals[1]. Among these, only about 100 species blood-feed preferentially on humans[1]. When biting insects evolve to prefer humans, they can spread diseases such as malaria and dengue fever with devastating efficiency. The mosquito *Aedes aegypti* provides one of the best examples of specialization on humans. It originated as a wild, animal-biting species in the forested areas of sub-Saharan Africa, where the subspecies *Ae. aegypti formosus* is still often found living in forests and biting non-human animals today[2-4]. The derived non-African subspecies *Ae. aegypti aegypti*, in contrast, has evolved to specialise in biting humans and thus has become the major worldwide vector of dengue and yellow fevers[2-4].

The evolutionary adaptations that help subspecies *aegypti* exploit humans are most clearly seen where it has been reintroduced along the coast of East Africa and is known as the 'domestic' form. Researchers investigating the outbreak of an unknown illness in Tanganyika in 1952 discovered homes heavily populated by brown-pigmented 'domestic' mosquitoes[5]. Subsequent work in the Rabai region of Kenya in the 1960s and 1970s showed that domestic mosquitoes readily entered homes[6], preferred to lay eggs in nutrient-poor river and rain water stored in containers indoors[7,8], were resistant to starvation as larvae[9], and had evolved a strong preference for biting humans[7,10,11]. Black-pigmented populations of the native African subspecies *formosus*, known in Rabai as the 'forest' form, were found just hundreds of meters away avoiding homes, laying their eggs in tree holes and rock pools outdoors, and preferring to bite non-human animals. These differences translated into marked divergence in capacity to spread human diseases, including Chikungunya (the unknown illness from 1952), yellow fever, prevalent in Africa and South America since the 16th century, and dengue fever, currently infecting almost 400 million people around the world each year[12].

Remarkably, the domestic and forest forms in Rabai remained distinct in nature but were interfertile in captivity[8], providing a rare opportunity to investigate the genetic basis and evolution of traits that adapt mosquitoes to humans. Here we find that human host preference in domestic mosquitoes is strongly correlated with functional genetic variation in an odorant receptor, *Or4*, which recognises a component of human body odour.

Domestic and forest mosquitoes continue to coexist in Rabai, Kenya

Forest and domestic mosquitoes were last documented in Rabai, Kenya in the 1970s, and we returned there in 2009 to determine whether they still exist. We collected *Ae. aegypti* larvae and pupae in water-storage containers inside approximately one in every five homes visited (Fig. 1a-b). We also collected eggs, larvae, and pupae of several mosquito species, including *Ae. aegypti*, outdoors in natural and artificial containers in villages and nearby forest (Fig.

1a, c). From these collections, we established 29 laboratory colonies, each descending from fewer than 20 males and females collected in the same house or outdoor location (Fig.1a).

Previous reports described differences in body colour between the forms[3,6]. Indeed, females from all outdoor colonies and some indoor colonies were black, resembling forest mosquitoes. Those from the remaining indoor colonies were brown, resembling domestic mosquitoes (Fig. 1d). Differences in thorax colour were maintained across multiple laboratory generations (Fig. 1e-h; Extended Data Fig. 1). Black and brown colonies also differed in abdominal scaling (Fig. 1i). A single indoor colony, K14, showed a mix of black and brown morphologies (Fig. 1e-i). Black mosquitoes resembled a subspecies *formosus* colony from inland Africa (Uganda), while brown mosquitoes resembled a subspecies *aegypti* colony from Asia (Thailand) (Fig. 1e-i). In light of these morphological differences, and genetic differences among the field-collected progenitors of our colonies[13], we hereafter refer to black and brown colonies as forest and domestic, respectively. In summary, mosquitoes fitting the morphological description of the two forms continue to coexist in Rabai, Kenya, 35 years after they were last documented.

Domestic mosquitoes have evolved strong preference for humans

We used three assays to characterise the preference of forest and domestic forms for humans *versus* non-human animals. We offered guinea pig as a non-human host because it is among the diverse hosts to which forest mosquitoes respond[10,11]. In a biting assay where females are exposed directly to live hosts (Fig. 2a), forest females preferred the guinea pig (Fig. 2b) and domestic females weakly preferred the human (Fig. 2b). Domestic females were approximately twice as likely to respond overall (Fig. 2c), possibly reflecting adaptation to indoor environments, and by extension laboratory settings[14]. When host cues were presented in an olfactometer (Fig. 2d), forest and domestic females again showed significantly different preferences and response rates (Fig. 2e, f), with domestic females displaying a strong preference for humans. All colonies fell into two discrete behavioural clusters corresponding precisely to the forest and domestic designations that we made on the basis of morphology (Fig. 2g). Behaviourally, forest colonies resembled subspecies *formosus* from Uganda, and domestic colonies resembled subspecies *aegypti* from Thailand.

We further confirmed these results with host-scented nylon sleeves supplemented with equal amounts of carbon dioxide (CO_2), a potent activator of mosquito host-seeking[15] (Fig. 2h). Three domestic colonies retained their strong preference for humans, while three forest colonies ranged from no preference to moderate preference for guinea pig (Fig. 2h).

To extend our results to other non-human hosts, we also assessed the preference of a small subset of colonies for human *versus* chicken in the live host assay, obtaining qualitatively similar results (Fig. 2i). Our findings confirm that domestic mosquitoes have evolved a marked preference for human body odour.

Human preference is associated with changes in *OR* expression

Novel chemosensory preferences in insects are sometimes accompanied by changes in the peripheral chemosensory system[16-20]. We reasoned that altered gene expression in antennae

may contribute to preference, and profiled differential gene expression in this major olfactory organ using RNAseq. To identify general differences between forms, we compared forest *versus* domestic colonies (Fig. 2h). To determine which of these differences are genetically associated with host preference[21], we crossed two representative colonies and compared pools of strongly human- *versus* guinea pig-preferring F2 hybrids (Fig. 3a-b).

959 antennal genes were differentially expressed in colonies (Fig. 3c, e), 46 genes were differentially expressed in F2 pools (Fig. 3d-e), and 14 genes were differentially expressed in the same direction in both comparisons (Fig. 3e and Supplementary Table 1). Odorant receptors (*ORs*), a family of insect chemosensory receptors[22] were dramatically overrepresented among differentially expressed genes (Fig. 3f, $P<0.0001$). Two other families of chemosensory genes, the ionotropic receptors (*IRs*)[23] and odorant-binding proteins (*OBPs*)[24] were less enriched or not enriched, respectively (Fig. 3f). A selective role for the OR pathway in helping mosquitoes distinguish among hosts is consistent with previous work in a laboratory strain of *Ae. aegypti*[25].

Of the 14 genes significant in both colony and F2 comparisons, two were *ORs* (Fig. 3f), *Or4* and *Or103*. Both were upregulated in human-preferring mosquitoes (Fig. 3g, solid lines). *Or4* was also the second most highly expressed ligand-selective *OR* in the antennae of domestic females overall (Fig. 3g), and we chose this gene for further study.

Or4 recognises sulcatone, a component of human body odour

An olfactory receptor could modulate host preference by mediating attraction or repulsion to specific host odours, so we asked whether Or4 is activated by a component of human or guinea pig odour. We expressed the genome reference allele of *Or4* heterologously in a *Drosophila* olfactory neuron lacking a ligand-selective OR[26], and tested responses to fractionated host odour from guinea pigs and humans (Fig. 4a). Or4 did not respond to any fraction of guinea pig odour (data not shown), but responded consistently to a fraction of human odour corresponding to 6-methyl-5-hepten-2-one, commonly called sulcatone (Fig. 4b).

Sulcatone is a volatile odorant repeatedly identified in human body odour[27-30]. While sulcatone is also emitted by a variety of other animals[31-34] and plants[35-37], it appears to reach uniquely high levels in humans (Fig. 4c-e). It was abundant in the odour of nylon sleeves worn by five humans but undetectable or at low concentration in unworn sleeves and sleeves worn by four guinea pigs (Fig. 4 c-d). For reference, another widespread volatile, benzaldehyde, did not differ significantly between samples (Fig. 4c-d). We also found approximately four times more sulcatone in the body odour of live humans than in the odour of a live chicken or the hair of horses, cows, and sheep (Fig. 4e). Collectively, these results suggest that increased expression of *Or4* may help mosquitoes distinguish humans from non-human animals by conferring sensitivity to sulcatone. Interestingly, the malaria mosquito *Anopheles gambiae* has at least 4 *ORs* that are strongly activated by sulcatone[38], but none is closely related to *AaegOr4* (ref. [39]). These two species diverged from each other 150 million years ago and evolved to specialize in biting humans independently[40].

Preference for humans is tightly linked to increases in *Or4* ligand-sensitivity and expression

Evolution of preference for human hosts could occur not only via changes in *Or4* expression, but also via changes in the *Or4* coding region that affect protein function. Natural variation in chemoreceptor proteins has previously been shown to alter ligand-sensitivity[41] and odour perception[42]. We found extensive variation in *Or4*, with seven major alleles present in the two parent colonies and F2 hybrids (Fig. 5a-c). The domestic parent, K14, was dominated by the closely related A and B alleles (Fig. 5a-b) and a highly divergent G allele (Fig. 5b). The K27 forest parent, in contrast, harboured 5 distinct alleles at low to moderate frequency (Fig. 5b). RNAseq data from 8 additional colonies suggest that these patterns apply globally. Human-preferring colonies derived from Kenya, Thailand, USA, and West Africa were dominated by Alike alleles, while animal-preferring colonies from Kenya and Uganda were all highly variable (Extended Data Fig. 2). Although all alleles were present in F2 mosquitoes, they were inherited at different frequencies by human- and guinea pig-preferring individuals (Fig. 5c-d). Moreover, F2s tended to carry alleles characteristic of the parent with similar preference (compare Fig. 5b and d), suggesting that some aspect of allele-specific function affects preference.

We next asked how allelic variation affects Or4 receptor function. The protein-coding sequence of *Or4* is remarkably variable among alleles with differences in 13 of 406 residues on average, and 26 residues in the most extreme case (Fig. 5a; Extended Data Fig. 3). Given this high level of variation, we conducted a molecular analysis to confirm that all alleles correspond to a single copy gene (Extended Data Fig. 4). We tested the function of each allele in *Drosophila* and found that A, B, C, F, and G were highly sensitive to sulcatone, while D and E were much less sensitive (Fig. 5e-f). Variation in spontaneous activity mirrored the variation in odour-evoked activity (Fig. 5g).

We also asked whether *Or4* alleles vary in expression and could thus help explain the upregulation of this locus in human-preferring mosquitoes. We reanalysed the RNAseq data, parsing gene expression according to major alleles in each F2 pool and normalising by allele frequency to isolate levels of allele-specific expression. Two major conclusions emerged. First, all alleles are expressed at higher levels when carried by human-preferring F2s than when carried by guinea pig-preferring F2s (Fig. 5h, compare red to blue). This difference suggests that a genetic element unlinked from *Or4* contributes to upregulation in human-preferring mosquitoes. Second, the rank order of gene expression of each allele was preserved, regardless of whether it was carried by human- or guinea pig-preferring mosquitoes (Fig. 5h, compare alleles within red or blue). For example, B is always expressed at the highest level and F at the lowest level. These consistent and significant differences among alleles suggest that genetic elements linked to *Or4* and varying among alleles also contribute to changes in expression.

