Arnal et al. BMC Research Notes 2014, 7:321
http://www.biomedcentral.com/1756-0500/7/321

BMC Research Notes

RESEARCH ARTICLE Open Access

Genetic polymorphism at an odorant receptor gene (Or39) among mosquitoes of the *Anopheles gambiae* complex in Senegal (West Africa)

Audrey Arnal[1*], Pierre Kengne[1,2], Cecile Brengues[1], Kounbobr Roch Dabire[3], Abdoulaye Diabate[3], Hubert Bassene[4] and Frederic Simard[1]

Abstract

Background: Olfaction plays a significant role in insect behavior during critical steps of their life-cycle, such as host-seeking during foraging or the search for a mate. Here, we explored genetic polymorphism within and divergence between sibling species of the African malaria mosquito, *Anopheles gambiae sensu lato* in the gene sequence and encoded peptides of an odorant receptor, Or39. This study included sympatric specimens of *An. gambiae sensu stricto, An. coluzzii* and *An. arabiensis* sampled together in the village of Dielmo, Senegal.

Results: A 1,601 bp genomic sequence composed of 6 exons and 5 introns was obtained for Or39 from 6–8 mosquitoes in each of the 3 species. DNA sequence analysis revealed a high level of molecular polymorphism ($\pi = 0.0154$; Haplotype diversity = 0.867) and high overall genetic differentiation between taxa ($Fst > 0.92$, $P < 0.01$). In total, 50 parsimony informative sites were recorded. Throughout the whole dataset, there were 13 non-synonymous mutations resulting in aminoacid changes in the encoded protein. Each of the 6 different identified peptides was species-specific and none was shared across species. Most aminoacid changes were located on the intracellular domains of the protein. However, intraspecific polymorphisms in *An. gambiae* and *An. arabiensis* as well as species-specific mutations also occurred in the first extracellular domain.

Conclusions: Although obtained from a limited number of specimens, our results point towards genetic differences between cryptic species within the *An. gambiae* complex in a gene of biological relevance that might be of evolutionary significance when exposed to disruptive selective forces.

Keywords: Mosquito, Malaria, *Anopheles gambiae*, Speciation, Olfactory receptor

Background

The *Anopheles gambiae sensu lato* complex groups together 8 sibling species, including 2 of the most powerful African human malaria vectors, *An. arabiensis* (Patton, 1905) and *An. gambiae sensu stricto* (Gile, 1902). The latter comprises 2 incipient species named *An. gambiae* (formerly, *An. gambiae* S form) and *An. coluzzii* (formerly, *An. gambiae* M form) which are genetically and biologically isolated from one another through assortative mating [1-6]. Population genetics and genomics studies revealed little genetic divergence throughout the genome, except in a few discrete regions with low recombination, a pattern that is compatible with retention of ancestral polymorphism after recent speciation and/or non negligible levels of residual gene flow between incipient species [3,7-11]. Indeed, although hybrids are rarely observed in the wild, no intrinsic fitness reduction was found when hybrids are artificially produced in the laboratory [12]. Thus, prezygotic barriers are believed to play a major role in fostering divergence between *An. coluzzii* and *An. gambiae* through strong assortative mating. These mosquitoes mate in swarms and further

* Correspondence: au.arnal@gmail.com
[1]Institut de Recherche pour le Développement (IRD), UMR IRD 224-CNRS 5290-UM1-UM2 MIVEGEC (Maladies Infectieuses et Vecteurs: Ecologie, Genetique, Evolution et Contrôle), team BEES (Biology, Ecology and Evolution of vector Systems), 911 Avenue Agropolis, BP 64501, Montpellier, cedex 5 34394, France
Full list of author information is available at the end of the article

BioMed Central

© 2014 Arnal et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.

studies using the previous molecular forms' nomenclature have shown that swarming and mating mainly involved mosquitoes of the same form [13-15] and that males and females engage in close-range acoustic interactions by shifting their flight tones to match each other prior to copulation [16]. However, the cues perceived by virgin mosquitoes for long distance orientation towards mating areas in search for a mate are still largely unknown. Close-range interactions between potential mating pairs are still incompletely understood as well. These processes may be mediated by volatile compounds like pheromones, which were shown to be involved in the mating behavior of several biting flies [17,18]. The olfactory system, including odorant receptors, would therefore play a critical role in detecting biologically active compounds such as volatiles emanating from potential mates, other insects (conspecifics, predators), hosts or candidate oviposition sites [19-21]. Odorant receptor (OR) genes have been identified in genomic regions of high differentiation between *An. gambiae* and *An. coluzzii* [3,22-24]. One of these OR genes, AgOr39 [AGAP 002639 in VectorBase, thereafter Or39] located on chromosome 2R, is highly polymorphic. Recent analyses of a c.a. 400 bp partial sequence of the gene suggested directional selection acting on this gene in Cameroon's *An. coluzzii*, where reproductive isolation between *An. gambiae* and *An. coluzzii* is highest [8,9,24]. This pattern was not observed in specimens collected from Mali [24]. Therefore, the authors suggested that this locus might be related to specific processes of ecological divergence prompting assortative mating among *An. gambiae* and *An. coluzzii* sympatric populations in only a limited area within their overlapping distribution range [24]. Different populations of the 2 incipient species are indeed known to exhibit different levels of reproductive isolation in different geographic locations across their range [10,11,25].

Here, we provide new data on Or39 molecular polymorphism by i) expanding the breadth of the sequencing effort to cover a larger portion of the gene, including 6 exons and 5 introns, and ii) exploring molecular polymorphism and divergence in this gene among specimens of *An. gambiae*, *An. coluzzii* and *An. arabiensis* collected in Senegal (West Africa), a geographic area where the 3 species co-exist and where genetic admixture between *An. gambiae* and *An. coluzzii* has been reported [11,25,26].

Methods

Study site and mosquitoes

Fieldwork was carried out in agreement with the procedures of the National Ethics Committee in Senegal (clearance N° 1971 MPM/DS/DER). Our study complies with the Convention on Biological Diversity (http://www.cbd.int/convention/) and the Convention on International Trade in Endangered Species of wild fauna and flora threatened with extinction (http://www.cites.org/).

Mosquitoes used for the study were collected by landing catches on adult volunteers between July and October 2007 (rainy season) in the village of Dielmo (13°45′N, 16° 25′W) situated 285 km southeast of Dakar (Senegal) and about 10 km north of the Gambian border. Dielmo is a village of about 350 inhabitants bordered by a semi-permanent freshwater river. In Dielmo, *An. arabiensis*, *An. coluzzii* and *An. gambiae s.s.* occur together all year round, although with seasonal fluctuation in their relative abundance and prevalence of *An. gambiae*/*An. coluzzii* hybrids reaching 3% during the rainy season [26]. After collection, mosquitoes were identified in the field as *An. gambiae s.l.* using morphological keys [27,28]. Only females of *An. gambiae s.l.* were included in subsequent analyses.

DNA extraction and molecular identification

Genomic DNA was extracted from legs or whole mosquito bodies using Cetyl Trimethyl Ammonium Bromide (CTAB) following the protocol of Morlais *et al.* [29]. DNA pellets were dried out and re-suspended in 20 μl of nuclease-free water and stored at −20°C. Mosquito species within the *An. gambiae s.l.* complex were identified using a standard PCR-RFLP protocol for molecular identification of *An. gambiae s.s.*, *An. coluzzii*, *An. arabiensis* as well as F1 hybrid specimens [30].

PCR amplification and sequencing of Or39

PCR reactions were carried out using 10–20 ng of template DNA in 25 μl reaction containing 0.5 Unit Taq polymerase (Quiagen, Courteboeuf, France) in manufacturer's buffer, 1.5 mM MgCl2, 200 μM each dNTP (PE Applied Biosystems) and 10 pmol each forward and reverse primers.

Primers were designed from the *An. gambiae s.s.* gene annotation AGAP002639 in VectorBase (https://www.vectorbase.org/) using Primer3 software [31]. A 1.6 kb region encompassing the whole transcribed region of the Or39 gene was amplified using forward primer Or39F (5′-GGTGCTGCAGCTTCTAATC-3′) and reverse primer Or39R (5′-CAAAAAGGACTTCATCAGTG-3′). Cycling conditions for amplification included denaturation at 94°C for 5 min, followed by 35 cycles at 94°C for 30 s, 50°C for 30 s and 72°C for 1 min, with a final extension step at 72°C for 7 min. PCR products were examined on a 1.5% agarose gel, and cloned using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen, Paisley, UK). Individual transformed colonies were selected and fragments of the appropriate size were sequenced using PE BigDye Terminator Ready Reaction Kit (PE Applied Biosystems) on an ABI 3130XL apparatus (Applied Biosystems, France)

according to manufacturer's instructions. To avoid sampling bias, a single allele (haplotype sequence) was arbitrarily selected from each specimen for analysis. Furthermore, 2 to 3 clones of the expected size were randomly sequenced per specimen, in order to check for PCR accuracy prior to cloning through double sequencing of the same allele (see [32] for more details).

Sequence analysis

Sequences were visually checked and aligned using Seqscape software (Applied Biosystems, France). Aligned DNA sequences were imported into MEGA6 [33] to compute basic sequence statistics, assess genetic distance between groups and build a Neighbour-Joining tree. To determine whether signatures of natural selection were present in our dataset, we analyzed the relative rates of non-synonymous and synonymous substitution (dN/dS) and performed the MacDonald-Krietman test of neutral evolution as implemented in MEGA6. Genetic differentiation between mosquito populations was assessed by Fst estimates and tested by permutation tests in MEGA6. The hydrophobic regions on the encoded receptor protein were identified using Kyte-Doolittle hydropathy algorithm [34] as implemented through the website of the Pasteur Institute (http://mobyle.pasteur.fr, TopPred 1.10, [35]).

Results

Species diversity and sequence analysis

Field collections returned 224 females of *Anopheles s.l.* among which 42% (N = 94) were *An. arabiensis*, 39% (N = 87) were *An. gambiae*, 11% (N = 24) were *An. coluzzii* and, 1% (N = 2) were *An. gambiae*/*An. coluzzii* F1 hybrids. The remaining 7% (N = 17) failed to amplify and were not considered in the study. All 3 species were found together in the collections from July to September, during the peak of the malaria transmission period. The 2 hybrid specimens were collected in July, and no *An. coluzzii* specimen was collected in October and November.

A 1,608-bp DNA sequence was obtained for the entire Or39 gene from 7 *An. gambiae* and 6 *An. coluzzii* specimens. The sequences matched the reference genomic sequence published from the PEST (Pink Eye STandard) strain of *An. gambiae* s.s. [AGAP002639] in VectorBase. The exon/intron structure of the gene was determined by reference to the cDNA sequence available in VectorBase (Transcript ID: AGAP0002639-RA). Full-length Or39 sequences consisted of 6 exons separated by 5 non-coding introns. Sequences obtained from 8 *An. arabiensis* specimens were shorter, due to a 7-bp deletion shared by all specimens, located in a non-coding region of the gene (Intron 1, between positions 430–437). Only partial sequence fragments were retrieved from the two hybrid specimens, which were not included in further analyses.