To determine which characteristics of *Or4* alleles are tied to behaviour, we asked whether ligand-sensitivity and/or expression can account for host preference-based differential inheritance of alleles (Fig. 5d). Remarkably, both factors had significant effects on relative allele frequency in F2s and together explained 92% of the variation (Fig. 5i). This striking

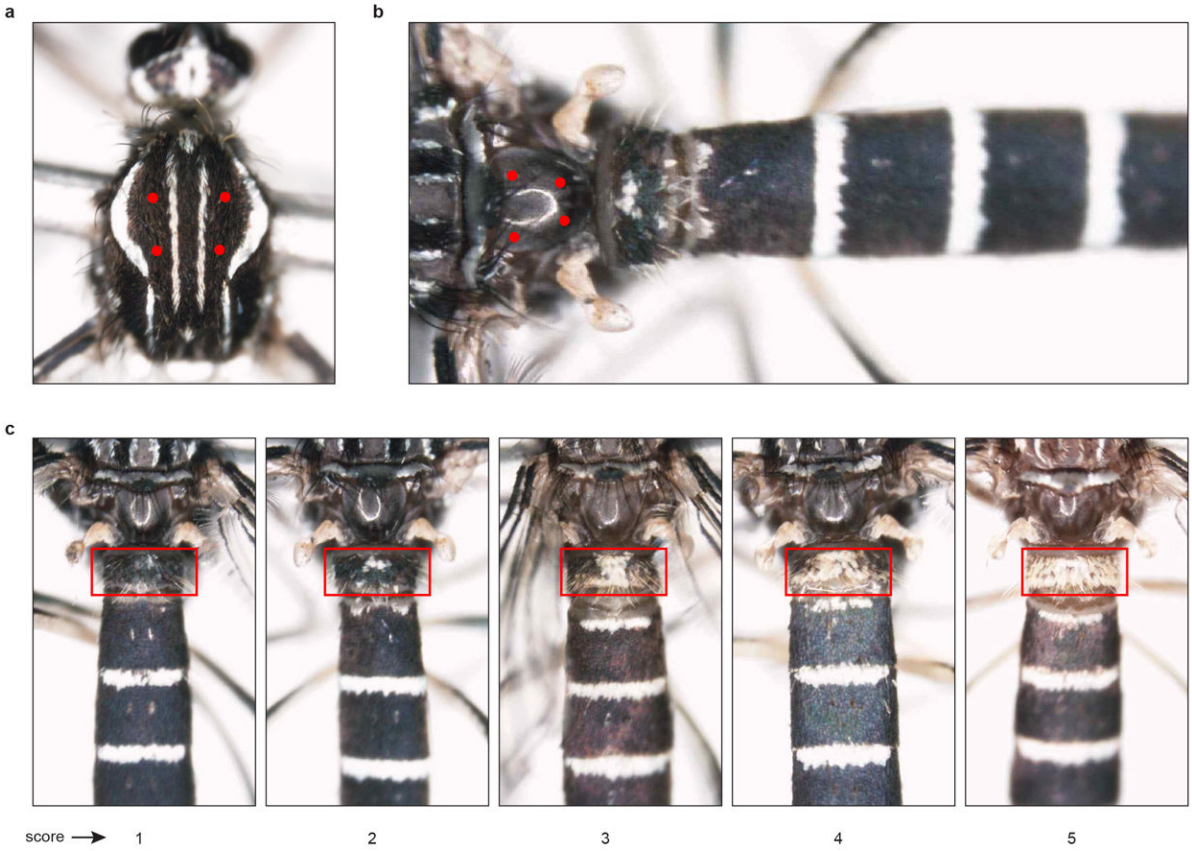
relationship suggests that expression and sensitivity to sulcatone have independent and additive effects on preference. It also suggests that high levels of both are required for strong human preference. For example, E is one of the most highly expressed alleles, yet it has weak sensitivity to sulcatone and is biased towards guinea pig-preferring mosquitoes (Fig. 5i). Conversely, allele F has high sensitivity to sulcatone, yet it is expressed at extremely low levels and is also slightly biased towards guinea pig-preferring mosquitoes (Fig. 5i).

Discussion

We have re-established the Rabai forest and domestic mosquito study system for investigation of the striking evolutionary adaptations that help domestic *Ae. aegypti* females find, bite, and thereby spread disease to humans. We show that preference for humans is tightly linked to increases in both the expression and ligand-sensitivity of odorant receptor *Or4*. These changes may help mosquitoes distinguish humans from non-human animals by increasing behavioural sensitivity to the signature human odorant sulcatone. Interestingly, sulcatone has been described as a mosquito repellent when added to human odour at certain concentrations[43-45] and sometimes as an attractant when added at low concentrations or delivered alone[29,45]. This raises the intriguing possibility that while a baseline level of sulcatone signals humanness, mosquitoes may prefer humans that have lower levels of sulcatone over those with high levels. In other words, as is true for many odours, it is possible to have too much of a good thing. We further note that sulcatone is unlikely to be the only odorant that makes us smell human, nor *Or4* the only gene contributing to human preference in domestic *Ae. aegypti*. Interestingly, guinea pig odour perfumed with sulcatone was not preferred over the odour of guinea pig alone by human-preferring mosquitoes (Extended Data Fig. 5). We strongly suspect that evolutionary changes at other loci, including some of the other hits from our antennal RNAseq analysis, may also play a role.

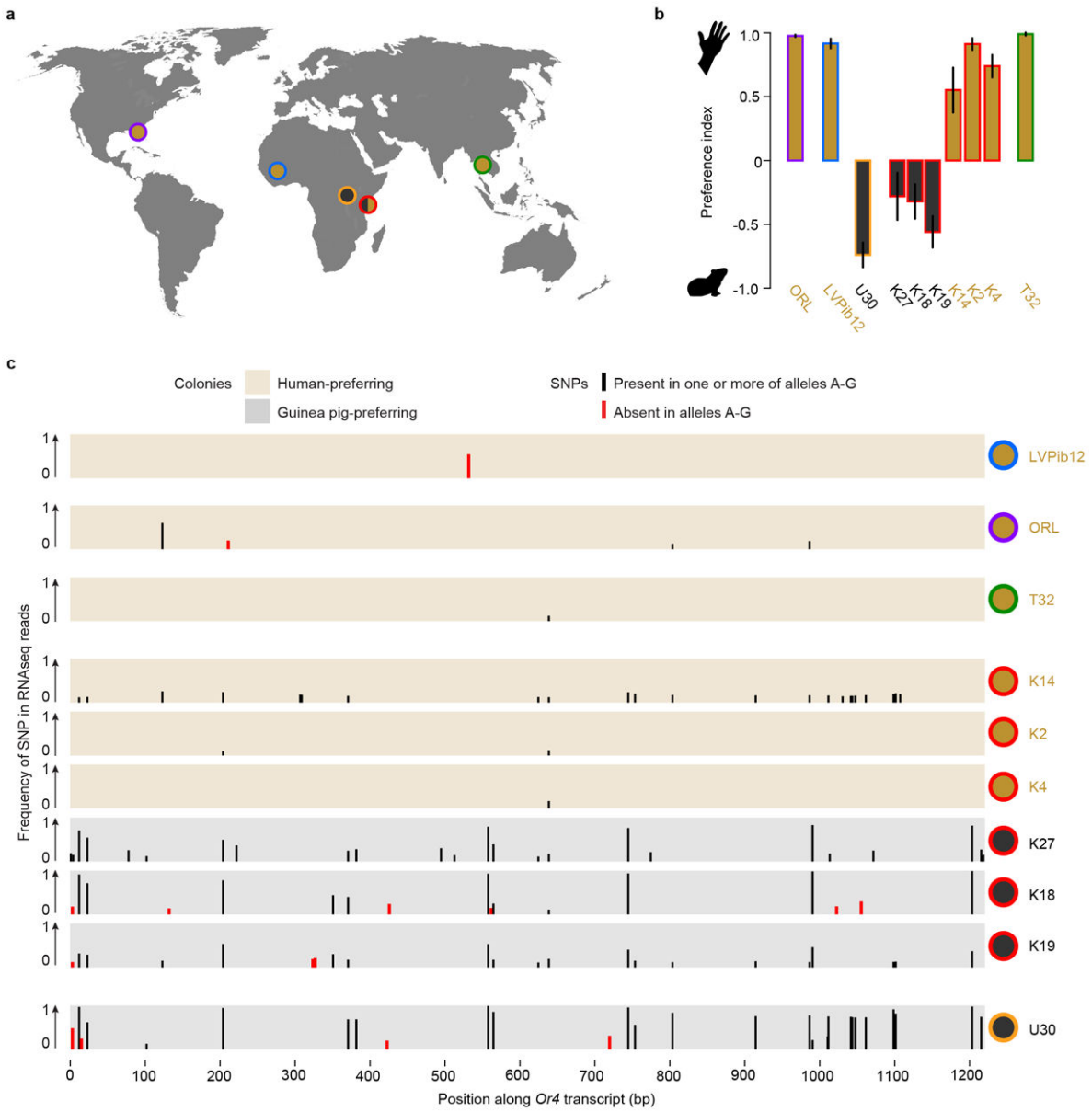
Our results also provide insight into the molecular basis of behavioural evolution. Despite exciting progress in this area[46,47], examples of specific genes associated with behavioural change are extremely rare[48]. Previous authors documented changes in the peripheral olfactory system of organisms with novel host preference[16,18,19], but direct links between these changes and behaviour have been missing. We have established a clear genetic association between such changes and behaviour. This work begins to unravel the molecular genetic basis of an important evolutionary shift in insect host preference. More generally, such host shifts not only impact the efficiency of mosquitoes as vectors of infectious disease, but contribute to the economic damage caused by agricultural pests[49] and play a key role in the formation of new species[50].

Extended Data



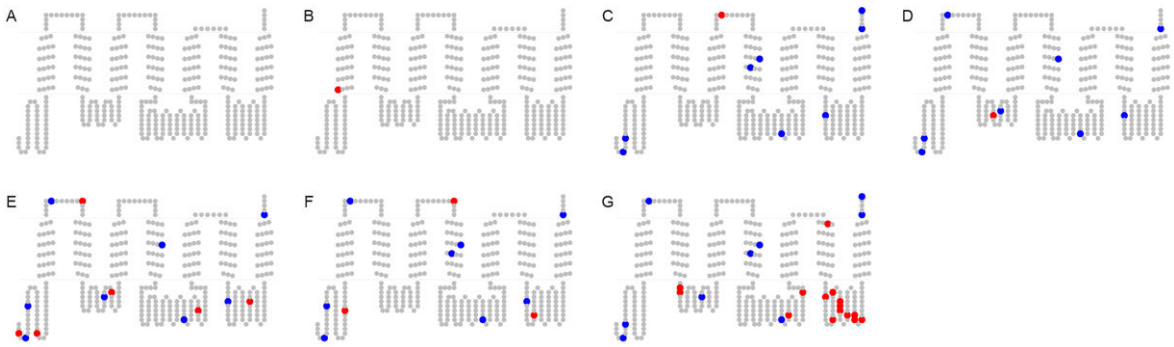
Extended Data Figure 1. Measuring colour and scaling of adult female Ae. aegypti mosquitoes

a, Representative photograph used to measure scale colour (Fig. 1e, g). Red dots mark the approximate position of 4 points where the colour of dark scales on the scutum was assessed. b, Representative photograph used to measure cuticle colour (Fig. 1f,h). Red dots mark the approximate position of 4 points where the colour of bare cuticle on the circular postnotum was assessed. c, Representative photographs used to assess the extent of white scaling on the first abdominal tergite (Fig. 1i), outlined with the red rectangle. Each individual is representative of the scaling score shown at the bottom.



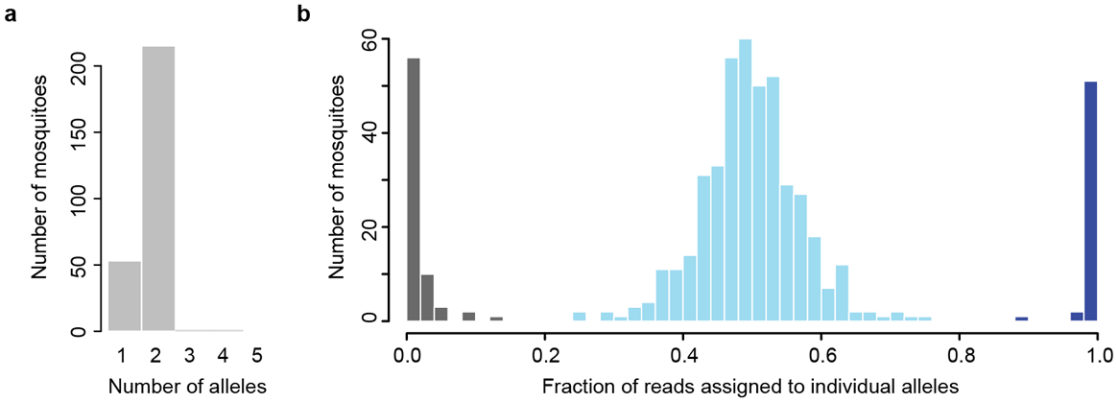
Extended Data Figure 2. Or4 coding sequence variation in human- and guinea pig-preferring colonies from around the world

a, Geographical origin of colonies characterized in b and c. Circle fill colour indicates preference of strains. Circle outline colour indicates origin: Purple–lab strain derived from USA, blue–reference genome strain derived from West Africa, orange–Uganda, red–Kenya, green–Thailand. b, Host preference assayed in the live host olfactometer. Data for Thailand, K14, K2, K4, K27, K18, K19, and Uganda are reprinted from Fig. 2e. c, Frequency of non-synonymous single nucleotide polymorphisms (SNPs) in female antennal RNAseq reads. SNPs are defined as differences from the A reference allele. SNPs with frequency ≤ 0.1 are not shown. Vertical black and red lines indicate SNPs that were present and absent, respectively, in the major alleles subject to functional analysis.



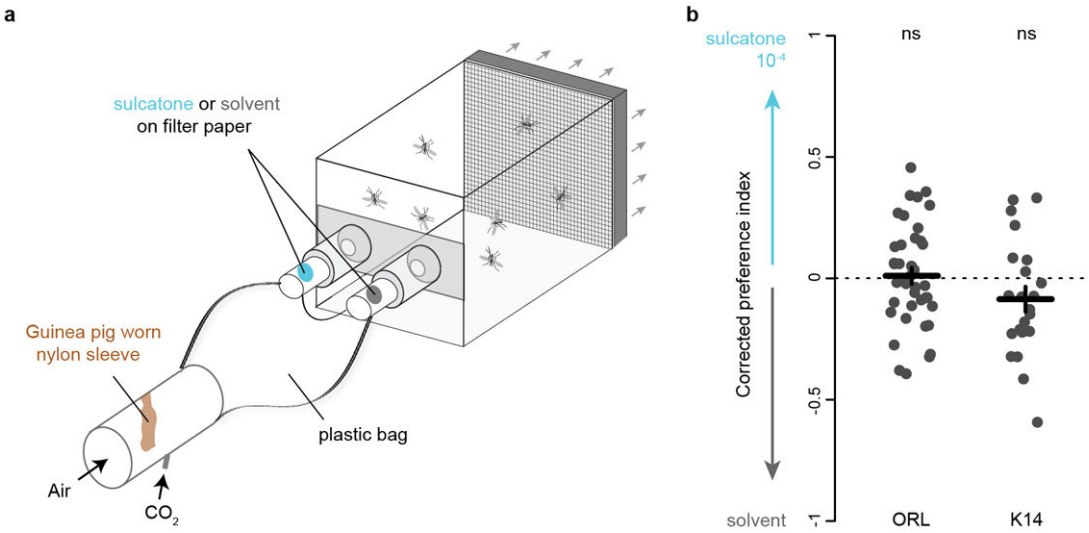
Extended Data Figure 3. Amino acid differences of major Or4 protein alleles

Dots represent amino acid differences with respect to the genome reference, not an inferred ancestor. Red dots indicate differences that are unique to the given allele. Blue dots indicate differences that are shared among multiple alleles. Snake plots are based on the predicted orientation and location of transmembrane domains Extracellular loops are oriented up and cytoplasmic loops are oriented down. Allele names are indicated to the left of each snake plot.



Extended Data Figure 4. Evidence that Or4 is a single copy gene

a, Histogram showing the number of alleles represented in the Or4-derived PacBio reads obtained for each of 270 parent and F2 hybrid mosquitoes. Alleles were only considered if they received at least 5% of an individual's reads. b, Histogram showing the fraction of reads from individual mosquitoes assigned to individual alleles. For all 270 mosquitoes, individual alleles were represented by either very few reads (grey bars – inferred to result from allele or barcode assignment errors or polymerase chain reaction contaminants), approximately half the reads (light blue bars – inferred to represent the two alleles in heterozygotes), or over 98% of all reads (dark blue bars – inferred to represent the single allele carried by homozygotes).



Extended Data Figure 5. Response of human-preferring mosquitoes to sulcatone-scented guinea pig odour

a, Olfactometer apparatus in which 50 mosquitoes per trial were given a choice between guinea-pig odour/CO2 mix supplemented with solvent on one side and sulcatone 10-4 on the other side. b, Corrected preference for sulcatone vs. solvent ports is indicated. Data were corrected for the daily average left-right side bias observed across 2 - 3 solvent vs. solvent tests conducted on each day of testing. An index value of 1 indicates strong preference for the sulcatone side, while -1 indicates strong preference for the solvent side. Neither mosquito colony showed a preference significantly different from zero (1-sample t-test P = 0.76 for ORL, P = 0.11 for K14). The trials for each colony were performed across 4 - 8 days (n = 40 for ORL and n = 22 for K14).

Methods

Ethics and regulatory—Mosquitoes were collected and exported from Kenya with approval of the director of the Kenya Medical Research Institute under the study approved by the Scientific Steering Committee and Ethical Review Committee (SSC No. 1679). Live mosquito eggs were imported to the USA with permits issued by the United States Department of Agriculture and the Centers for Disease Control and Prevention. The use of non-human animals in host preference tests at Rockefeller was approved and monitored by The Rockefeller University Institutional Animal Care and Use Committee (Protocol 11487). The participation of humans in blood-feeding mosquitoes during routine colony maintenance and as subjects in host preference tests at Rockefeller was approved and monitored by The Rockefeller University Institutional Review Board (IRB protocol LVO-0652). The protection of human subjects and ethical work with animals at the Swedish University of Agricultural Sciences (SLU) was in accordance with the Central Ethical Review Board and the Committee for Laboratory Animal Science in Sweden, respectively. All human subjects gave their informed consent to participate in the work carried out at The Rockefeller University and at SLU.

Field collection and creation of laboratory colonies—We collected mosquito (*Ae. aegypti*) eggs, larvae, and pupae in indoor and outdoor habitats in the Rabai region of Kenya in January 2009 (39° 34-36' E, 3° 55-57' S). Indoor collections in the Rabai region of Kenya (Fig. 1a) were made with the verbal permission of homeowners by visually scanning artificial containers used to store water with flashlights and removing larvae and pupae with a nylon sieve. Artificial containers harbouring mosquitoes included plastic buckets, metal jerry cans, and traditional earthenware pots (Fig. 1b). Outdoor collections were made in both village environments and nearby forest fragments along the Kombeni River (Fig. 1a) in two ways. First, larvae and pupae were removed from artificial containers left outdoors using a sieve (Fig. 1c), and from natural containers such as tree holes (Fig. 1c, top) using a turkey baster or small plastic pipette. In some cases, tree holes retained water from the last rains, while in others we introduced well water two days prior to collection to induce hatching of dormant eggs. Second, freshly laid eggs were collected in oviposition traps nailed to trees (Fig. 1c, bottom) or left on the ground for 3 days. Traps comprised black plastic cups (13 cm diameter, 15 cm tall) half-filled with water and lined with coarse brown seed-germination paper (76 pound, Anchor Paper) (Fig. 1c, bottom). Adult females attracted to the cups laid their eggs on the wet paper, which was then removed and dried to prevent embryos from hatching. When mosquitoes were collected as larvae or pupae, individuals originating in the same home/container or cluster of nearby homes/containers were reared to adulthood, mated with each other, blood-fed, and induced to oviposit in a field laboratory so that eggs of the first laboratory generation could be dried and exported to the USA. 29 laboratory colonies were established (K1 - K29), each founded by between 1 and 14 females (median 4) collected within 0-50 m of each other in the field (Fig. 1a) and mated to males collected from the same area. The only exception was K27, which was established using males from a forest tree hole that yielded no females. We therefore mated the males with females collected outdoors in a village 2 km away (Fig. 1a). The exact number of females that had the opportunity to contribute to the six Kenyan colonies characterized in Fig. 1g-I, 2g-h, and 3c-d is as follows: K2 – 3, K4 – 6, K14 – 2, K18 – 2, K19 – 6, K27 – 9. However, some of these females may not have laid eggs and thus ultimately not contributed genetic material to the colony. We also established one colony with eggs sent to us from Bundibugyo, Uganda and another with eggs sent from Rayong, Thailand. We denote the generation of a lab colony with lower case 'g' followed by a number; for example, g1 refers to the first laboratory generation.

Colony maintenance and insect rearing—Mosquitoes were maintained at 25 - 28°C with 70 - 80% relative humidity under a 14 h light: 10 h dark cycle (lights on 8 a.m.). Eggs were hatched by submersion in a broth containing deoxygenated deionised water and powdered Tetramin tropical fish food (Tetra). Larvae were cultured in deionised water and fed Tetramin tablets. Adults were maintained in large plastic cages (30 × 30 × 30 cm, BioQuip) and given unlimited access to 10% sucrose. In each generation, 75 - 250 adult females from each colony were blood-fed on a human volunteer. To minimise the potential for natural selection on host preference, a human arm was offered 1 - 2 times each day until 90% or more of the females had taken a full blood-meal. Eggs were collected in 96 ml black plastic soufflé cups (Solo Cup Company) lined with seed-germination paper and filled with 30 ml of "soil water" prepared by incubating deionised water with commercial potting soil

at room temperature in an open vessel for 1 - 10 weeks. Eggs were dried to prevent hatching and stored at 18°C, 85% relative humidity for 6 - 12 months. Eggs from domestic colonies K1 - K6 were less tolerant to drying and storage than eggs from forest colonies. They were therefore hatched after only 1 - 4 months, resulting in shorter generation times.

Morphological analysis—We characterised the scale and cuticle colour of mosquitoes from 14 colonies in 2010 (g3 to g7; n = 10 - 15 mosquitoes per colony), and a partially overlapping set of 8 colonies in 2012 (g5 to g11; n = 10 - 15 mosquitoes per colony). We collected adult female mosquitoes within 48 h of eclosion and stored them at -20°C for up to 4 weeks. Immediately upon removal from the freezer, we inserted an insect pin laterally through the thorax and positioned each specimen using a pinned specimen manipulator (Rose Entomology) one at a time under a microscope (Nikon Eclipse SMZ1500) fitted with a ring light and a digital camera (Nikon digital sight DS-2Mv controlled by NIS-Elements F v3.0 software). We took photographs in two characteristic positions, one highlighting the dark and light scales that decorate the scutum (Extended Data Fig. 1a) and the other showing the bare cuticle of the postnotum on the posterior face of the thorax plus the first 3-4 segments of the abdomen (Extended Data Fig. 1b). We measured scale and cuticle colour on the dark parts of the scutum and postnotum, respectively, in Adobe Photoshop (v. CS6, Adobe Systems Inc.) by sampling RGB values at four characteristic positions using the colour picker tool (Extended Data Fig. 1a-b), and converted RGB numbers to hue, saturation, value (HSV) using R software (v. 2.15.0, http://www.r-project.org/). The light environment and camera settings were held constant for the duration of each analysis, but differed slightly between analyses, such that values from 2010 and 2012 are not directly comparable.

We assessed the extent of white scaling on the first abdominal tergite in the 2012 analysis using an ordinal scale from 1 to 5 (Extended Data Fig. 1c) as follows: 1, up to a few scattered white scales; 2, small patch of white scales at midline; 3, contiguous patch of white scales at midline stretching from top to bottom of tergite and covering up to 60% of visible area; 4, contiguous patch of white scales covering 60 - 90% of visible area; 5, contiguous patch of white scales covering >90% of visible area. These scores correspond roughly to letters F through K in the scheme of McClelland[51,52]. A single experimenter scored all mosquitoes blind to their identity.

Host preference assays—Mosquitoes used in host preference assays were adult females 1-3 weeks post eclosion that had been given the opportunity to mate, but had not taken a blood-meal. Females were sorted briefly under cold anaesthesia (4°C) and deprived of access to food or water 16 - 24 h before testing.

Biting assay: This assay was used to test mosquito preference for human *versus* guinea pig (Fig. 2a) and was based on a previously described landing assay[53]. For each trial, we allowed approximately 50 females to acclimate overnight in a large custom-made cage (61 × 61 × 91.5 cm) constructed with aluminium screening on 3 sides and clear vinyl on the fourth side for easy viewing (BioQuip). The following morning, we simultaneously introduced a human arm (33 year old female) and an anaesthetised guinea pig (pigmented strain, one of two females, 2 - 6 months old) through cloth sleeves at opposite ends of the cage and rested

them on the floor of the cage ~60 cm apart (Fig. 2a), and recorded the number of mosquitoes that blood-fed within 10 min. We defined blood-feeding as landing on the host, inserting the proboscis, and drawing enough blood into the abdomen that it was visible to the naked eye of the observer. Preference index was calculated as number of mosquitoes that blood-fed on the human minus the number that blood-fed on the guinea pig divided by the total number of mosquitoes that blood-fed on either host. Overall response was the fraction of mosquitoes that blood-fed on either host. We tested g1 - g2 females from 16 colonies, 1 - 3 trials per colony and assessed the significance of the difference between forest and domestic colonies using two-sided, two-sample t-tests, where each colony served as a single data point.