Genetic polymorphism and divergence

DNA polymorphism analysis identified 68 (4.25%) variable sites along the sequence, of which 50 were parsimony informative (Figure 1). Average nucleotide diversity was lowest in *An. coluzzii* ($\pi = 0.00033$) and one order of magnitude higher in *An. gambiae* and *An. arabiensis* ($\pi = 0.0012$ and 0.0025, respectively) (Table 1). However, haplotype diversity was lower in *An. arabiensis* (Hd = 0.464) than in *An. gambiae* and *An. coluzzii* (Hd = 0.533 and 0.714, respectively). Throughout the whole gene region, there were 33 SNPs differences between *An. gambiae* and *An. arabiensis,* and 35 between *An. arabiensis* and *An. coluzzii*. These mutations were located

Haplotype	n	Base Number																																																	
		81	114	124	145	309	347	378	385	386	396	404	413	415	425	445	455	461	481	533	557	635	638	661	663	665	668	687	694	695	699	706	729	752	756	764	773	776	794	803	920	986	995	998	1145	1251	1339	1355	1394	1507	1554
AGAP002639		A	G	G	G	T	G	G	C	A	G	A	A	C	A	C	T	C	C	T	T	G	T	G	G	C	A	A	C	C	A	G	A	C	C	G	C	G	G	C	A	C	G	G	C	G	A	A	T	A	A
An. gambiae H1	4	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	.	.	C	A	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	C	A	.	.	.	.	.	.	G	.	G	.	.
An. gambiae H2	1	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	T	.	C	A	.	.	.	.	.	G	T	.	.	.	.	.	.	.	.	.	C	A	.	.	.	.	G	.	G	.	G	.	.
An. gambiae H3	1	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	T	.	C	A	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	C	A	.	.	.	.	G	.	G	.	G	.	.
An. gambiae H4	1	T	.	.	.	.	C	.	.	.	.	.	.	.	.	.	A	.	T	.	C	A	.	.	.	.	.	G	T	.	.	.	.	.	.	.	.	.	C	A	.	.	.	.	G	.	G	.	G	.	.
An. coluzzii H1	2	C	.	A	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	C	A	.	T	C	.	.	A	.	G	.	G	G
An. coluzzii H2	4	C	.	A	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	C	C	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	C	A	.	T	C	.	.	A	.	G	.	G	G
An. arabiensis H1	6	T	.	.	A	C	.	.	T	G	T	T	T	.	C	T	A	G	.	C	.	A	C	A	T	T	G	.	.	T	C	A	C	A	T	A	.	T	.	.	T	.	.	T	.	.	.	G	.	.	G
An. arabiensis H2	1	T	A	.	A	C	.	A	T	G	T	T	T	.	C	T	A	G	.	C	.	A	C	A	T	T	G	.	.	T	C	A	C	A	T	A	.	T	.	.	T	.	.	T	.	.	.	G	.	.	G
An. arabiensis H3	1	T	A	.	A	C	.	A	T	G	T	T	T	.	C	T	A	G	.	C	.	A	C	A	T	T	G	.	.	T	C	A	C	A	T	A	.	T	.	.	T	T	.	.	.	.	.	.	.	.	G

Figure 1 Summary sequence alignment of the different Or39 haplotypes detected in mosquito specimens collected in the locality of Dielmo in Senegal during the 2007 rainy season. Only parsimony informative sites are shown (N = 50). Nucleotides are numbered according to the published reference sequence from the PEST strain of *An. gambiae* (VectorBase, AGAP002639). Dots indicate identity to the reference sequence. The coding regions (exons) are shaded.

Table 1 Summary statistics for polymorphism in the Or39 receptor gene within sympatric populations of *An. gambiae, An. coluzzii* and *An. arabiensis* from Dielmo, Senegal

Population	n	size (bp)	s	h	π	Hd	D
An. gambiae	7	1601	4	4	0.0012	0.714	0.79 ns
An. coluzzii	6	1601	1	2	0.0003	0.533	0.85 ns
An. arabiensis	8	1601	13	3	0.0025	0.464	−1.06 ns
All	21	1601	61	9	0.0154	0.867	1.23 ns

n, number of DNA sequences; s, number of segregating sites; h, number of haplotypes based on the number of segregating sites; π, nucleotide diversity; Hd, haplotype diversity; D, Tajima's (1989) statistic; ns, not significant.

in both exon and intron domains (Figure 1). A total of 16 SNPs were observed between *An. gambiae* and *An. coluzzii*, including 11 in exons. Four of these were replacement mutations. Within the coding region (Total exon length = 1,224 bp), non-synonymous diversity (dN, π nonSyn = 0.005) was 6 times lower than synonymous diversity (dS, π Syn = 0.029), although the difference was not significant (P> > 0.05, two-sided Fisher test of neutral evolution available in MEGA6). There was no evidence for deviation from neutral expectations in any species (McDonald-Kreitman test, P > 0.39).

At the DNA level, divergence between the 3 species was highly significant (P < 0.01) with Fst estimates above 0.92 across the whole gene.