Live host olfactometer assay: We used a live host olfactometer assay to test mosquito preference for human *versus* non-human animals in the absence of visual cues, as previously described[25] (Fig. 2d). 50 - 100 females acclimated in a dual-port olfactometer for 15 min. We then activated a fan and opened a sliding door exposing the mosquitoes to streams of air that had passed over the arm of a human volunteer (one of three females, 22 - 35 years old) or an awake guinea pig (pigmented strain, one of two females, 6 - 24 months old) or chicken (domestic strain, sex unknown, 2 - 4 weeks old). During an 8 min trial, mosquitoes choosing to fly upwind toward the odour of either host were trapped in small ports. The breath of both hosts provided a source of carbon dioxide (CO_2) in their respective air streams. The comfort of the animal was assured by rest and feeding periods outside of the olfactometer between trials and the minimization of any stress while in the olfactometer. Neither the animals nor the human volunteer were bitten by mosquitoes during this assay. We screened g1 - g5 females from 26 colonies, 1 - 7 trials per colony in the human *versus* guinea pig comparison and a subset of 5 colonies, 1 trial per colony for human *versus* chicken. A preference index equal to the number of mosquitoes entering the human trap minus those that entered the animal trap divided by the total number of mosquitoes entering either trap was calculated. Overall response was the fraction of mosquitoes that entered either trap. We assessed the significance of the difference between forest and domestic colonies using two-sided, two-sample t-tests, where the mean for each colony served as a single data point.

Nylon sleeve olfactometer assay: We substituted host-scented nylon sleeves for live hosts in the olfactometer to isolate the effects of specific host odours on mosquito preference in the context of a controlled amount of CO_2 (Fig. 2h, top). This assay was conducted as described[25], except that instead of pushing carbon-filtered air through the olfactometer with pumps, ambient air was gently pulled over the nylon sleeves and into the main compartment with a fan in exactly the same way as the live host assay. Both air streams were supplemented with an equal amount of CO_2, resulting in a final concentration of 0.2 - 0.3%. Human- and guinea pig-scented sleeves comprised 35 cm long sections of women's sheer nylon stockings that had been worn on a human arm (one of two females 22 - 35 years old) or guinea pig torso (pigmented strain, one of two females, 6 - 24 months old) for 24 h during normal daily activity. Conditioned sleeves were stored at -20°C for up to a month before use, and individual sleeves were used in only one trial per colony. We tested g2 - g8 females from 6 colonies, 4 - 17 trials per colony. We quantified preference and overall response as described for the live host olfactometer assay. We used a one-way ANOVA followed by Tukey's test to assess significance of variation among colonies.

Sulcatone perfuming experiments: The olfactometer apparatus was modified from the main figures as follows: the air stream was carbon-filtered and simultaneously pushed by a pump and pulled by an exhaust fan. The stream first mixed with CO_2 and picked up the odour of a guinea pig-worn nylon sleeve in a large cylinder before entering a plastic bag and splitting into two streams. The two streams then passed through two smaller cylinders containing a 55 mm diameter filter circle (Whatman) loaded with 50 µl of either solvent (paraffin oil) or 10^{-4} sulcatone, a concentration chosen because this is within the range of concentrations eluted from human-worn nylon sleeves and above the response threshold of most *AaegOr4* alleles as assessed in the *Drosophila* ab3A neuron. The air then entered two mosquito traps and flowed into a large mosquito chamber. In each trial, 50 1 - 4 week old female mosquitoes from the human-preferring Orlando (ORL) lab strain or human-preferring K14 Rabai strain were given a choice between the guinea pig odour/CO_2 mix supplemented with solvent on one side and sulcatone 10^{-4} on the other side. We used a 1-sample t-test to evaluate if preference was different from zero or no preference. -1 indicates strong preference for the solvent side and +1 indicates strong preference for sulcatone.

Isolation of human- and guinea pig-preferring F2 hybrids—We generated a large population of F2 mosquitoes by crossing ~150 individuals from domestic colony K14 (g5) with ~150 individuals from forest colony K27 (g5) in both directions and then interbreeding ~2000 of their F1 progeny (Fig. 3a). We chose colonies K14 and K27 as parents because they showed the most extreme preferences in the nylon sleeve assay (Fig. 2h). We reared ~2500 F2 females and identified the most human- and guinea pig-preferring among them using repeated tests in the live host olfactometer. F2 individuals derived from the reciprocal parent crosses were pooled for testing. The live host olfactometer was used as described above, except that human arm odour was not supplemented with human breath, and 200 - 300 females were tested at a time. In the first round of testing, mosquitoes that entered the human or guinea pig trap were separated. In subsequent rounds, mosquitoes that had previously responded to human were tested separately from mosquitoes that previously responded to guinea pig, and only individuals that responded to the same host as in previous trials were retained. Mosquitoes were allowed to rest in the insectary with unlimited access to 10% sucrose for 48 hr between trials. This process resulted in a pool of 141 females that responded to the human three times in a row, and 117 females that responded to the guinea pig twice in a row. We then returned the pools to the olfactometer one last time to quantify their preference (Fig. 3b), but retained all females in each pool regardless of their behaviour in this final test. We also tested individuals from the two parent colonies each day as positive controls. Each F2 hybrid pool was split into two equal-sized groups for the preparation of replicate RNAseq libraries.

RNAseq experiments—Antennal RNAseq libraries were prepared for the two F2 hybrid pools and 8 colonies using Illumina TruSeq RNA Sample Prep Kit v. 2 and mRNA-Sequencing Sample Prep Kit, respectively. Libraries were sequenced on Illumina GEX and HiSeq 2000 sequencing systems. We prepared antennal RNAseq libraries for the two F2 hybrid pools and the following colonies: K2 (g8), K4 (g8), K14 (g5), K18 (g5), K19 (g4), K27 (g5), Thailand (g4), and Uganda (g3). Colony females were 2 weeks post eclosion, had been given the opportunity to mate, but had not taken a blood-meal. F2 hybrids had

additionally experienced 3 - 4 rounds of preference testing and were 2 - 3 weeks post eclosion. One library was prepared for each colony (n = 8 libraries total), while two replicate libraries were prepared for each F2 hybrid pool (n = 4 libraries total), split for this purpose into two equal-sized groups as described above. For each library, we removed whole antennae, including pedicel and flagellum, from 55 - 150 cold-anaesthetised females with fine forceps under a stereo-microscope and placed them in a microcentrifuge tube held at -76°C in an ethanol/dry ice batch. We stored tubes at -80°C and then extracted total RNA using an RNeasy kit (Qiagen). For colonies, we prepared sequencing libraries from 1 - 3 μg total RNA with an mRNA-Sequencing Sample Prep Kit (Illumina). We selected 200 base pair (bp) inserts on a 2% agarose gel both prior to PCR enrichment, per kit instructions, and after PCR enrichment, to further narrow the insert size distribution. For F2 hybrid pools, we prepared sequencing libraries from 1.5 μg total RNA using a TruSeq RNA Sample Prep Kit v. 2 (Illumina) with a 4 min fragmentation time, resulting in a broad range of insert sizes (mean ~330 bp, s.d. ~90 bp). Colony libraries were each sequenced in one lane of an Illumina GEX and one lane of an Illumina HiSeq 2000, generating 20 - 30 million 76 bp and 65 - 90 million 101 bp single-end reads respectively. Barcoded F2 hybrid libraries were pooled, and the resulting pool was sequenced in four lanes of an Illumina HiSeq 2000, generating 175 - 240 million 101 bp single-end reads per library.

We used TopHat2 v. 2.0.9 (ref. [54]) to align unfiltered RNAseq reads to the *Ae. aegypti* AaegL1.3 reference transcriptome, custom-revised to include community annotations for four chemosensory gene families: olfactory receptors[39], ionotropic receptors[23], gustatory receptors[55], and odorant binding proteins[56]. We chose lenient alignment parameters allowing 3 mismatches per segment, 12 mismatches per read, and gaps of up to 3 bp (-N 12 --segment-mismatches 3 --read-gap-length 3 --read-edit-dist 15) to minimise the possibility that sequence divergence between forest or domestic mosquitoes and the reference genome would bias expression estimates. These settings resulted in the successful alignment of 47 - 56% of raw reads (43 - 62 million per colony). Default alignment parameters allowing 2 mismatches per read and gaps up to 2 bp resulted in the same qualitative conclusions. We then used Cuffdiff2 v. 2.1.1 (ref. [57]) to test for differential expression between forest and domestic colonies (n = 3 - 4 colonies each with each colony treated as a biological replicate), and between human- and guinea pig-preferring F2 hybrids (n = 2 replicate libraries each) with the multiple mapping correction and a false discovery rate set at 0.05. For colonies, we compared all four forest colonies (K18, K19, K27, Uganda) to all four domestic colonies (K2 K4, K14, Thailand) or just the three most guinea pig-preferring forest colonies (K19, K27, Uganda) to the three most human-preferring domestic colonies (K4, K14, Thailand). Results were similar and we present only the latter comparison in Figure 3c. However, data from the two excluded colonies (K2 and K18) are used in Extended Data Figure 2. We explored and visualised data with CummeRbund v. 2.0.0 (http://compbio.mit.edu/cummeRbund/) and custom R scripts (R software v. 2.15.0, http://www.r-project.org/).

Transgenic *Drosophila* strains—*UAS-AaegOr4* contained the full length coding sequence of each major allele cloned into the EcoRI site of the *pUAST attB* vector[58] using the In-Fusion HD Cloning Kit (Clontech). Constructs were injected into the *Drosophila melanogaster* ZH-86Fb strain [*P{ry[+t7.2]=hsp70-flp}1, y[1] w[*]; M{3xP3-*

RFP.attP}ZH-86Fb; M{vas-int.B}ZH-102F] (ref. [58]) using the phiC31-based integration system targeting the 86Fb-attP docking site on chromosome III by Genetic Services Inc. The resulting transgenic strains were outcrossed into a w^{1118}; double balancer background by selecting against markers in the parent strain. *UAS-AaegOr4* transgenes were crossed into a strain lacking *DmelOr22a/b* (*Dhalo*) and carrying *Or22a-GAL4* (ref.[26,59]). Electrophysiological recordings were obtained from flies of genotype *w;Dhalo/Dhalo;Or22a-GAL4/UAS-AaegOr4*.

Human and animal volatile collections—We collected headspace volatile extracts from nylon sleeves previously worn by each of 5 human volunteers (3 female, 22 - 47 years old) and 4 guinea pigs for 24 h (see nylon sleeve host preference assay), and unworn control sleeves (Fig. 4c-d). Four sleeves per human or guinea pig or control replicate were enclosed in a 0.5 l glass jar. A charcoal filtered air stream (1.5 l min^{-1}) was drawn by a diaphragm vacuum pump (KNF Neuberger) over the sleeves, from the bottom to the top of the jar, onto an air filter for 3 h. The air filters were made of Teflon tube (4 × 50 mm), holding 50 mg Porapak Q adsorbent (80/100 mesh, Altech) between glass wool plugs. The filters were rinsed with 4 ml acetone (>99.9% pure, HPLC Grade, Chromasolv Plus, Sigma Aldrich), redistilled ethanol and pentane (>99% pure, GC grade, Sigma Aldrich) before use. Adsorbed volatiles were desorbed by eluting with 500 µl pentane, and condensed to 20% of their original volume, under a stream of nitrogen. Heptyl acetate (50 ng, 99.8% pure, Sigma Aldrich) was added as an internal quantification standard. Two negative technical controls were included: a filter control sample obtained by running an elution step on cleaned adsorbent filters and a solvent only control.

We also collected headspace volatiles from live human volunteers and live non-human animals or hair from non-human animals. Human body volatiles were collected as described[60] by placing naked volunteers in customised heat-sealed cooking bags, introducing synthetic air into the bags, and extracting it with pumps through columns containing Porapak Q for a period of 2.5 h (n = 20 humans). Volatiles were collected from one live chicken placed on a metal mesh in an airtight desiccator covered with a black cloth. Charcoal filtered air was introduced (1 l min^{-1}) via a Teflon tube and pumped out of the desiccator via a glass splitter connected to four Porapak Q adsorbent columns (0.25 l min^{-1} each) for 1 hr. Volatiles from cattle, horse, and sheep were collected by placing 20 g of freshly shaved hair or wool from multiple individuals in a 0.5 l glass wash bottle (n = 10 cattle, n = 2 horses, n = 5 sheep). Charcoal filtered air was drawn by pumps (0.1 l min^{-1}) through the bottle onto a Porapak Q adsorbent column over 24 h. Before use, the adsorbent columns were rinsed with 1 ml each of methanol, dichloromethane, and pentane. Trapped volatiles were desorbed by eluting each column with 600 µl of pentane (*puriss p.a.*, Sigma-Aldrich Chemie GmbH). Heptyl acetate was added to each extract as an internal quantification standard (500 ng for human, 2 µg for chicken, 100 ng for other animals). Extracts from the 20 individual humans were pooled before the addition of the internal standard. Final collections were concentrated under a gentle stream of nitrogen before analysis.

Gas chromatography-coupled single sensillum recording (GC-SSR)—We used GC-SSRs as described[61] to screen human and guinea pig odour for individual volatiles that activated *Or4* allele A, which was heterologously expressed in the *Drosophila* ab3A neuron (Fig. 4a-b). We separated volatiles from host odour extracts using an Agilent 6890 gas chromatograph (GC) (Agilent Technology) fitted with a fused silica capillary column (30 m × 0.25 mm i.d.) coated with non-polar HP-5 stationary phase (d.f.=0.25 μm), and using hydrogen gas as the mobile phase (45 cm s^{-1}). Aliquots of the extracts (5-7.5 μl) were injected splitless for 30 s, with the injector maintained at 225°C. The GC oven temperature was programmed from 30°C (3 min hold), followed by a ramp of 8°C/min to 225°C, and held isothermal for 10 min. The GC was fitted with a make-up gas fed (4 psi N_2) four-way cross (Graphpack® 3D/2 Crosspiece Sulfinert™, Gerstel) at the end of the column, delivering half of the effluent to the flame ionization detector and the other half to the air stream passing over the antenna of an immobilised fly via a Gerstel ODP-2 transfer line maintained at 135°C for 15 min and increased at 8°C min^{-1}.

We monitored the response to GC-separated volatiles of *Or4*-expressing ab3A neurons of a fly restrained under a Nikon Eclipse microscope (E600-FN8). Using a piezoelectric micromanipulator (DC-3K, Märzhauser), an electrolytically sharpened tungsten microelectrode was introduced into the shaft or base of an ab3 sensillum and the reference tungsten electrode was inserted into the eye of the fly. The recording electrode was connected to a preamplifier (x10, Syntech) and the electrical signals were fed through an analogue-digital signal converter (IDAC-4, Syntech) and then visualised and recorded on a computer using Autospike software (Syntech). The mounted fly was placed in a continuous humidified charcoal-filtered airstream, into which GC-separated volatiles were introduced, delivered at 1 m s^{-1} via a glass tube (6mm i.d.).

We performed 3 replicates of GC-SSR for each of the following host odour collections: human body headspace, human-worn sleeve headspace, guinea pig-worn sleeve headspace. Compounds were considered bioactive if their elution corresponded to changes in ab3A neuron activity in all 3 replicates. Bioactive compounds were then identified via gas chromatography-mass spectrometry (GC-MS) as described below.

Chemical analysis of volatile collections—We used gas chromatography-mass spectrometry (GC-MS) to identify the single bioactive component of human sleeve odour noted in GC-SSR experiments. Human sleeve volatile extract was injected (2 μl) into a combined Agilent 6890N gas chromatograph and 5975 mass spectrometer (Agilent Technology) fitted with an HP-5 column and programmed as for the GC-SSR analyses. The active compound was identified by comparison with reference mass spectra in our custom made library and commercially available libraries (NIST05 and Wiley). The putative identification of sulcatone was confirmed by parallel injections of synthetic reference compounds with authentic samples on the GC-MS. We also used GC-MS to quantify the sulcatone and benzaldehyde content of human and animal volatile collections by ratio of their corresponding peak areas to that of the internal standard, heptyl acetate. Absolute quantities were divided by the total time over which the respective collections were made to calculate emission rate. Emission rates of human, guinea pig, and control sleeve extracts were compared using pairwise non-parametric Wilcoxon tests with Bonferroni correction.

***Or4* cDNA cloning—**We prepared two antennal cDNA libraries for each parent colony. For the first library we started with the same total antennal RNA used to prepare RNAseq libraries. We treated the RNA with DNase using the TURBO DNA-free Kit (Ambion) and synthesised single-stranded cDNA using Superscript III (Invitrogen) according to instructions. For the second library, we started with total antennal RNA prepared as described for RNAseq experiments, except the dissected females came from the subsequent generation (K14 g6, K27 g6). We again treated with DNase, but then isolated mRNA using the Oligotex mRNA Mini Kit (Qiagen) before synthesizing cDNA with Superscript III. We amplified the full coding sequence of *Or4* from each cDNA library with AccuPrime *Taq* DNA polymerase (Invitrogen) using primers designed to anneal to the 5' and 3' untranslated regions: 5'-CGGAGTTTCCTTCGTCAAGA-3' (forward), 5'-TCGACCACTCCTATACATCGC-3' (reverse). Gel-extracted PCR amplicons (MinElute Gel Extraction Kit, Qiagen) were then cloned into pCR4 using the TOPO TA Cloning Kit (Invitrogen) and sequenced (Genewiz Inc.). We sequenced 48 clones for each cDNA library, or 96 clones per parent colony. All 192 sequences fell within one of 8 distinct allele classes, 7 of which were subsequently named alleles A-G (see below). 3' RACE reactions produced no additional sequences, providing further evidence that these alleles were representative of parent colony diversity.

***Or4* haplotype network and outgroup—**We used the program SplitsTree4[62] to infer the haplotype network shown in Fig. 5a using the distance-based SplitDecomposition algorithm. The out group sequence from *Aedes albopictus* was identified via blast search of allele A to an *Ae. albopictus* draft genome assembly, which was partly supported by the Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (2009) (Xiaoguang Chen, personal communication).

***Or4* genotyping and definition of major *Or4* alleles—**Single Molecule Real Time (SMRT) sequencing (Pacific Biosciences) of barcoded PCR amplicons was used to genotype 34 females from each parent colony and 100 - 102 females from each F2 hybrid pool. DNA was extracted from bodies (parent colonies) or single legs (F2 hybrids) using a DNeasy Blood & Tissue Kit (Qiagen). The individuals providing DNA were a subset of those whose antennae had been dissected for RNAseq. We then amplified a 786 bp fragment of *Or4* with KOD Hot Start DNA Polymerase (Novagen) using barcoded primers complementary to invariant regions of the 2nd and 3rd coding exons: 5'-barcode-GTTGACCTATTGCGTTTTCG-3' (forward), 5'-barcode-GCACATCAGAACAGAACTTGC-3' (reverse). The 48 forward and 48 reverse barcode sequences, provided by Pacific Biosciences, were 16 bp long and custom-paired in 92 unique combinations. We were thus able to create 3 independent pools, each containing the purified PCR amplicons of up to 92 mosquitoes. A sequencing library was prepared from each pool and sequenced on 1-2 SMRT cells of a Pacific Biosciences RS Sequencer by staff at the Cold Spring Harbor Laboratory's PacBio core facility. We obtained 20,000 - 30,000 750-850 bp Circular Consensus Sequence (CCS) reads per pool.

We deconvoluted CCS read barcodes using the PacBioBarcodeIDCCS.py python script (http://pacificbiosciences.github.com/DevNet/), retaining only those reads with high quality

assignments (barScore> 38). This resulted in an average of 158 reads per individual. We then assigned reads to alleles via blastn search[63] to a custom database containing the eight *Or4* coding sequences cloned from cDNA. Since SMRT sequencing has a relatively high error rate, we did not require a perfect match between reads and assigned alleles, instead simply assigning reads to the allele with the best blastn score. Indeed, a fraction of reads from all mosquitoes had mismatches with respect to their assigned alleles, but the median number of mismatches per individual was almost always zero. In a few cases, individuals had reads that differed from the alleles to which they had been assigned by a consistent number of mismatches, suggesting real variation rather than sequencing error. We determined that these corresponded to rare alleles not identified by cDNA cloning and represented at a frequency of < 0.1 in both parents and F2 pools and did not study them further.

The genotyping strategy described above could not differentiate two closely related alleles, subsequently named A and B (see below) that differed by a single SNP located outside the PCR amplicon. We therefore conducted secondary genotyping on all parent and F2 mosquitoes that were shown to carry one of these two alleles (n = 152). For each individual, we used a LightScanner (Idaho Technology Inc.) to characterize the melting curves of a 196 bp PCR amplicon from a 3'-blocked, unlabelled, 21 bp oligonucleotide probe complementary to allele B in the region surrounding the diagnostic SNP. Melting curves were characteristic to genotype, allowing easy discrimination of individuals carrying 0, 1, or 2 copies of allele B. Reactions were prepared and analysed according to manufacturer instructions using the following primers and probe: 5'-CAAGGTCTTGCAAATGATCGGTAA-3' (forward), 5'-CGATGTTGATGATCTGACCGAAA-3' (reverse), 5'-AAGTCCAGTTCCGGTTTCGTGamino-modifier-'3' (probe).

Seven of the 8 full-length *Or4* alleles cloned from cDNA were present at a frequency of ≥ 0.1 in at least one parent colony or F2 hybrid pool. We defined these as major alleles and named them A through G in order of increasing distance from the genome reference allele A, and deposited them in GenBank (accession numbers KF801614-KF801615 and KF801617-KF801621). The 8th allele discovered by cDNA cloning was present at a frequency of < 0.1 and was not studied further.

We quantified the relative frequency of each major allele in F2 hybrids using a frequency index equal to frequency in human-preferring F2s minus frequency in guinea pig-preferring F2s divided by the sum of the frequencies in the two types of F2s.

Confirmation of *Or4* as a single copy gene—The *Ae. aegypti* reference genome contains another olfactory receptor, named *Or5*, that is 96% identical to the *Or4* gene across the coding sequence, leading to a predicted Or5 protein that is 97% identical to the Or4 protein. Although we did not sample any sequences identical to *Or5* in our field-derived colonies, some of the alleles we sampled were more similar to *Or5* than to *Or4*, raising the possibility that they actually belong to this second hypothetical locus. To confirm that all sampled alleles segregate at a single locus, we reanalysed our PacBio genotyping data, focusing on the number of alleles carried by individual mosquitoes. As described above, we

amplified a diagnostic segment of *Or4*/*Or5* from the genomic DNA of 270 females and sequenced ~150 pieces of DNA from each resulting amplicon. The reads for each mosquito were then assigned to alleles based on sequence. The vast majority of reads for each mosquito corresponded to a single allele (homozygotes) or were evenly split between two alleles (heterozygotes) (Extended Data Fig. 4). No mosquito had significant numbers of reads assigned to more than two alleles. We conclude that *Or4* is a single locus with multiple diverse alleles and suspect that the gene annotated as *Or5* represents a misassembly of *Or4* allelic sequences segregating within the genome reference strain.

Or4 transmembrane prediction and snake plots—For Extended Data Figure 3, we used TMHMM (v. 2.0)[64] to predict the location of transmembrane domains in the A reference of allele of Or4. The output was manually edited to remove a predicted transmembrane domain that occurred in an anomalous position relative to predictions for the olfactory co-receptor Orco and other ligand-selective ORs. Snake plots were generated using TOPO2 (ref. [65]).

Single sensillum electrophysiology—We used single sensillum recording (SSR) to characterise the spontaneous activity and odour-evoked responses of 7 major *Or4* alleles heterologously expressed in *Drosophila* ab3A olfactory sensory neurons. Recordings were conducted as described previously[41] using 5 - 11 day old females and a modified method for immobilizing the third antennal segment. We glued a small wedge-shaped plastic Lego piece (rectangular footprint 15 mm × 10 mm with angled surface rising at 45° to 10 mm above base) to a glass microscope slide such that a coverslip affixed to its upper face with double stick tape would protrude upwards at a fixed angle of 45°. We then immobilised the fly in a pipette tip mounted on a ball of dental wax opposite the coverslip as described[66]. After bringing the base of the fly's antennae into contact with the edge of the coverslip, we used two small drops of UV glue (KOA 300, Kexmert) to secure the rim of the cut pipette tip on either side of the protruding fly head to the edge of the coverslip and cured it for 10 s with a 405 nm violet laser pointer. We then used a glass micropipette to gently lower one antenna towards the angled coverslip until the arista contacted a small drop of UV glue placed directly below, and immediately cured the glue with the laser. These modifications prevented the fly preparation and antenna from drifting away from the coverslip over the course of a recording. Neither transgenic nor wild type ab3 sensilla responded to the odour of the glue.