Phylogenetic relationships among taxa

Figure 2 shows a Neighbour-Joining tree constructed with Kimura-2-parameter genetic distances, retained as the best model based on the Bayesian Information Criterion (BIC) scores used in MEGA6, between the 10 haplotypes shown in Figure 1. Specimens segregated unambiguously into the 3 known taxa, *An. gambiae*, *An. coluzzii* and *An. arabiensis*. Note that the PEST strain sequence maps as expected, in between the *An. gambiae* and *An. coluzzii* clusters. The PEST genome is indeed known to be a composite genome of the two cryptic species.

Peptide analysis

The open reading frame of the 1400 bp cDNA encoded a 407 amino acid sequence for Or39. Throughout the whole dataset, there were 13 non-synonymous replacement mutations resulting in aminoacid changes in the encoded protein. The amino acid sequences obtained from deducted cDNA sequences resulted in 6 distinct peptides: 3 in *An. arabiensis*, 2 in *An. gambiae* and a single peptide encoded by *An. coluzzii* DNA (Figure 3). Note that all peptides identified from wild mosquito specimens were distinct from the protein sequence deducted from the PEST strain DNA. Moreover, each peptide was species-specific and they were not shared across species. There were 4 amino acid changes between *An. gambiae* and *An. coluzzii* resulting in a Methionin-to-Valine substitution at position 42, a Serine-to-Threonine substitution at position 116, a Histidine-to-Arginine substitution at position 339 and a Glutamine-to-Glycine substitution at position 343 (Figure 3).

Structural analysis of the encoded protein

The structural analysis of the encoded protein explored through the Kyte-Doolittle hydropathy algorithm returned a single most likely topology for Or39 with 7 transmembrane domains that were reported at the same position in the 3 species (Figure 3). Given the known membrane conformation for odorant receptors with the N-terminus internal and the C-terminus external [37], the topology with 4 intra-cellular domains and 4 extra-cellular domains is most likely. Accordingly, most aminoacids changes on the receptor were located on the intracellular domains of the protein. However, intraspecific