We prepared serial dilutions of sulcatone (C.A.S. 409-02-9, Sigma Aldrich) v/v in paraffin oil and loaded 30 µl aliquots into odour delivery pipettes on the day of recording as described[66]. During recording, we applied a continuous stream of charcoal-filtered air to the insect antenna. For each stimulus, we first cleared the odour delivery pipette of accumulated volatiles by redirecting a fraction of the air stream through it and away from the preparation for a 1 s pulse. Fifteen seconds later, we then delivered the stimulus by redirecting air through the pipette for another 1 s pulse, but this time with the tip inserted back into the air stream flowing over the antenna. We used each pipette no more than twice.

We applied dilutions to a single ab3 sensillum per fly in the following order: solvent, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}. To increase resolution for the most biologically relevant doses,

we sometimes stimulated a second sensillum on the same fly with 10^{-5} and 10^{-4} dilutions only. Final sample sizes were n = 13 - 17 sensilla per genotype for 10^{-5} and 10^{-4} and n = 8 - 10 sensilla per genotype for solvent and all other dilutions. We identified ab3 sensilla by size, location on the antenna, lack of A cell response to the ab2A ligand ethyl acetate (10^{-2}) (C.A.S. 141-78-6, Sigma Aldrich), and strong B cell response to its cognate ligand, 2-heptanone(10^{-4})[67] (C.A.S. 110-43-0, Sigma Aldrich). We confirmed the *Dhalo/Dhalo* genotype of each fly by lack of A cell response to ethyl hexanoate (10^{-4}) (C.A.S. 123-66-0, Sigma Aldrich).

We recorded and processed neuronal activity using AutoSpike software (Syntech). Spontaneous activity was averaged over a 14 s period for each sensillum. Evoked response was calculated by subtracting average spontaneous activity in the 3 s before stimulus onset from average activity in the 1 s after stimulus onset. The distance air travels through the odour delivery system causes a delay between digitally recorded onset and the time the stimulus reaches the antenna. We therefore defined stimulus onset on the basis of the excitatory response elicited in the ab3B cell by 10^{-2} sulcatone; this invariably occurred at 200 ms after the computer-recorded onset. We compared spontaneous and evoked activity across fly genotypes using 1-way ANOVA followed by Tukey's test. We fit monotone cubic splines to evoked response data for each genotype (Fig. 5f) using the R function smooth.monotone.

Analysis of allele-specific expression—We estimated the allele-specific expression of the seven major *Or4* alleles in F2 hybrid RNAseq libraries as follows. First, we altered our reference transcriptome by removing reference transcripts for *Or4* and *Or5* and adding separate transcript sequences for each major *Or4* allele A-G. We then realigned unfiltered RNAseq reads to this transcriptome using Bowtie2 v. 2.1.0 (ref. [68]) and estimated expression using eXpress v. 1.5 (ref. [69]) specifying *Or4* alleles as haplotypes of a single locus. We used the Bowtie2 alignment parameters recommended for use with eXpress, allowing an unlimited number of hits per read (-a −rdg 6,5 −rfg 6,5 −score-min l,-0.6,-0.4). eXpress then uses a probabilistic model to weight the hits from which it may be derived (ref. [69]). The overall alignment rate was 47 - 51% for all libraries. For each F2 pool, we estimated expression for the two replicate libraries separately and then averaged them. The replicate estimates for a single pool were generally closer to each other than either was to estimates from the other pool. Expression values for each allele are partly a function of the frequency of that allele among the 110 - 125 individuals whose antennae were dissected for RNA extraction. We therefore normalised each allele's expression in a given F2 pool by its frequency among the individuals that made up that pool to generate final estimates of allele-specific expression. We used a 2-way ANOVA to test for significant differences between the estimates derived from human- *versus* guinea pig-preferring pools and among the estimates for individual alleles.

Statistical analysis—All statistical analyses were performed with JMP software v. 8 (SAS Institute, Inc.) or R software v. 2.15.0 (http://www.r-project.org/). For all parametric tests, including t-tests and analyses of variance, data were tested and met the assumptions of normality and homogeneity of variance.

References cited

51. McClelland GAH. A preliminary study of genetics and abdominal colour variations in *Aedes aegypti* (L.). Ann Trop Med Parasitol. 1960; 54:305–320.
52. McClelland G. Worldwide survey of variation in scale pattern of abdominal tergum of *Aedes aegypti* (L) (Diptera-Culicidae). T Roy Ent Soc London. 1974; 126:239–259.
53. Mukwaya L. Genetic control of feeding preferences in the mosquitoes *Aedes* (*Stegomyia*) *simpsoni* and *aegypti*. Physiol Entomol. 1977; 2:133–145.
54. Kim D, et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14:R36. [PubMed: 23618408]
55. Kent LB, Walden KKO, Robertson HM. The Gr family of candidate gustatory and olfactory receptors in the yellow-fever mosquito *Aedes aegypti*. Chem Senses. 2008; 33:79–93. [PubMed: 17928357]
56. Zhou J, He X, Pickett J, Field L. Identification of odorant-binding proteins of the yellow fever mosquito *Aedes aegypti*: genome annotation and comparative analyses. Insect Mol Biol. 2008; 17:147–163. [PubMed: 18353104]
57. Trapnell C, et al. Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature Biotechnol. 2012; 31:46–53. [PubMed: 23222703]
58. Bischof J, Maeda RK, Hediger M, Karch F, Basler K. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. Proc Natl Acad Sci USA. 2007; 104:3312–3317. [PubMed: 17360644]
59. Hallem E, Ho M, Carlson J. The molecular basis of odor coding in the *Drosophila* antenna. Cell. 2004; 117:965–979. [PubMed: 15210116]
60. Harraca V, Ryne C, Birgersson G, Ignell R. Smelling your way to food: can bed bugs use our odour? J Exp Biol. 2012; 215:623–629. [PubMed: 22279069]
61. Stensmyr MC, Giordano E, Balloi A, Angioy AM, Hansson BS. Novel natural ligands for *Drosophila* olfactory receptor neurones. *J Exp Biol*. 2003; 206:715–724.
62. Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006; 23:254–267. [PubMed: 16221896]
63. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215:403–410. [PubMed: 2231712]
64. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol. 2001; 305:567–580. [PubMed: 11152613]
65. Johns, SJ. TOPO2, Transmembrane protein display software. http://www.sacs.ucsf.edu/TOPO2/
66. Pellegrino M, Nakagawa T, Vosshall LB. Single sensillum recordings in the insects *Drosophila melanogaster* and *Anopheles gambiae*. J Vis Exp. 2010; 36:1–5. [PubMed: 20164822]
67. Hallem EA, Carlson JR. Coding of odors by a receptor repertoire. Cell. 2006; 125:143–160. [PubMed: 16615896]
68. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–359. [PubMed: 22388286]
69. Roberts A, Pachter L. Streaming fragment assignment for real-time analysis of sequencing experiments. Nat Methods. 2013; 10:71–73. [PubMed: 23160280]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Mara Lawniczak, Kevin Lee, Michael Nitabach, and the Vosshall lab for discussion and comments on the manuscript; Julia Brown and Jeffrey Powell for discussion and coordination of field collections; Willem Takken for advice regarding many aspects of this work; John-Paul Mutebi, Barry Miller, and Alongkat Ponlawat for live specimens from Uganda and Thailand; Deborah Beck, Kara Nygaard, Katya Prakash, and Laura Seeholzer for expert technical assistance. We also thank Xiaoguang Chen for pre-publication access to a draft *Ae. albopictus*

genome assembly, and Jeff Liesch for access to Orlando strain RNAseq data. We received valuable advice on collecting and working with forest and domestic forms of *Ae. aegypti* from Milan Trpis, John Peterson, and Phil Lounibos, and on the design and use of two-port olfactometers from Ulrich Bernier and Vadim Sherman. This work was funded in part by a grant to Richard Axel and L.B.V. from the Foundation for the National Institutes of Health through the Grand Challenges in Global Health Initiative. This work was supported in part by the following National Institutes of Health grants: K99 award from NIDCD to C.S.M. (DC012069), an NIAID Vectorbase DBP subcontract to L.B.V. (HHSN272200900039C), and a CTSA award from NCATS (5UL1TR000043). R.I. received support from the Swedish Research Council and SLU: Insect Chemical Ecology and Evolution (IC-E3). L.B.V. is an investigator of the Howard Hughes Medical Institute.

References

1. Lehane, MJ. The biology of blood-sucking in insects. Cambridge University Press; 2005.
2. Christophers SR. Aedes aegypti, the yellow fever mosquito: Its life history, bionomics, and structure. 1960
3. Mattingly PF. Genetical aspects of the *Aedes aegypti* problem. I: Taxonomy and bionomics. Ann Trop Med Parasitol. 1957; 51:392–408. [PubMed: 13498658]
4. Brown JE, et al. Human impacts have shaped historical and recent evolution in *Aedes aegypti*, the dengue and yellow fever mosquito. Evolution. 201310.1111/evo.12281
5. Lumsden WH. An epidemic of virus disease in Southern Province, Tanganyika Territory, in 1952-53. II. General description and epidemiology. Trans R Soc Trop Med Hyg. 1955; 49:33–57. [PubMed: 14373835]
6. Trpis M, Hausermann W. Demonstration of differential domesticity of *Aedes aegypti* (L) (Diptera, Culicidae) in Africa by mark release recapture. Bull Entomol Res. 1975; 65:199–208.
7. Petersen, JL. Ph D thesis. University of Notre Dame; 1977. Behavioral differences in two subspecies of Aedes aegypti (L.) (Diptera: Culicidae) in East Africa.
8. Trpis M, Hausermann W. Genetics of house entering behavior in East African populations of *Aedes aegypti* (L) (Diptera-Culicidae) and its relevance to speciation. Bull Entomol Res. 1978; 68:521–532.
9. Saul SH, Novak RJ, Ross QE. The role of the preadult stages in the ecological separation of 2 subspecies of *Aedes aegypti*. Am Midl Nat. 1980; 104:118–134.
10. Gouck HK. Host preferences of various strains of *Aedes aegypti and Aedes simpsoni* as determined by an olfactometer. B World Health Organ. 1972; 47:680–683.
11. McKenna, RJ. Ph D thesis. University of California; Davis: 1973. Attraction of seven strains of Aedes aegypti to man and guinea pig in the laboratory.
12. Bhatt S, et al. The global distribution and burden of dengue. Nature. 2013; 496:504–507.10.1038/nature12060 [PubMed: 23563266]
13. Brown J, et al. Worldwide patterns of genetic differentiation imply multiple 'domestications' of *Aedes aegypti*, a major vector of human diseases. Proc Biol Sci. 2011; 278:2446–2454. [PubMed: 21227970]
14. Mukwaya LG. Host preference in *Aedes* (*Stegomyia*) mosquitoes in Uganda. II. Studies on indoor and outdoor biting and resting behaviour with special reference to *Aedes aegypti L*. Acta Trop. 1974; 31:165–176. [PubMed: 4152901]
15. Dekker T, Geier M, Carde R. Carbon dioxide instantly sensitizes female yellow fever mosquitoes to human skin odours. J Exp Biol. 2005; 208:2963–2972. [PubMed: 16043601]
16. Dekker T, Ibba I, Siju KP, Stensmyr MC, Hansson BS. Olfactory shifts parallel superspecialism for toxic fruit in *Drosophila* melanogaster sibling, *D sechellia*. Curr Biol. 2006; 16:101–109. [PubMed: 16401429]
17. Matsuo T, Sugaya S, Yasukawa J, Aigaki T, Fuyama Y. Odorant-binding proteins OBP57d and OBP57e affect taste perception and host-plant preference in *Drosophila sechellia*. PLoS Biol. 2007; 5:e118. [PubMed: 17456006]
18. Linz J, et al. Host plant-driven sensory specialization in *Drosophila erecta*. Proc Biol Sci. 2013; 280:2013.0626.

19. Rinker DC, Zhou X, Pitts RJ, Rokas A, Zwiebel LJ. Antennal transcriptome profiles of anopheline mosquitoes reveal human host olfactory specialization in *Anopheles gambiae*. BMC genomics. 2013; 14:749. [PubMed: 24182346]
20. Wada-Katsumata A, Silverman J, Schal C. Changes in taste neurons support the emergence of an adaptive behavior in cockroaches. Science. 2013; 340:972–975. [PubMed: 23704571]
21. Michelmore RW, Paran I, Kesseli RV. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA. 1991; 88:9828–9832. [PubMed: 1682921]
22. Robertson HM, Warr CG, Carlson JR. Molecular evolution of the insect chemoreceptor gene superfamily in *Drosophila melanogaster*. Proc Natl Acad Sci USA. 2003; 100(Suppl 2):14537–14542. [PubMed: 14608037]
23. Croset V, et al. Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction. PLoS Genet. 2010; 6:e1001064. [PubMed: 20808886]
24. Fan J, Francis F, Liu Y, Chen JL, Cheng DF. An overview of odorant-binding protein functions in insect peripheral olfactory reception. Genet Mol Res. 2011; 10:3056–3069. [PubMed: 22180039]
25. DeGennaro M, et al. *orco* mutant mosquitoes lose strong preference for humans and are not repelled by volatile DEET. Nature. 2013; 498:487–491. [PubMed: 23719379]
26. Dobritsa AA, van der Goes van Naters W, Warr CG, Steinbrecht RA, Carlson JR. Integrating the molecular and cellular basis of odor coding in the *Drosophila* antenna. Neuron. 2003; 37:827–841. [PubMed: 12628173]
27. Cork A, Park KC. Identification of electrophysiologically-active compounds for the malaria mosquito, *Anopheles gambiae*, in human sweat extracts. Med Vet Entomol. 1996; 10:269–276. [PubMed: 8887339]
28. Labows J, Preti G, Hoelzle E, Leyden J, Kligman A. Analysis of human axillary volatiles: compounds of exogenous origin. J Chromatogr. 1979; 163:294–299. [PubMed: 541384]
29. Bernier UR, Kline DL, Barnard DR, Schreck CE, Yost RA. Analysis of human skin emanations by gas chromatography/mass spectrometry. 2. Identification of volatile compounds that are candidate attractants for the yellow fever mosquito (*Aedes aegypti*). Anal Chem. 2000; 72:747–756. [PubMed: 10701259]
30. Syed Z, Leal WS. Acute olfactory response of *Culex* mosquitoes to a human- and bird-derived attractant. Proc Natl Acad Sci USA. 2009; 106:18803–18808. [PubMed: 19858490]
31. Birkett MA, et al. The role of volatile semiochemicals in mediating host location and selection by nuisance and disease-transmitting cattle flies. Med Vet Entomol. 2004; 18:313–322. [PubMed: 15641996]
32. Jumean Z, Gries R, Unruh T, Rowland E, Gries G. Identification of the larval aggregation pheromone of codling moth, *Cydia pomonella*. J Chem Ecol. 2005; 31:911–924. [PubMed: 16124259]
33. Nielsen BLL, Jerôme N, Saint-Albin A, Rampin O, Maurin Y. Behavioural response of sexually naïve and experienced male rats to the smell of 6-methyl-5-hepten-2-one and female rat faeces. Physiol Behav. 2013; 120:150–155. [PubMed: 23911690]
34. Stoeffler M, Maier TS, Tolasch T, Steidle JLM. Foreign-language skills in rove-beetles? Evidence for chemical mimicry of ant alarm pheromones in myrmecophilous *Pella* beetles (Coleoptera: Staphylinidae). J Chem Ecol. 2007; 33:1382–1392. [PubMed: 17558536]
35. Rines HW, French RC, Daasch LW. Nonanal and 6-methyl-5-hepten-2-one: endogenous germination stimulators of uredospores of *Puccinia graminis* var. tritici and other rusts. J Agric Food Chem. 1974; 22:96–100. [PubMed: 4855793]
36. Socaci SA, et al. Chemometric discrimination of different tomato cultivars based on their volatile fingerprint in relation to lycopene and total phenolics content. Phytochem analysis. 2013
37. Webster B, et al. Identification of volatile compounds used in host location by the black bean aphid, *Aphis fabae*. J Chem Ecol. 2008; 34:1153–1161. [PubMed: 18584254]
38. Carey AF, Wang G, Su CY, Zwiebel LJ, Carlson JR. Odorant reception in the malaria mosquito *Anopheles gambiae*. Nature. 2010; 464:66–71.10.1038/nature08834 [PubMed: 20130575]

39. Bohbot J, et al. Molecular characterization of the *Aedes aegypti* odorant receptor gene family. Insect Mol Biol. 2007; 16:525–537. [PubMed: 17635615]
40. Krzywinski J, Grushko OG, Besansky NJ. Analysis of the complete mitochondrial DNA from Anopheles funestus: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. Mol Phylogenet Evol. 2006; 39:417–423.10.1016/j.ympev.2006.01.006 [PubMed: 16473530]
41. Pellegrino M, Steinbach N, Stensmyr MC, Hansson BS, Vosshall LB. A natural polymorphism alters odour and DEET sensitivity in an insect odorant receptor. Nature. 2011; 478:511–514. [PubMed: 21937991]
42. Keller A, Zhuang H, Chi Q, Vosshall LB, Matsunami H. Genetic variation in a human odorant receptor alters odour perception. Nature. 2007; 449:468–472. [PubMed: 17873857]
43. Logan J, et al. Arm-in-cage testing of natural human-derived mosquito repellents. Malar J. 2010; 9:239. [PubMed: 20727149]
44. Logan JG, et al. Identification of human-derived volatile chemicals that interfere with attraction of *Aedes aegypti* mosquitoes. J Chem Ecol. 2008; 34:308–322. [PubMed: 18306972]
45. Menger DJ, JJ VANL, Takken W. Assessing the efficacy of candidate mosquito repellents against the background of an attractive source that mimics a human host. Med Vet Entomol. 201410.1111/mve.12061
46. Greenwood AK, Wark AR, Yoshida K, Peichel CL. Genetic and neural modularity underlie the evolution of schooling behavior in threespine sticklebacks. Curr Biol. 2013; 23:1884–1888. [PubMed: 24035541]
47. Weber JN, Peterson BK, Hoekstra HE. Discrete genetic modules are responsible for complex burrow evolution in *Peromyscus* mice. Nature. 2013; 493:402–405. [PubMed: 23325221]
48. Martin A, Orgogozo V. The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution. 2013; 67:1235–1250. [PubMed: 23617905]
49. van der Goes van Naters W, Carlson JR. Insects as chemosensors of humans and crops. Nature. 2006; 444:302–307.10.1038/nature05403 [PubMed: 17108954]
50. Coyne, JA.; Orr, HA. Speciation. Sinauer Associates; 2004.

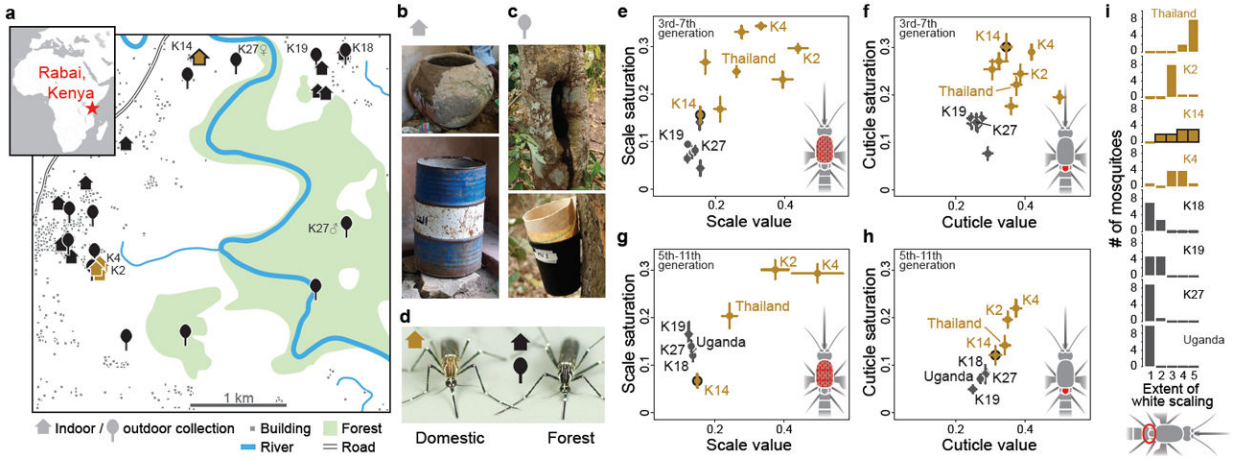


Figure 1. Field collection of forest and domestic forms of *Ae. aegypti* in Rabai, Kenya

a, Map of Rabai, Kenya showing collection sites, with six colonies labelled. **b-c**, Typical indoor (**b**) and outdoor (**c**) water containers or traps. **d**, Examples of characteristic differences in body colour between domestic (left) and forest (right) females, selected from the extremes of variation. **e-h**, Summary of colour variation in scales (**e**,**g**) and cuticle (**f**,**h**) among laboratory colonies tested after 3 - 7 (**e-f**) and 5 - 11 generations (**g-h**), including colonies from Thailand and Uganda. Cartoons at the bottom of **e-i** indicate in red where morphology was sampled. See also Extended Data Figure 1. Data are plotted as mean ± s.e.m. (n = 10 - 15 mosquitoes/colony). **i**, Extent of white scaling on the first abdominal tergite (n = 10 mosquitoes/colony). In all figures, black and brown indicate forest and domestic colonies, respectively, except K14 in brown with black outline to represent its mixed morphology.

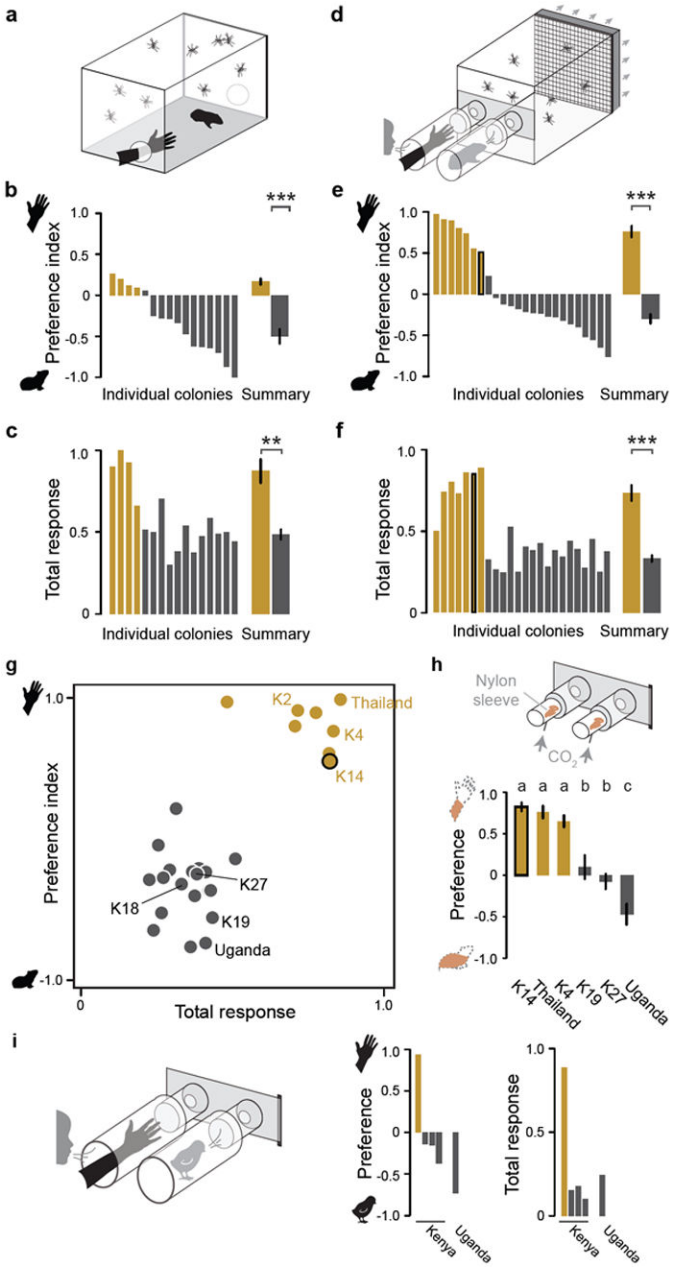


Figure 2. Forest and domestic females differ in host preference

a-c, Biting assay. **a**, Assay schematic. **b-c**, Host preference (**b**) and total response of mosquitoes choosing either host (**c**) of 16 Kenyan colonies (n =1 - 3 trials/colony). **d-f**, Live host olfactometer assay. **d**, Assay schematic. **e-f**, Host preference (**e**) and total response of mosquitoes choosing either host (**f**) of 24 Kenyan colonies (n = 1 - 7 trials/colony). Thin bars in **b, c, e**, and **f** are colony means and thick bars are summary mean of individual colony means ± s.e.m. (two-sample *t*-test treats each colony as single data point). **g**, Summary of live host olfactometer behaviour of 24 Kenyan plus Thai and Ugandan colonies. **h**, Nylon sleeve olfactometer assay. Top: Schematic of stimulus ports. Bottom: Preference of 6 colonies (mean ± s.e.m., n = 5-17 trials/colony). Bars labelled with different

letters are significantly different (one-way ANOVA $P < 0.0001$, followed by Tukey's test). **i**, Live host olfactometer assay (human *versus* chicken). Top: Schematic of host ports. Bottom: Preference and response of 5 colonies (n = 1 trial/colony). In all figures, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