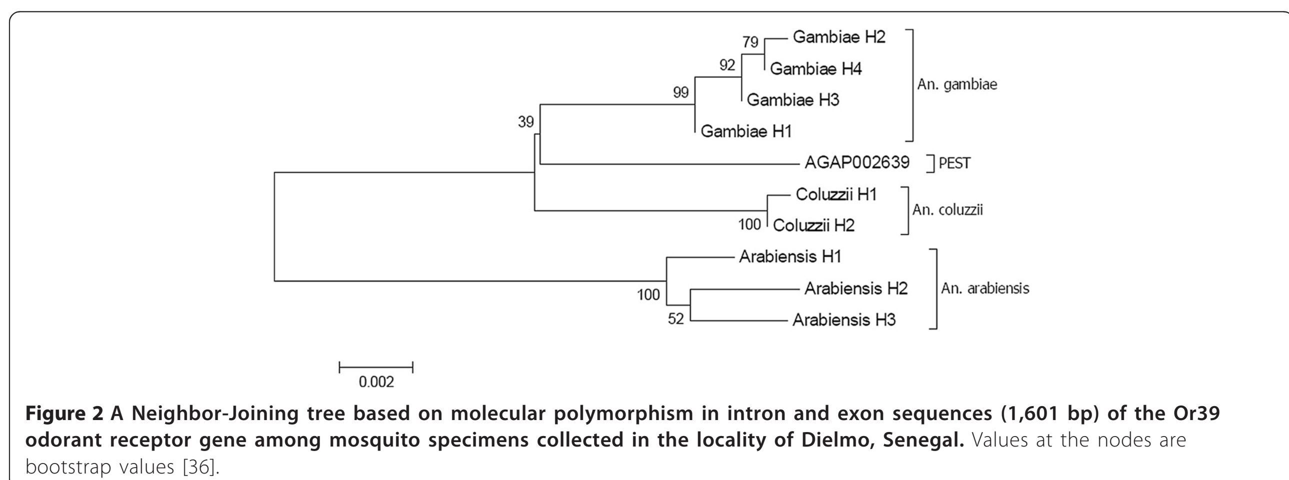

Figure 2 A Neighbor-Joining tree based on molecular polymorphism in intron and exon sequences (1,601 bp) of the Or39 odorant receptor gene among mosquito specimens collected in the locality of Dielmo, Senegal. Values at the nodes are bootstrap values [36].

```
                     1                                                           60
An. gambiae_P1       MVSFGAAASN PRDAETTPQW DIFKLQRKIL LIFGVWPADR LVRPWYVKVL IAINLATLAI
An. gambiae_P2       MVSFGAAASN PRDAETTPQW DIFKLQRKIL LIFGVWPADR LVRPWYVKVL IAINLATLAI
An. coluzzii         MVSFGAAASN PRDAETTPQW DIFKLQRKIL LIFGVWPADR LMRPWYVKVL IAINLATLAI
An. arabiensis_P1    MVSFGAAASN PRDAETTPQW DIFKLQRKIL LIFGVWPADR LVRPWYVKML IAINLATLAI
An. arabiensis_P2    MVSFGAAASN PRDAETTPQW DIFKLQRKIL LIFGVWPADR LVRPWYVKML IAINLATLAI
An. arabiensis_P3    MVSFGAAASN PRDAETTPQW DIFKLQRKIL LIFGVWPADR LVRPWYVKML IAINLATLAI
AGAP002639           MVSFGAAASN PRDAETTPHW DIFKLQRKIL LIFGVWPADR LVRPWYVKVL IAINLATLAI
                     ********** ********:* ********** ********** *:******:* **********

                     61                                                          120
An. gambiae_P1       CMVGEFLHGL YAYQEGDLSE SIESICPTVA RISGFLRMVF YLANEEKIHR VLNNITKTLQ
An. gambiae_P2       CMVGEFLHGL YAYQEGDLSE SIESICPTVA RISGFLRMVF YLANEEKIHR VLNNITKTLQ
An. coluzzii         CMVGEFLHGL YAYQEGDLSE SIESICPTVA RISGFLRMVF YLANEEKIHR VLNNISKTLQ
An. arabiensis_P1    CMVGEFLHGL YAYQEGDLSE SIESICPTVA RISGFLRMVF YLANEEKIHR VLNNISKTLQ
An. arabiensis_P2    CMVGEFLHGL YAYQEGDLSE SIESICPTVA RISGFLRMVF YLANEEKIHR VLNNISKTLQ
An. arabiensis_P3    CMVGEFLHGL YAYQEGDLSE SIESICPTVA RISGFLRMVF YLANEEKIHR VLNNISKTLQ
AGAP002639           CMVGEFLHGL YAYQEGDLSE SIESICPTVA RISGFLRMVF YLANEEKIHR VLNNISKTLQ
                     ********** ********** ********** ********** ********** *****:****

                     121                                                         180
An. gambiae_P1       DEHPREHTIT KQLTTLGQQF TFYLFLMMFF AACLYGVTPF FIMIYNWVQG QRPLVKLLPF
An. gambiae_P2       DEHPREHIIT KQLTTLGQQF TFYLFLMMFF AACLYGVTPF FIMIYNWVQG QRPLVKLLPF
An. coluzzii         DEHPREHTIT KQLTTLGQQF TFYLFLMMFF AACLYGVTPF FIMIYNWVQG QRPLVKLLPF
An. arabiensis_P1    EEHPREHTIT KQLTTLGQQF TFYLFLMMFF AACLYGVTPF FIMIYNWVQG QRPLVKLLPF
An. arabiensis_P2    EEHPREHTIT KQLTTLGQRF TFYLFLMMFF AACLYGVTPF FIMIYNWVQG QRPLVKLLPF
An. arabiensis_P3    EEHPREHTIT KQLTTLGQQF TFYLFLMMFF AACLYGVTPF FIMIYNWVQG QRPLVKLLPF
AGAP002639           DEHPREHTIT KQLTTLGQQF TFYLFLMMFF AACLYGVTPF FIMIYNWVQG QRPLVKLLPF
                     :******:** ********:* ********** ********** ********** **********

                     181                                                         240
An. gambiae_P1       KLALPFDSQN LFHFVVTTLF LNYASAPTIT SQSGSDALFS GVCLYIYGQF QAIRLEVEAL
An. gambiae_P2       KLALPFDSQN LFHFVVTTLF LNYASAPTIT SQSGSDALFS GVCLYIYGQF QAIRLEVEAL
An. coluzzii         KLALPFDSQN LFHFVVTTLF LNYASAPTIT SQSGSDALFS GVCLYIYGQF QAIRLEVEAL
An. arabiensis_P1    KLALPFDSQH LFHFVVTTLF LNYASAPTIT SQSGSDALFS GVCLYIYGQF QAIRLEVEAL
An. arabiensis_P2    KLALPFDSQH LFHFVVTTLF LNYASAPTIT SQSGSDALFS GVCLYIYGQF QAIRLEVEAL
An. arabiensis_P3    KLALPFDSQH LFHFVVTTLF LNYASAPTIT SQSGSDALFS GVCLYIYGQF QAIRLEVEAL
AGAP002639           KLALPFDSQN LFPFVVTTLF LNYASAPTIT SQSGSDALFS GVCLYIYGQF QAIRLEVEAL
                     *********: **:******* ********** ********** ********** **********

                     241                                                         300
An. gambiae_P1       AETVDRNTLK SSAAETHRIN LELRRISKRH QAIIDLVSEV RSAFTPNVLL VYTATAIIMC
An. gambiae_P2       AETVDRNTLK SSAAETHRIN LELRRISKRH QAIIDLVSEV RSAFTPNVLL VYTATAIIMC
An. coluzzii         AETVDRNTLK SSAAETHRIN LELRRISKRH QAIIDLVSEV RSAFTPNVLL VYTATAIIMC
An. arabiensis_P1    AETVDRNTLK SSAAETHRIN LELRRISKRH QAIIDLVSDV RSAFTPNVLL VYTATAIIMC
An. arabiensis_P2    AETVDRNTLK SSAAETHRIN LELRRISKRH QAIIDLVSEV RSAFTPNVLL VYTATAIIMC
An. arabiensis_P3    AETVDRNTLK SSAAETHRIN LELRRISKRH QAIIDLVSDV RSAFTPNVLL VDTATAIIMC
AGAP002639           AETVDRNTLK SSAAETHRIN LELRRISKRH QAIIDLVSEV RSAFTPNVLL VYTATAIIMC
                     ********** ********** ********** ********:* ********** *:********

                     301                                                         360
An. gambiae_P1       IVCIAMLVVE GIYKLTYLPY AFAELTLLFL YSYSGTIIRD SSGAVQTVAY DFPWYRYDRN
An. gambiae_P2       IVCIAMLVVE GIYKLTYLPY AFAELTLLFL YSYSGTIIRD SSGAVQTVAY DFPWYRYDRN
An. coluzzii         IVCIAMLVVE GIYKLTYLPY AFAELTLLFL YSYSGTIIHD SSEAVQTVAY DFPWYRYDRN
An. arabiensis_P1    IVCIAMLVVE GIYKLTYLPY AFAELTLLFL YSYSGTIIRD SSEAVQTVAY DFPWYRYDRN
An. arabiensis_P2    IVCIAMLVVE GIYKLTYLPY AFAELTLLFL YSYSGTIIRD SSEAVQTVAY DFPWYRYDRN
An. arabiensis_P3    IVCIAMLVVE GIYKLTYLPY AFAELTLLFL YSYSGTIIRD SSEAVQTVAY DFPWYRYDRN
AGAP002639           IVCIAMLVVE GIYKLTYLPY AFAELTLLFL YSYSGTIIRD SSEAVQTVAY DFPWYRYDRN
                     ********** ********** ********** ********:* **:******* **********

                     361                                            407
An. gambiae_P1       TRHLIQMMMI RAQYGSNVDV PFFETSMASF SAIVRTASSY ITLMKSFL
An. gambiae_P2       TRHLIQMMMI RAQYGSNVDV PFFETSMASF SAIVRTASSY ITLMKSFL
An. coluzzii         TRHLIQMMMI RAQYGSNVDV PFFETSMASF SAIVRTASSY ITLMKSFL
An. arabiensis_P1    TRHLIQMMMI RAQYGSNVDV PFFETSMASF SAIVRTASSY ITLMKSFL
An. arabiensis_P2    TRHLIQMMMI RAQYGSNVDV PFFETSMASF SAIVRTASSY ITLMKSFL
An. arabiensis_P3    TRHLIQMMMI RAQYGSNVDV PFFETSMASF SAIVRTASSY ITLMKSFL
AGAP002639           TRHLIQMMMI RAQYGSNVDV PFFETSMASF SAIVRTASSY ITLMKSFL
                     ********** ********** ********** ********** ********
```