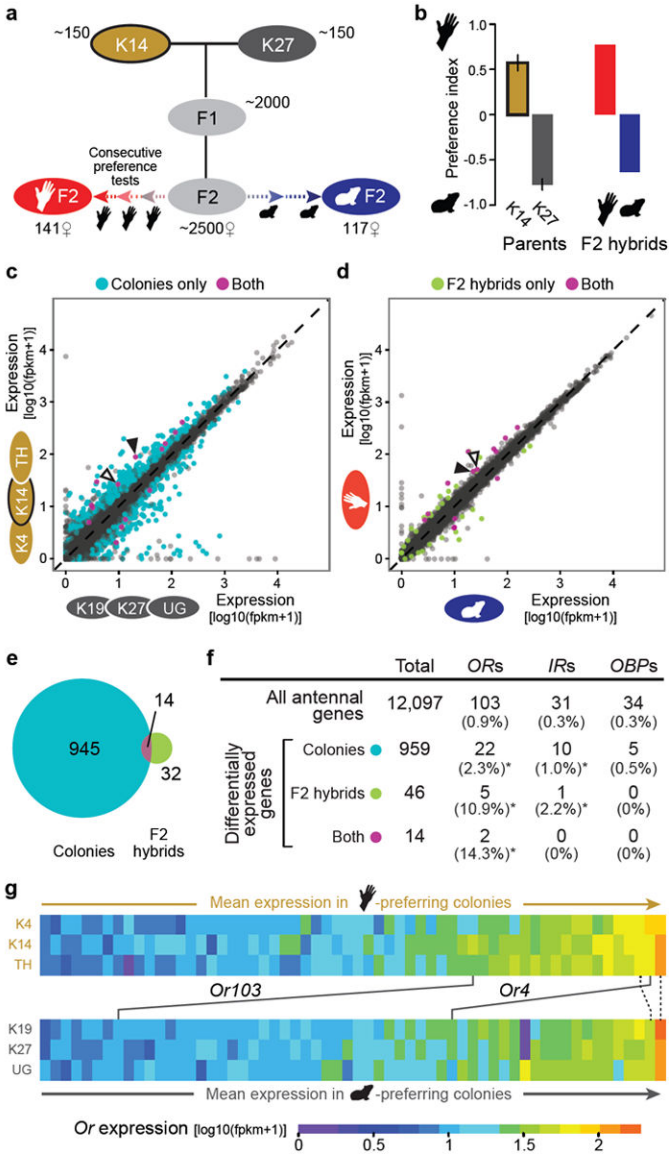


Figure 3. Antennal gene expression is significantly associated with preference for humans

a, Schematic of isolation of strongly human- and guinea pig-preferring F2 females. **b**, Live host olfactometer preference of parent colonies (n = 4 - 7 trials, mean ± s.e.m.) and F2 hybrids (n = 1). **c-d**, Differential antennal gene expression in colonies (**c**) and F2 hybrids (**d**) measured via RNAseq (FDR = 0.05). Filled and open arrowheads point to *Or4* and *Or103*, respectively. **e**, Summary of differential gene expression in **c-d. f**, Numbers of detectable (fpkm≥ 1) (top row) or differentially expressed (bottom 3 rows) olfactory receptors (ORs), ionotropic receptors (IRs), and odorant-binding proteins (OBPs) (% of total genes shown in parentheses). Asterisks indicate significant enrichment relative to % of all detectable genes (χ^2 $P < 0.0001$). **g**, Heat maps of ligand-selective *OR* genes with log10fpkm expression > 0.83 in domestic (top) or > 0.84 in forest (bottom) colonies, with two *ORs* differentially (solid lines) or not differentially expressed (dashed lines) indicated.

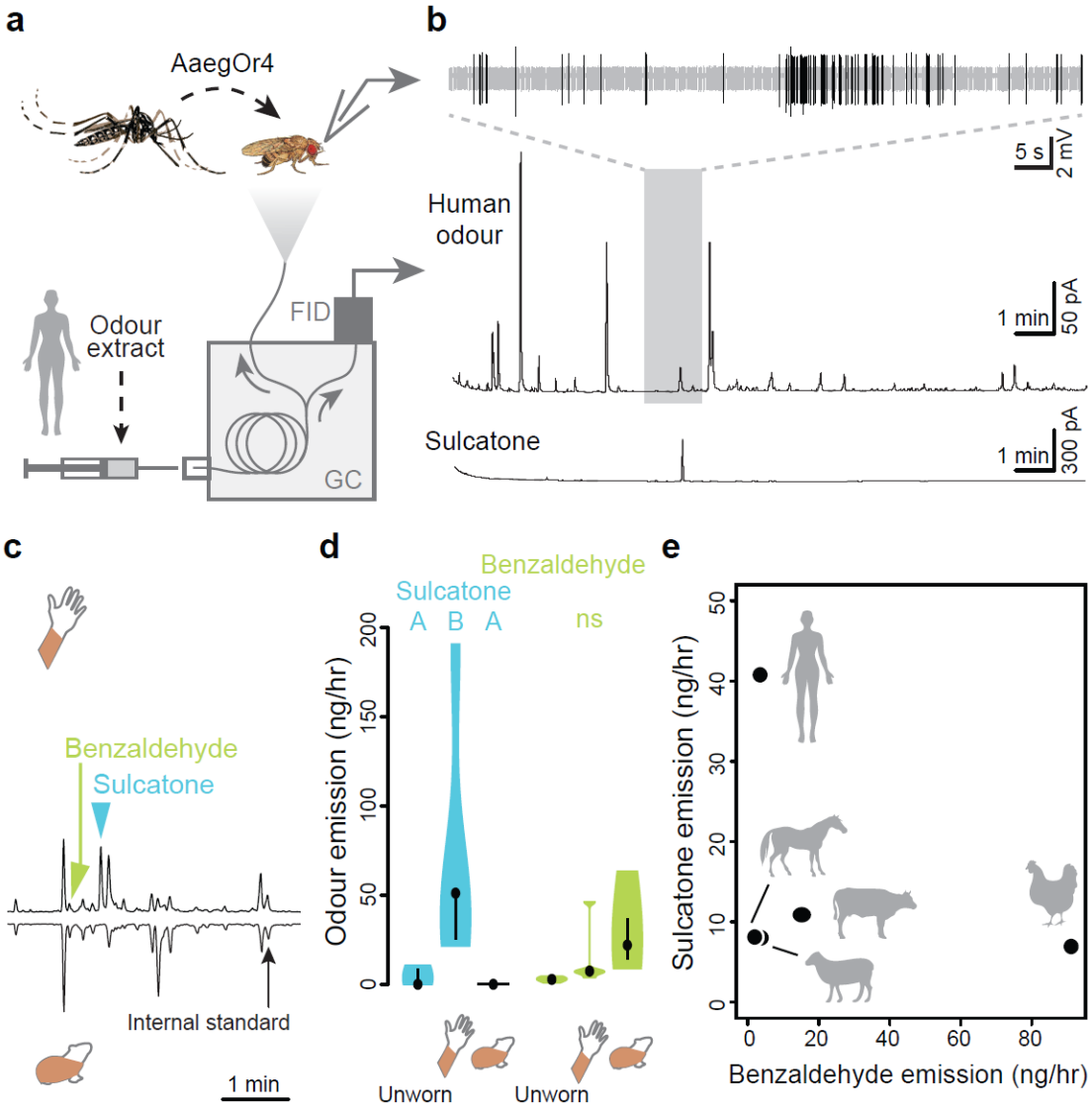


Figure 4. *Or4* responds to sulcatone, a human odorant

a, Schematic of gas chromatography-coupled single sensillum recording (GC-SSR). FID = flame ionization detector. **b**, Sample GC-SSR output, with SSR trace at the top showing large-amplitude spikes (black) from the *Or4*-expressing neuron and FID output showing elution times of human volatiles (middle trace) and sulcatone (bottom trace). **c**, Magnified sections of sample total ion chromatogram traces from GC-mass spectrometry (GC-MS) analysis of human (top) and guinea pig (bottom) sleeves. **d**, Summary of sulcatone and benzaldehyde emission from unworn, human, and guinea pig sleeves (n = 4 - 5 replicates per group; violin plots clipped at the range of data; median: black dot; 1st and 3rd quartile: bounds of black bar). Violins labelled with different letters are statistically different (pairwise Wilcoxon tests with Bonferroni correction; ns = not significant). **e**, Summary of sulcatone and benzaldehyde levels in human and animal odour collections (n = 1 collection from 1 - 20 individuals per species).

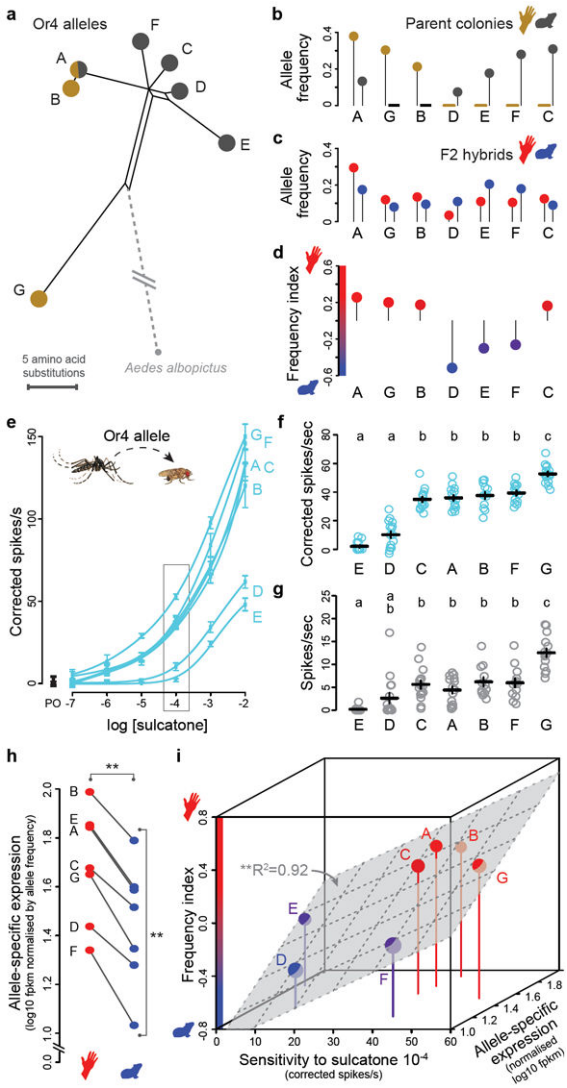


Figure 5. Tight linkage of *Or4* allelic expression and sulcatone sensitivity to preference for humans

a, Protein haplotype network showing seven *Or4* alleles segregating within the K27 forest (black) and K14 domestic (brown) parent colonies with a related *Ae. albopictus* gene shown for reference. Allele A is the reference allele tested in Fig 4a-b. **b-c**, Frequency of *Or4* alleles in K14 and K27 parent colonies (n = 34) (**b**) and human- and guinea pig-preferring F2 hybrids (n = 100) (**c**). In **c**, *Or4* alleles were inherited at significantly different frequencies by F2s with different host preference (χ^2; $P = 0.0002$). **d**, Relative frequency of *Or4* alleles in human- and guinea pig-preferring F2 hybrids. Data replotted from **c**. Index equal to the difference in frequency divided by the sum of the frequencies. Index bounds of -1 and 1 indicate the allele was only present in guinea pig or human-preferring F2s, respectively. **e**, Sulcatone dose-response curves, with mean ± s.e.m at each concentration. PO, paraffin oil. **f-g**, Summary of responses to 10^{-4} sulcatone (**f**) and spontaneous activity (**g**). Alleles marked with different letters are significantly different (one-way ANOVA $P < 0.0001$ for both variables followed by Tukey's test; n = 13 - 17 sensilla/allele). **h**, Allele-

specific expression derived from human- and guinea pig-preferring F2 hybrids. Variation was significant both between F2 groups (two-way ANOVA; $P = 0.004$), and among alleles (two-way ANOVA; $P = 0.005$). **i**, Joint effects of allele-specific ligand-sensitivity and expression on the F2 frequency index. Multiple regression: $P = 0.006$ for ligand-sensitivity, $P = 0.01$ for expression, $R^2 = 0.92$.