Figure 3 (See legend on next page.)

(See figure on previous page.)
Figure 3 Alignment of the predicted amino acid sequences for Or39 odorant receptor of *An. gambiae, An. coluzzii* and *An. arabiensis* from Dielmo, Senegal. Reference amino acid sequence for the PEST strain of *An. gambiae* was obtained from the published cDNA sequence (Accession N° XM_312289.1). Shaded areas represent hydrophobic trans-membrane domains of the protein. Stars denote identity to the reference sequence and dots show amino acid changes. Within-species polymorphisms are bolded and boxed; Polymorphic sites among species are highlighted in reversed fonts.

polymorphisms in *An. gambiae* and *An. arabiensis* as well as species-specific mutations also occur in the first (long) extracellular domain (Figure 3).

Discussion and conclusion

The study of reproductive behavior in *An. gambiae s.l.* is of major importance in the understanding of how the main vectors of malaria in Africa have evolved and how selective pressures operate to foster divergence and/or gene flow within as well as among these cryptic taxa.

In the village of Dielmo in Senegal, *An. arabiensis, An. coluzzii* and *An. gambiae s.s.* occur in sympatry throughout the year, although with seasonal fluctuation in their relative frequencies [27,38]. The prevalence of F1 hybrids between *An. gambiae* and *An. coluzzii* (ie, formerly referred to as M/S hybrids) was also shown to vary both geographically and temporally throughout their common distribution area in West and Central Africa [10,11,25,26]. The ecological determinants of such a dynamic hybridization process are currently unknown and may be under the control of a few genes mapping in areas of high genomic divergence between *An. coluzzii* and *An. gambiae* that have been called 'speciation islands' [2,3,22]. This is the case of Or39 lying in a genomic region of high differentiation on chromosome 2R. Indeed, our results detected high levels of genetic differentiation throughout this gene among the sympatric mosquito populations we have sampled, with fixed differences in coding regions resulting in amino-acid changes in the encoded protein. There was however no sign of diversifying selection acting on this gene, at least in our limited dataset which resulted in low statistical power of neutrality tests. Meanwhile, the three species explored in this study exhibited different mature peptides. Most of the aminoacid changes observed between species were located on the intracellular domains of the protein, and might indeed reflect random changes accumulated since lineage splitting within the *An. gambiae s.l.* complex. However, the nucleotide sequence of the first, and longest (*i.e.,* 72 aminoacids) extracellular domain of the protein was different between species, and polymorphic within the *An. gambiae* and *An. arabiensis* samples. At this stage of our analysis, we can only speculate on the role these amino acid changes may play for the perception of different olfactory stimuli by the mosquitoes through specific ligand-receptor interactions, thereby enhancing reproductive isolation when disruptive selective forces apply. The molecular mechanisms that are involved in such processes will have to be investigated, as well as their biological relevance to be evaluated through functional studies. Moreover, further knowledge on the mechanisms involved in mating behavior and the genetics and ecology of mate choice in this major disease vector is needed for the development of alternative vector control strategies based on population replacement/suppression.

Availability of supporting data

The data set supporting the results of this article is available in the treebase repository under reference 15820; http://purl.org/phylo/treebase/phylows/study/TB2:S15820.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

PK, AD, KRD and FS initiated and designed the study. HB and PK organized and conducted the field work. PK, CB and AA carried out lab work and analyzed data. AA, PK and FS wrote the manuscript. All authors read, revised and approved the final manuscript.

Acknowledgments

This study was financially supported by the Multilateral Initiative on Malaria (MIM) Project A60098 through the UNICEF/UNDP/World Bank/WHO Special Program for Research and Training in Tropical Diseases (TDR) to KRD and by the "Institut de Recherche pour le Developpement" (IRD), France. We thank Dr Antonio-Nkondjio Christophe for comments on an earlier version of this work. The authors extend their gratitude to authorities and the populations of Dielmo who made this study possible.

Author details

[1]Institut de Recherche pour le Développement (IRD), UMR IRD 224-CNRS 5290-UM1-UM2 MIVEGEC (Maladies Infectieuses et Vecteurs: Ecologie, Genetique, Evolution et Contrôle), team BEES (Biology, Ecology and Evolution of vector Systems), 911 Avenue Agropolis, BP 64501, Montpellier, cedex 5 34394, France. [2]Organisation de Coordination pour la lutte contre les Endémies en Afrique Centrale (OCEAC), P.O. Box 288, Yaoundé, Cameroon. [3]Institut de Recherche en Sciences de la Santé (IRSS), P.O. Box 545, Bobo-Dioulasso, Burkina Faso. [4]UMR 198 URMITE, Campus International IRD/UCAD de Hann, Dakar, Senegal.

Received: 17 December 2013 Accepted: 21 May 2014
Published: 30 May 2014

References

1. Coetzee M, Hunt RH, Wilkerson R, Della Torre A, Coulibaly MB, Bezansky N: ***Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex.** *Zootaxa* 2013, **3619**:246–274.
2. White BJ, Collins FH, Besansky NJ: **Evolution of *Anopheles gambiae* in Relation to Humans and Malaria**. In *Annual Review of Ecology, Evolution, and Systematics*, Volume 42. Edited by Futuyma DJ, Shaffer HB, Simberloff D.; 2011:111–132.

3. White BJ, Cheng C, Simard F, Costantini C, Besansky NJ: **Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*.** *Mol Ecol* 2010, **19:**925–939.
4. Lehmann T, Diabate A: **The molecular forms of *Anopheles gambiae*: a phenotypic perspective.** *Infect Genet Evol* 2008, **8:**737–746.
5. della Torre A, Tu ZJ, Petrarca V: **On the distribution and genetic differentiation of Anopheles gambiae s.s. molecular forms.** *Insect Biochem Mol Biol* 2005, **35:**755–769.
6. della Torre A, Fanello C, Akogbeto M, Dossou-yovo J, Favia G, Petrarca V, Coluzzi M: **Molecular evidence of incipient speciation within Anopheles gambiae s.s. in West Africa.** *Insect Mol Biol* 2001, **10:**9–18.
7. Stump AD, Fitzpatrick MC, Lobo NF, Traore S, Sagnon NF, Costantini C, Collins FH, Besansky NJ: **Centromere-proximal differentiation and speciation in *Anopheles gambiae*.** *Proc Natl Acad Sci U S A* 2005, **102:**15930–15935.
8. Weetman D, Wilding CS, Steen K, Pinto J, Donnelly MJ: **Gene flow–dependent genomic divergence between *Anopheles gambiae* M and S forms.** *Mol Biol Evol* 2012, **29:**279–291.
9. Reidenbach KR, Neafsey DE, Costantini C, Sagnon F, Simard F, Ragland GJ, Egan SP, Feder JL, Muskavitch MAT, Besansky NJ: **Patterns of genomic differentiation between ecologically differentiated M and S forms of *Anopheles gambiae* in West and Central Africa.** *Genome Biol Evol* 2012, **4:**1202–1212.
10. Lee Y, Marsden CD, Norrisa LC, Colliera TC, Maina BJ, Fofanad A, Cornel AJ, Lanzaro GC: **Spatiotemporal dynamics of gene flow and hybrid fitness between the M and S forms of the malaria mosquito, *Anopheles gambiae*.** *Proc Natl Acad Sci* 2013, **110:**19854–19859.
11. Nwakanma DC, Neafsey DE, Jawara M, Adiamoh M, Lund E, Rodrigues A, Loua KM, Konate L, Sy N, Dia I, Awolola TS, Muskavitch MA, Conway DJ: **Breakdown in the process of incipient speciation in *Anopheles gambiae*.** *Genetics* 2013, **193:**1221–1231.
12. Diabate A, Dabire RK, Millogo N, Lehmann T: **Evaluating the effect of postmating isolation between molecular forms of *Anopheles gambiae* (*Diptera : Culicidae*).** *J Med Entomol* 2007, **44:**60–64.
13. Diabaté A, Baldet T, Brengues C, Kengne P, Dabire KR, Simard F, Chandre F, Hougard JM, Hemingway J, Ouedraogo JB, Fontenille D: **Natural swarming behavior in the molecular M form of *Anopheles gambiae*.** *R Soc Trop Med Hyg* 2003, **97:**1–4.
14. Diabate A, Dabire RK, Kengne P, Brengues C, Baldet T, Ouari A, Simard F, Lehmann T: **Mixed swarms of the molecular M and S forms of *Anopheles gambiae* (*Diptera : Culicidae*) in sympatric area from Burkina Faso.** *J Med Entomol* 2006, **43:**480–483.
15. Diabate A, Dao A, Yaro AS, Adamou A, Gonzalez R, Manoukis NC, Traore SF, Gwadz RW, Lehmann T: **Spatial swarm segregation and reproductive isolation between the molecular forms of *Anopheles gambiae*.** *Proc R Soc B Biol Sci* 2009, **276:**4215–4222.
16. Pennetier C, Warren B, Dabire KR, Russell IJ, Gibson G: **"Singing on the wing" as a mechanism for species recognition in the malarial mosquito *Anopheles gambiae*.** *Curr Biol* 2010, **20:**278.
17. Cabrera M, Jaffe K: **An aggregation pheromone modulates lekking behavior in the vector mosquito *Aedes aegypti* (*Diptera : Culicidae*).** *J Am Mosq Control Assoc* 2007, **23:**1–10.
18. Yuval B: **Mating systems of blood-feeding flies.** In *Annual Review of Entomology*, Volume 51. ; 2006:413–440.
19. Fox AN, Pitts RJ, Robertson HM, Carlson JR, Zwiebel LJ: **Candidate odorant receptors from the malaria vector mosquito *Anopheles gambiae* and evidence of down-regulation in response to blood feeding.** *Proc Natl Acad Sci U S A* 2001, **98:**14693–14697.
20. Gimonneau G, Bouyer J, Morand S, Besansky NJ, Diabate A, Simard F: **A behavioral mechanism underlying ecological divergence in the malaria mosquito Anopheles gambiae.** *Behav Ecol* 2010, **21:**1087–1092.
21. Wang G, Carey AF, Carlson JR, Zwiebel LJ: **Molecular basis of odor coding in the malaria vector mosquito *Anopheles gambiae*.** *Proc Natl Acad Sci U S A* 2010, **107:**4418–4423.
22. Turner TL, Hahn MW, Nuzhdin SV: **Genomic islands of speciation in *Anopheles gambiae*.** *PLoS Biol* 2005, **3:**1572–1578.
23. Neafsey DE: **SNP genotyping defines complex gene-flow boundaries among African malaria vector mosquitoes.** *Science* 2010, **330:**1477.
24. Turner TL, Hahn MW: **Locus- and population-specific selection and differentiation between incipient species of *Anopheles gambiae*.** *Mol Biol Evol* 2007, **24:**2132–2138.
25. Caputo B, Santolamazza F, Vicente JL, Nwakanma DC, Jawara M, Palsson K, Jaenson T, White BJ, Mancini E, Petrarca V, Conway DJ, Besansky NJ, Pinto J, della Torre A: **The "Far-West" of *Anopheles gambiae* molecular forms.** *PLoS One* 2011, **6:**e16415.
26. Ndiath MO, Brengues C, Konate L, Sokhna C, Boudin C, Trape JF, Fontenille D: **Dynamics of transmission of *Plasmodium falciparum* by *Anopheles arabiensis* and the molecular forms M and S of *Anopheles gambiae* in Dielmo, Senegal.** *Malaria J* 2008, **7:**136.
27. Gillies MT, Coetzee M: **A supplement to the *Anophelinae* of Africa South of Sahara.** *Publ S Afr Inst Med Res* 1987, **54:**343.
28. Gillies MT, De Meillon B: **The *Anophelinae* of Africa South of the Sahara (Ethiopian zoogeographical region).** In *The South African Institute for Medical Research, Johannesburg.* 2nd edition. ; 1968:1–343.
29. Morlais I, Poncon N, Simard F, Cohuet A, Fontenille D: **Intraspecific nucleotide variation in *Anopheles gambiae*: new insights into the biology of malaria vectors.** *Am J Trop Med Hyg* 2004, **71:**795–802.
30. Fanello C, Santolamazza F, della Torre A: **Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP.** *Med Vet Entomol* 2002, **16:**461–464.
31. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers.** *Methods Mol Biol (Clifton, NJ)* 2000, **132:**365–386.
32. Simard F, Licht M, Besansky NJ, Lehmann T: **Polymorphism at the defensin gene in the *Anopheles gambiae* complex: testing different selection hypotheses.** *Infect Genet Evol* 2007, **7:**285–292.
33. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S: **MEGA6: molecular evolutionary genetics analysis version 6.0.** *Mol Biol Evol* 2013, **30:**2725–2729.
34. Kyte J, Doolittle RF: **A simple method for displaying the hydropathic character of a protein.** *J Mol Biol* 1982, **157:**105–132.
35. Von Heijne G: **Membrane protein structure prediction. Hydrophobicity analysis and the 'positive inside' rule.** *J Mol Biol* 1992, **225:**487–494.
36. Kengne P, Awono-Ambene P, Nkondjio CA, Simard F, Fontenille D: **Molecular identification of the *Anopheles nili* group of African malaria vectors.** *Med Vet Entomol* 2003, **17:**67–74.
37. Tsitoura P, Andronopoulou E, Tsikou D, Agalou A, Papakonstantinou MP, Kotzia GA, Labropoulou V, Swevers L, Georgoussi Z, Iatrou K: **Expression and membrane topology of *Anopheles gambiae* odorant receptors in lepidopteran insect cells.** *PLoS One* 2010, **5:**e15428.
38. Fontenille D, Lochouarn L, Diagne N, Sokhna C, Lemasson JJ, Diatta M, Konate L, Faye F, Rogier C, Trape JF: **High annual and seasonal variations in malaria transmission by *Anophelines* and vector species composition in dielmo, a holoendemic area in senegal.** *Am J Trop Med Hyg* 1997, **56:**247–253.

doi:10.1186/1756-0500-7-321
Cite this article as: Arnal *et al.*: Genetic polymorphism at an odorant receptor gene (Or39) among mosquitoes of the *Anopheles gambiae* complex in Senegal (West Africa). *BMC Research Notes* 2014 **7**:321.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit