

First report of an exophilic *Anopheles arabiensis* population in Bissau city, Guinea-Bissau: recent introduction or sampling bias?

Vasco Gordicho^{1,2}
Email: vascogordicho@hotmail.com

José L Vicente^{1,2}
Email: JoseLV@ihmt.unl.pt

Carla A Sousa¹
Email: CASousa@ihmt.unl.pt

Beniamino Caputo³
Email: beniamino.caputo@uniroma1.it

Marco Pombi³
Email: Marco.Pombi@uniroma1.it

João Dinis⁴
Email: jdinis66@gmail.com

Gonçalo Seixas^{1,2}
Email: gseixas@ihmt.unl.pt

Katinka Palsson⁵
Email: katinkap@kth.se

David Weetman⁶
Email: D.Weetman@liverpool.ac.uk

Amabélia Rodrigues⁴
Email: a.rodrigues@bandim.org

Alessandra della Torre³
Email: alessandra.dellatorre@uniroma1.it

João Pinto^{1,2,*}
Email: jpinto@ihmt.unl.pt

¹ UEI Parasitologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira 100, 1349-008 Lisbon, Portugal

² Centro de Malária e outras Doenças Tropicais Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira 100, 1349-008 Lisbon, Portugal

³ Istituto Pasteur-Fondazione Cenci-Bolognetti, Dipartimento di Sanità Pubblica e Malattie Infettive, Università di Roma “La Sapienza”, Piazzale Aldo Moro 5, 00185 Rome, Italy

⁴ Instituto Nacional de Saúde Pública, Avenida Combatentes da Liberdade da Pátria, Apartado 861, 1004 Bissau Codex, Guinea Bissau

⁵ Royal Institute of Technology, SE-100 44, Stockholm, Sweden

⁶ Department of Vector Biology, Liverpool School of Tropical Medicine, Liverpool L3 5QA, UK

* Corresponding author. UEI Parasitologia Médica, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Rua da Junqueira 100, 1349-008 Lisbon, Portugal

Abstract

Background

The malaria vector *Anopheles arabiensis* exhibits greater behavioural and ecological plasticity than the other major vectors of the *Anopheles gambiae* complex, which presents challenges for major control methods. This study reports for the first time the presence of *An. arabiensis* in Antula, a suburb of Bissau city, the capital of Guinea Bissau, where high levels of hybridization between *Anopheles coluzzii* and *An. gambiae* have been reported. Given that previous surveys in the area, based on indoor collections, did not sample *An. arabiensis*, the possibility of a recently introduced exophilic population was investigated.

Methods

Larval and adult mosquito collections were carried out in Antula at the end of the rainy season of 2010. *Anopheles gambiae* species composition, determined by rDNA-IGS and SINE200X6.1 markers, was compared with four previously collected samples dating back to 1993. Analysis of ten microsatellites was used to estimate levels of genetic diversity, relatedness and to investigate demographic stability.

Results

Anopheles arabiensis comprised 54.0% of larvae and 25.6% of adults collected in 2010, but was absent in all previous collections, a highly unlikely observation by chance if the population was stable. This species had the lowest levels of genetic diversity, highest relatedness and, along with *An. gambiae*, exhibited evidence of a recent population expansion.

Conclusions

Results point to the presence of a previously undetected outdoor population of *An. arabiensis* in Antula, which appears to have expanded recently, highlighting the importance of complementing indoor-based mosquito collections with sampling methods targeting outdoor

adults and immature stages for a more complete assessment of mosquito biodiversity. A change in temporal dynamics in the species complex composition was also detected. Coupled with previous evidence of asymmetric introgression from *An. coluzzii* to *An. gambiae*, this suggests that the study area may be subject to ecological changes with a potential impact on both the genetics of these species and on malaria transmission.

Background

For more than a century some of the sibling species of the *Anopheles gambiae* complex have been recognized as the most important Afrotropical malaria vectors, responsible for hundreds of thousands of deaths each year [1]. They are *Anopheles gambiae* and *Anopheles arabiensis*, which share a very widespread sympatric range. Of these two species, *An. gambiae* is considered a more efficient malaria vector due to its higher anthropophily.

A process of ecological speciation occurred within *An. gambiae* leading to the differentiation of two taxonomic units, preliminarily named M and S molecular forms [2] and lately elevated to species status, with the M-form being named as *Anopheles coluzzii* and the S-form keeping the original designation *An. gambiae* [3]. The two species co-exist in sympatry in West and Central Africa and hybrids between them are rare throughout most of their distribution range [4], although assortative mating may periodically or locally break down [5-9]. *Anopheles coluzzii* and *An. gambiae* differ in a few bio-ecological features (see [10] for a review). One of the most notable differences is the greater propensity of *An. coluzzii* to exploit larval habitats of more permanent nature (e.g., rice fields) due to superior predator avoidance when compared to *An. gambiae*, which predominates in temporary larval habitats [11-13]. Both types of larval habitats may also be colonized by *An. arabiensis* although recent evidence suggests that this species shares similar ecological requirements with *An. gambiae* in exploiting temporary larval habitats [13].

Guinea Bissau has recently become a focus of interest for studies aimed at better understanding the process of ecological speciation which led to the differentiation of *An. coluzzii* and *An. gambiae*, where exceptional hybrid rates up to *ca.* 25% have led to the hypothesis that the country may contain the core of a secondary contact region between the two species [5,7,8,14,15]. Deeper studies in this geographical region are expected to help clarify the reproductive isolating mechanisms between the two species, as well as to identify possible ecological determinants of their breakdown.

Previous reports on the *An. gambiae* complex species distribution in Guinea Bissau [16-19] show that *An. gambiae* was widespread throughout the country, while *An. arabiensis* was only present in the northern inland region, where drier open savannah and shrubland landscapes prevail. In addition, *Anopheles melas* was reported in the coastal region, characterized by flooded areas of mangrove and forest, as expected, based on its specific adaptation to brackish water larval habitats. These studies were mostly based on indoor-resting mosquito collections by pyrethrum spray catches or hand aspirations and, thus, only provide a picture of the endophagic and endophilic fraction of the local anopheline species composition in the region. In this context, larval collections may deliver a more unbiased sampling, irrespective of the feeding or resting patterns [20]. Moreover, the older studies predate the description of *An. gambiae* molecular forms. Nevertheless, the sympatric presence of *An. coluzzii*, *An. gambiae* and *An. melas* in the coastal area, where the capital city

Bissau is located, has been shown in mosquito samples collected in 1995 and 1996 [21] that were subsequently identified to molecular form by Oliveira *et al.* [5].

In this study, *An. gambiae* complex species composition was assessed in Antula, a suburb of the capital city Bissau, in indoor-collected adult samples as well as in larval samples. The primary objective was to assess larval spatial segregation possibly associated to niche partitioning between *An. coluzzii* and *An. gambiae* in this secondary contact zone. In the course of the analysis *An. arabiensis* was identified for the first time in the area, which predominated in larval samples. Microsatellite data was used to determine whether the presence of this vector could be a recent introduction or a well-established exophilic population that had not been sampled earlier in the area, owing to dependence on indoor adult collections.

Methods

Mosquito collections

The study took place in Antula (11°50'N, 15°30'W), a semirural suburb surrounded by flooded plains, mangrove swamps and subsistence agriculture plots, located about 5 km north of the centre of Bissau, Guinea Bissau's capital city [22]. The suburb is bordered eastwards by a large rice field. The majority of houses are clay-walled and thatch-roofed dwellings. Domestic animals (pigs, goats, chicken, and cattle) are frequent and are sometimes kept inside houses [5,22]. The use of bed nets for protection against malaria transmission has been implemented in the study site at least since 1995 [21] and insecticide-treated nets have been introduced in the area of Bissau since 2005 [23,24]. The climate of this region is tropical humid, with a rainy season from May to October and a dry season from November to April. Mosquito sampling was carried out at the end of the rainy season, between 8 and 28 October, 2010.

Mosquito larval collections were performed using dippers and pipettes from permanent and temporary larval habitats (Figure 1). The temporary larval habitats consisted of two main rain-water puddles bordered by smaller pools of various origins, such as tyre tracks or footprints, located on a dirt road. The permanent larval habitat was the rice field located to the east of the suburb. Larval samples were identified to the subfamily and anopheline mosquitoes were kept individually in 0.5 ml tubes filled with 80% ethanol. Adult mosquitoes were collected indoors by two methods: i) CDC light traps [25]; and, ii) resting collections, either with mechanical aspirators or by pyrethrum spray catches. Adults were identified to species or species complex with the help of digital keys [26] and kept individually in silica gel filled 0.5 ml tubes.

Figure 1 Map of Antula, Bissau (Guinea Bissau) showing the collection sites for larvae and adults (adapted from Google Maps).

In addition to 2010 collections, samples of indoor-resting *An. gambiae* s.l. females collected in the same study site, *i.e.* Antula, in September/October 1993, October 1995, November 1996, and August/September 2007 were also analysed for a temporal comparison of the *An. gambiae* complex species composition in the area. Further details on these samples are described elsewhere [5,21,27].

Molecular analyses

DNA extraction of larval samples was performed with the DNeasy® Blood & Tissue Kit (QIAGEN, Hilden, Germany) whereas the protocol described in Collins *et al.* [28] was used for adult mosquitoes.

Anopheles gambiae s.l. samples were identified to species by two complementary methods: i) the PCR-RFLP protocol of Fanello *et al.* [29] which targets species-specific polymorphisms at the intergenic spacer region of the ribosomal DNA (hereafter termed IGS); and, ii) a PCR assay targeting the insertion of the short interspersed element SINE200X6.1 (hereafter termed SINE) found to be fixed in *An. coluzzii* but absent in *An. gambiae s.s.* [30]. Specimens were identified as either *An. coluzzii* or *An. gambiae* if they had coincident species-specific patterns for both markers. Specimens exhibiting either a consistent *An. coluzzii/An. gambiae* pattern for both IGS and SINE or a discordant result between markers were considered as individuals of admixed ancestry [7,31].

A subsample of larvae was sequenced for a 658 bp fragment of the cytochrome oxidase I (COI) mitochondrial gene used in mosquito DNA barcoding. Amplification by PCR was performed with the universal primers HC02198 (TAAACTTCAGGGTGACCAAAAAATCA) and LCO1490 (GGTCAACAAATCATAAAGATATTGG) [32] under the conditions described by Herbert *et al.* [33]. Amplified products were cleaned with Qiaquick® PCR purification kit (Qiagen, Hilden, Germany) and sequenced in a DNA sequencing facility (STAB VIDA, Oeiras, Portugal). Forward and reverse sequences were aligned and corrected by hand using BIOEDIT v. 7.0.5.2 [34]. Species identification was performed using the barcoding identification engine BOLD v. 2 [35].

Ten autosomal microsatellite loci mapped on chromosome-3 of *An. gambiae* were genotyped [36,37] (see Additional file 1). Microsatellites located on chromosomes-X and -2 were not used to prevent bias due to gender (as larvae were not sexed) and selective pressures associated with paracentric inversions known to be frequent on chromosome-2 [38]. However, two of these microsatellites (AG3H119 and AG3H555) are located within inversion 3Ra (between divisions 31A and 34D), which is a polymorphic inversion in *An. arabiensis* [39]. Amplification was performed by PCR using forward primers labelled with 5' fluorescent dyes (FAM, NED or HEX, Applied Biosystems, Foster City CA, USA) as described previously [40]. Amplified fragments were separated by capillary electrophoresis in an automatic sequencer ABI 3730 (Applied Biosystems, Foster City CA, USA) at Yale University's DNA Analysis Facility (New Haven, CT, USA). Fragment sizes were scored using the software GeneMarker v1.4 (SoftGenetics, State College, PA, USA).

Data analysis

Ninety-five per cent confidence intervals of proportions based on the sample size were calculated with continuity correction as described by Newcombe [41] and implemented in VassarStats [42]. Comparisons between groups were made by Pearson's Chi-square tests on contingency tables or Fisher's exact tests in the case of low sample sizes.

Microsatellite genetic diversity per locus and sample was characterized by estimates of allele richness (A_R) [43], Nei's unbiased expected heterozygosity H_e [44] and inbreeding coefficient F_{IS} [45]. Calculations were performed in FSTAT v. 2.9.3.2 [46]. Significance of mean F_{IS}

values was assessed by randomization tests also available in FSTAT (1,400 replicates). Comparisons among samples of mean over-loci estimates of A_R , H_e , and F_{IS} were performed by calculating bootstrapped 95% confidence intervals (1,000 replicates) using the Bootstrap Plot for Central Tendency v. 1.0.13 [47]. Genotypic frequencies were tested against Hardy-Weinberg Equilibrium (HWE) by exact probability tests performed in GENEPOP v. 4.1 [48]. The same software was used to perform exact tests of genotypic linkage disequilibrium between pairs of loci in each sample. Presence of null alleles at each locus and sample was assessed using the software MICRO-CHECKER v.2.2.3 [49].

Microsatellite loci isolated from one species (focal species) may be less variable in closely related species, due to ascertainment bias during the selection of microsatellite loci, in which loci with the longest tracts of pure repeats are usually favoured in order to ensure polymorphism [50,51]. Since ascertainment bias is expected to increase with microsatellite length, interspecies diversity differences should be higher for loci with longer repeat tracts in *An. gambiae* and a stronger correlation between genetic diversity and repeat tract length is expected for the focal species. To test this hypothesis, estimates of allele richness were plotted against the length of the repeat tract (in repeat units) of the microsatellites described in the original clones [36,37]. Analyses were conducted with and without loci AG3H93 and 45C1 because microsatellites with interrupted repeat motifs tend to be less variable than loci with pure repeat motifs [50]. Spearman rank correlation coefficients were estimated in SPSS v.20.0 [52] for *An. arabiensis* (larvae and adults pooled) and a pooled sample of *An. coluzzii* and *An. gambiae* (adults and larvae) representing the focal species (the isolation of the microsatellites predates the description of molecular form divergence). Fisher's r-to-z tests, available in VassarStats, were used to assess the significance of the difference between two correlation coefficients.

In order to assess the degree of relatedness among individuals within samples, estimates of Queller and Goodnight [53] and Lynch and Ritland [54] relatedness coefficients were calculated using GENALEX v.6.5 [55]. Significance of the mean estimates for each sample was assessed by permutation tests (1,000 replicates) and bootstrapped 95% confidence intervals (1,000 replicates). In addition, the maximum-likelihood method implemented in ML-RELATE [56] was used to determine proportions of related individuals within samples. For each pair of individuals, log-likelihood estimates are calculated for four pedigree classes: unrelated, parent-offspring, full-siblings, and half-siblings. In loci displaying presence of null alleles, relatedness calculations were adjusted by including maximum likelihood estimates of the frequency of the putative null allele [57]. Individual pairs classified as relatives (i.e., PO, FS or HS) were summed in order to calculate the proportion of related individual pairs in each sample.

Single-sample estimates of current effective population size were calculated by the bias-corrected linkage disequilibrium method described in Waples and Do [58], as implemented in NeEstimator v.2 [59]. Because rare alleles may bias linkage disequilibrium N_e estimates, alleles with frequency below 0.05 at each locus were removed from the analysis.

Heterozygosity tests to detect deviations from mutation-drift equilibrium (MDE) were performed using BOTTLENECK v1.2.02 [60]. In these tests, two estimates of heterozygosity are compared: one based on allele frequencies assuming HWE (H_e) and the other based on the number of alleles and sample size assuming MDE (H_{eq}). At MDE, both estimates should be similar at the majority of the loci analysed (i.e., $H_e = H_{eq}$). In case of a population bottleneck, allelic diversity will decrease faster than heterozygosity (i.e. $H_e > H_{eq}$), while the

opposite (i.e., $H_e < H_{eq}$) is an indicator of a population expansion. Estimates of expected heterozygosity under MDE were performed using the stepwise mutation model (SMM) and a two-phased model (TPM) with 10 to 20% indels larger than the repeat unit.

Whenever multiple tests were performed, the sequential Bonferroni procedure was applied to adjust the nominal significance level ($\alpha = 0.05$) [61].

Results

A total of 305 anopheline larvae (95 from the rice field and 210 from the temporary puddles) and 339 adult females (294 from CDC light traps and 45 from indoor-resting collections) were collected in the 2010 mosquito survey. While all 210 larvae collected in the temporary puddles were successfully amplified for both IGS and SINE, no amplified product was obtained for both markers in 79 (83.2%) out of the 95 specimens collected in the rice field. A subsample of 30 negative larvae was sequenced for the mtDNA COI fragment to perform species identification in BOLD (see Additional file 2). Of these, 28 individuals were identified as *Anopheles coustani* (99.3-100.0% similarity), one as *Uranotaenia balfouri* (99.7% similarity) and one specimen was assigned to an undetermined *Anopheles* species (*Anopheles* MBI-14, 99.4% similarity). An NCBI BLAST of the mtDNA COI sequence for this specimen gave a 91.0% similarity with *Anopheles funestus*. In the adult sample, one (2.2%) indoor-resting and four (1.4%) CDC light trap collected specimens had a readable genotype only for either the IGS or the SINE markers. These specimens were removed from subsequent analyses. Molecular identification by both IGS and SINE was thus achieved for a total of 560 specimens (226 larvae and 334 adults) from the 2010 collection. Additionally, a total of 553 indoor-resting females from collections undertaken between 1993 and 2007 were identified to species by both markers.

Anopheles gambiae and *An. arabiensis* were the predominant species sampled in the 2010 collection (Table 1). *Anopheles arabiensis* comprised 54.0% of the overall larval sample but only 3.8 and 22.7% of adults collected by CDC light traps and indoor-resting capture, respectively. *Anopheles coluzzii* larvae were collected only in temporary puddles, reaching a frequency of 10.5% which was *ca.* two-fold greater than those recorded in the adult samples. The frequency of this species was below 5.0% in the adult samples. *Anopheles melas* was identified only in adult samples, with an overall frequency of 2.7%.

Table 1 Species distribution (in percentage) according to mosquito life-stage and collection site or method

		<i>N</i>	<i>Anopheles arabiensis</i>	<i>Anopheles coluzzii</i>	<i>Anopheles gambiae</i>	Admixed	<i>Anopheles melas</i>
Larvae	Puddles	210	52.9	10.5	24.8	11.9	0.0
			[45.9-59.7]	[6.8-15.6]	[19.2-31.3]	[8.0-17.3]	[0.0-2.2]
	Rice field	16	68.8	0.0	6.3	25.0	0.0
			[41.5-87.9]	[0.0-24.1]	[0.3-32.3]	[8.3-52.6]	[0.0-24.1]
	Subtotal	226	54.0	9.7	23.5	12.8	0.0
			[47.3-60.6]	[6.3-14.6]	[18.2-29.6]	[8.9-18.1]	[0.0-2.1]
Adults	LT	290	3.8	4.8	43.8	45.2	2.4
			[2.0-6.9]	[2.8-8.2]	[38.0-49.7]	[39.4-51.1]	[1.1-5.1]
	IR	44	22.7	4.5	43.2	25.0	4.5
			[12.0-38.2]	[0.8-16.7]	[28.7-58.9]	[13.7-40.7]	[0.8-16.7]
	Subtotal	334	6.3	4.8	43.7	42.5	2.7
			[4.0-9.6]	[2.9-7.8]	[38.4-49.2]	[37.2-48.0]	[1.3-5.2]
Total		560	25.5	6.8	35.5	30.5	1.6
			[22.0-29.4]	[4.9-9.3]	[31.6-39.7]	[26.8-34.6]	[0.8-3.1]

N: sample size; LT: CDC light trap collection; IR: indoor resting collection. In square brackets: 95% confidence intervals.

The relative proportions of *An. coluzzii*, *An. gambiae* and admixed individuals (i.e., excluding *An. arabiensis* and *An. melas*) were significantly different when larval and adult samples were compared ($\chi^2 = 26.70$; d.f. 2; $P < 0.001$; Additional file 3). *Anopheles gambiae* prevailed over *An. coluzzii* in both larval (51.0%) and adult (48.0%) samples. The relative proportion of *An. coluzzii* decreased from 22.1% in larvae to 5.3% in adults ($\chi^2 = 23.17$; d.f. 1; $P < 0.001$), while relative proportion of admixed individuals nearly doubled from larvae (27.9%) to adults (46.7%) ($\chi^2 = 11.28$; d.f. 1; $P < 0.001$).

The detection of *An. arabiensis* in the study area was unexpected, as this species was not identified in previous indoor-resting samples collected since 1993 (Figure 2). Given the sample sizes of each year, the upper 95% confidence levels (UCL) for *An. arabiensis* to be present but not sampled in these previous years (1993: UCL = 2.2%, $N = 213$; 1995: UCL = 0.9%, $N = 549$; 1996: UCL = 5.9%, $N = 78$; 2007: UCL = 2.9%, $N = 162$) do not overlap with the confidence interval obtained for the proportion of *An. arabiensis* recorded in 2010 (95%CI: 12.0-38.2%, $N = 44$). Figure 2 also shows an apparent increase of the relative frequency of *An. gambiae* with a concomitant decrease of *An. coluzzii*. With the exception of the sample of 1996 (43.6%), the proportion of *An. coluzzii* in indoor resting samples went from 26.3% in 1993 to 4.5% in 2010.

Figure 2 Temporal variation of species composition of the *Anopheles gambiae* complex in indoor-resting collections Y-axis: proportion of each species; X-axis: years of collection and sample sizes in brackets. * The 1995 sample includes data from Palsson *et al.* [21] from which a subsample of 100 individuals was identified by both IGS and SINE.

Microsatellite-based analyses

A total of 265 specimens collected in 2010 were genotyped for ten chromosome-3 microsatellite loci. Since the objective of this analysis was to perform interspecific comparisons, admixed individuals (as well as the few *An. coluzzii* and *An. melas* adults available) were excluded from the analysis.

Significant departures from HWE expectations were detected in five out of 50 tests performed, in the samples of *An. arabiensis* larvae (AGH758), *An. gambiae* adults (AG3H93) and *An. gambiae* larvae (AG3H119, AG3H242 and AG3H249) (Additional file 1). These departures were associated with positive F_{IS} values suggesting heterozygote deficits. Analysis performed by MICRO-CHECKER detected the presence of null alleles in three out of the five loci displaying heterozygote deficits. Loci AG3H119 and AG3H555, located within polymorphic inversion 3Ra in *An. arabiensis*, did not show any departures from HWE expectations in this species. These loci were not outliers for estimates of A_R , H_e or F_{IS} , suggesting that microsatellite polymorphism does not seem to have been affected by the possible presence of 3Ra inversion polymorphism in the *An. arabiensis* population sampled. There was no particular association between pairs of loci in any of the samples. Of the seven significant pairwise tests of linkage disequilibrium (out of 225 performed), four were detected in the sample of *An. arabiensis* larvae, one in *An. coluzzii* larvae and two in *An. gambiae* larvae. The low number of loci with heterozygote deficits and of linkage disequilibrium tests is consistent with each sample representing a single panmictic gene pool.

Mean estimates of genetic diversity were similar between larval and adult samples within the same species (Table 2). *Anopheles arabiensis* was less genetically diverse than *An. coluzzii* or

An. gambiae. Mean allele richness for this species was around six alleles per locus whereas it varied between nine and ten in *An. coluzzii* and *An. gambiae*, respectively. Similarly, mean expected heterozygosity in *An. arabiensis* was below 0.600 while values above 0.800 were recorded for *An. coluzzii* and *An. gambiae*. These differences were significant judging from the non-overlapping bootstrapped 95% confidence intervals.

Table 2 Mean estimates of genetic diversity and relatedness

		A_R	H_e	F_{IS}	R_{LR}	R_{QG}	R_{ML}
<i>Anopheles arabiensis</i>	Adults	5.7	0.583	-0.054	0.054	0.304	17.9
	($N = 20$)	[4.5-6.9]	[0.431-0.708]	[-0.132-0.017]	[0.042-0.063]	[0.279-0.327]	[12.9-24.3]
	Larvae	5.8	0.598	0.024	0.049	0.268	18.3
<i>Anopheles coluzzii</i>	($N = 122$)	[4.7-6.8]	[0.455-0.706]	[-0.013-0.064]	[0.047-0.050]	[0.264-0.272]	[17.5-19.3]
	Larvae	8.8	0.812	0.017	0.023	0.027	6.5
<i>Anopheles gambiae</i>	($N = 22$)	[7.5-10.5]	[0.745-0.870]	[-0.052-0.086]	[0.017-0.030]	[0.009-0.047]	[3.8-10.7]
	Adults	10.0	0.819	0.082	0.019	0.022	9.8
	($N = 48$)	[8.1-11.9]	[0.767-0.864]	[0.055-0.113]	[0.016-0.021]	[0.013-0.031]	[8.1-11.7]
	Larvae	9.7	0.812	0.079	0.015	0.020	10.0
	($N = 53$)	[7.9-11.7]	[0.756-0.863]	[0.022-0.130]	[0.013-0.019]	[0.012-0.028]	[8.5-11.8]

N : sample size; A_R : allele richness; H_e : expected heterozygosity; F_{IS} : inbreeding coefficient; R_{LR} : Lynch and Ritland [54] relatedness coefficient; R_{QG} : Queller and Goodnight [53] relatedness coefficient; R_{ML} : proportion (in percentage) of related pairs of individuals as determined by ML-RELATE. Significant estimates after correction of multiple tests for F_{IS} , R_{LR} and R_{QG} are in bold. 95% confidence intervals are in square brackets.

Positive Spearman's rho correlation coefficients between allele richness and repeat length of the original microsatellite clone were obtained for both *An. arabiensis* (rho = 0.648, $P = 0.043$) and *An. gambiae/An. coluzzii* (rho = 0.349, $P = 0.323$) (see Additional file 4). These coefficients were not statistically different (Fisher's r-to-z test: $z = 0.760$, $P = 0.447$), and when only microsatellites with pure repeat motifs were analysed, correlation coefficients were near-identical (*An. arabiensis*: rho = 0.686, $P = 0.060$; *An. gambiae/An. coluzzii*: rho = 0.713, $P = 0.047$; Fisher's r-to-z test: $z = 0.080$, $P = 0.936$).

There was no evidence for increased inbreeding in larval samples compared to the adult samples in both *An. arabiensis* and *An. gambiae* (Table 2). The latter species presented the highest and only significant mean F_{IS} estimates but these values were also practically identical between larvae and adults. Similarity between larval and adult samples was also evident in the values obtained for the estimators of relatedness in both species (Table 2; see Additional file 5). However, *An. arabiensis* exhibited the highest estimates for both relatedness coefficients and for the proportion of related individuals, when compared to the other two species. These differences were significant, judging from the non-overlapping confidence intervals, suggesting a higher degree of relatedness among *An. arabiensis* individuals in both larvae and adults (Table 2).

Mean estimates of effective population size varied between 32.8 in *An. arabiensis* adults and 192.4 in *An. gambiae* adults, with the latter estimate being the only with an unbounded 95% confidence interval (Table 3). However, all estimates had overlapping 95% confidence intervals indicative of no significant differences in N_e among samples. A significant departure from MDE was detected in both larval and adult samples of *An. arabiensis* (Table 3). This departure corresponded to a significant number of loci with an apparent deficit of heterozygotes when compared to expectations under MDE (i.e., $H_e < H_{eq}$), an indicator of recent population expansion, and was consistent in all mutation models. A signal of population expansion was also detected in *An. gambiae* but only under the SMM and TPM 10% mutation models, in the case of the adult sample, and under the SMM in the case of the larval sample. There was no evidence of departure from MDE in the *An. coluzzii* sample.

Table 3 Estimates of effective population size and heterozygosity tests

		N_e	Heterozygosity tests		
			SMM	TPM (10%)	TPM (20%)
<i>An. arabiensis</i>	Adults	32.8 [15.1-253.0]	9 <0.001	9 0.002	9 0.002
	Larvae	88.3 [54.7-166.4]	10 <0.001	10 <0.001	10 <0.001
<i>An. coluzzii</i>	Larvae	38.5 [23.2-86.7]	5 0.278	5 0.615	2 0.884
<i>An. gambiae</i>	Adults	192.4 [84.3-∞]	9 <0.001	8 0.003	8 0.007
	Larvae	114.8 [61.7-415.3]	9 0.001	7 0.012	6 0.097

N_e : linkage disequilibrium-based estimate of effective population size and Jackknife 95% confidence interval in square brackets. Heterozygosity tests: upper values are the number of loci (out of 10) in which $H_e < H_{eq}$; lower values are the P -value for the corresponding one-tailed Wilcoxon test. SMM: stepwise mutation model; TPM (10%): two-phased model with 10% of indels greater than one repeat; TPM (20%): two-phased model with 20% of indels greater than one repeat.

Discussion

This study reports for the first time the occurrence of the major malaria vector *An. arabiensis* in Bissau, the capital city of Guinea Bissau. This was an unexpected finding since adult mosquito surveys carried out in the same area (Antula) and with the same collection method (indoor resting) did not sample this species [16,18,19,21,22] (Figure 2). In fact, *An. arabiensis* was reported over 30 years ago in Guinea Bissau, but its distribution appeared to be limited to the northeast inland region of the country, characterized by a drier savannah ecosystem [16].

The new occurrence of *An. arabiensis* in the humid coastal region of Guinea Bissau could have resulted from a very recent (after 2007) introduction by sporadic migration of a single or few females from north-eastern inland populations. If this was the case, then the expectation would be for an *An. arabiensis* population with low genetic diversity, small effective population size and with a signal of population contraction as a consequence of a recent founder event. Allele richness and expected heterozygosity estimated for both larval and adult *An. arabiensis* samples were indeed lower than those of *An. coluzzii* and *An. gambiae*. *Anopheles arabiensis* also presented a higher degree of relatedness among individuals. However, estimates of current N_e obtained for *An. arabiensis* were comparable to those of *An. gambiae* and *An. coluzzii*. While these estimates may be affected by the relatively low number of microsatellites analysed and small sample sizes, *An. arabiensis* does not seem to have a dramatically reduced N_e when compared to its sibling species, and no signal of population contraction was detected by heterozygosity tests. Instead, an apparent heterozygote deficit, relative to equilibrium expectations, was detected consistently in both larval and adult *An. arabiensis* samples, suggesting population expansion. This result agrees with the apparent increase of the relative frequency of this species in 2010, which reached 22.7% (12.0-38.2%) in indoor-resting collections, a value far greater than the maximum upper level confidence limit estimated for previous collections (5.9%, in 1996). Therefore, although the possibility of a new colonization in Bissau cannot be excluded, a more likely explanation is that of an expanding *An. arabiensis* population that has been resident for some time and that other causes may underlie the reduced levels of diversity.

The unavailability of microsatellite loci isolated specifically from *An. arabiensis* precluded performing reciprocal tests [50] so that ascertainment bias cannot be fully rejected. However, correlation coefficients between allele diversity and microsatellite repeat tract length were similar for both *An. arabiensis* and *An. gambiae* + *An. coluzzii*, indicating that allele diversity was consistently lower in *An. arabiensis* irrespective of the originally cloned repeat tract length (which varied between six and 21 repeats). This similarity in the diversity vs tract length relationship across species argues against ascertainment bias as an explanation for reduced genetic diversity of the microsatellite loci in *An. arabiensis*.

Another hypothesis is that the low genetic diversity found in *An. arabiensis* reflects a small but now-expanding exophagic and exophilic population living at the edge of the species distribution. This population could have remained undetected in collections carried out before 2010 due not only to its low abundance, but also to inadequate sampling methodology. The difference in the relative frequency of *An. arabiensis* between larval and adult samples supports this hypothesis. In 2010, *An. arabiensis* was the most frequent species in larval collections (exceeding 50%) but this was not apparent in the indoor adult samples, particularly in CDC light trap collections, where its frequency was lower than 4%. This suggests the presence of a markedly exophagic and exophilic *An. arabiensis* in Antula,

difficult to catch using CDC light trap and resting collections performed indoors, which sample preferentially endophagic and endophilic fractions of a mosquito population. Bio-ecological studies point to a greater behavioural plasticity of *An. arabiensis* when compared to *An. coluzzii* and *An. gambiae* (see [1] for a review). Populations of this species frequently display higher degrees of exophagy and exophily, which is often associated with a higher propensity to feed on non-human hosts. Moreover, the introduction of insecticide-treated bed nets in the area [23,24] may also have contributed to the selection of outdoor behaviours in this species. Additional bio-ecological studies, involving outdoor collections and determination of host-preferences and gonotrophic state, will be required to confirm the extent of these behaviours in the *An. arabiensis* population of Antula. Such studies should also involve sampling in different collection sites in order to clarify if the reduced diversity is associated with the edge of the species distribution and the possible causes of the apparent population expansion detected in the 2010 sample.

There is also a possibility of the high proportion of *An. arabiensis* found in larvae having resulted from sampling a large number of siblings from a single or a few ovipositions, due to the relatively low number of larval habitats surveyed. However, this is not supported by estimates of genetic relatedness, which, although higher in *An. arabiensis*, were similar between adult and larval samples of this species as well as of *An. gambiae*, suggesting that larval and adult collections were equally representative of the population for both species. The proportion of related individuals determined in this study was lower than those reported for *An. gambiae* larvae from Kenya [62] and for adult samples from three African countries [63], but these studies have used different microsatellites and relatedness estimators. Furthermore, methods to estimate genetic relatedness require a large number of polymorphic loci (*ca.* 30-40) to fully discriminate individual pairs according to pedigree classes [56,64]. This was evident in the present analysis when parent-offspring relations at frequency between 0.1 and 2.5% (see Additional file 5) were identified in larval samples, which is a biologically unsound result. However, even when based on a few loci these approaches may still be useful when the primary goal is to compare the average relatedness within groups [64], as was the case of the present study.

The increasing relative frequencies of *An. gambiae* since 1993 and the high levels of genetic diversity with a signal of population expansion observed in 2010 suggest that this species is also expanding in the study area. This may be due to a higher fitness derived from integration of *An. coluzzii* genetic variants following asymmetric genetic introgression from *An. coluzzii* to *An. gambiae* in this secondary contact zone as previously shown [5,14] and here confirmed by the increased proportion of admixed individuals in the adult collections. Additional surveys targeting the dry season and involving genetic analyses are required to confirm whether the apparent dominance of *An. gambiae* and *An. arabiensis* over *An. coluzzii* in the study site is a stable situation or if it corresponds to a seasonal fluctuation. In fact, the frequency of *An. coluzzii* was higher than *An. gambiae* in the only adult sample collected at the onset of the dry season (*i.e.*, November 1996), contrasting with the other samples that were collected in the rainy season. This is possibly due to *An. coluzzii* higher capacity to explore more permanent larval habitats [11,13].

The initial goal of assessing spatial segregation at the larval stage between *An. coluzzii* and *An. gambiae* in a setting of high hybridization was hampered by the low numbers obtained for these species in the collections made in the permanent larval habitat. Of the 95 larvae collected in the rice field, only 16 belonged to the *An. gambiae* complex and 11 were identified as *An. arabiensis*. Morphological identification of mosquito species at the larval

stage sometimes implies mounting the biological material in slides for microscopic observation of diagnostic structures, which can be difficult under certain field conditions. In this context, DNA barcoding may be an effective alternative. This approach reliably identified a subsample of the larvae, revealing a large predominance of *An. coustani* in the rice field larval collection made. Members of this species complex are considered secondary malaria vectors due to a high degree of zoophily and exophagy. However, a few studies suggest a role in malaria transmission, which may justify further attention to this complex in future malaria vector surveys in area [65,66].

Conclusions

The finding of an *An. arabiensis* population in Antula, apparently displaying exophilic behaviour, has important implications for the epidemiology and control of malaria. Previous malaria control programmes, such as the Garki Project in the 1970s, have demonstrated that outdoor feeding and resting mosquito populations can undermine vector control [67,68] which is still based mainly on indoor measures. This is of particular relevance for malaria control efforts in Guinea Bissau given that free distribution of insecticide-treated nets has been the mainstay of vector control in the country since 2005 [23,24]. Finally, these results highlight the importance of complementing indoor mosquito sampling with alternative methods targeting outdoor adult mosquitoes and immature stages, for a more representative sampling of the mosquito biodiversity in a given region.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JLV, CAS, BC, MP, JD, KP, and JP carried out mosquito surveys. VG, JLV, BC, and DW performed species molecular identification and microsatellite genotyping. GS performed species identification by DNA barcoding. VG, DW and JP conducted data analysis. AR, AdT and JP conceived and coordinated the study. VG, DW, MP, AdT, and JP drafted the manuscript. All authors read and approved the final manuscript.

Acknowledgements

The authors wish to thank the people of Antula, Bissau for their hospitality and goodwill during mosquito collections. We acknowledge Prof Thomas G T Jaenson (Uppsala University, Sweden), Dr Francisco Dias and Mr Mario Joao Gomes (National Public Health Laboratory, Guinea Bissau), and Dr Rui Barreto dos Santos (Veterinary Services, Guinea Bissau) for providing some of the historical mosquito samples. This study received financial support from FP7/EC funded INFRAVEC project, FCT Portugal/FEDER (through Programme COMPETE) co-funds (PTDC/BIA-EVF/120407/2010), MIUR-FIRB “Futuro in Ricerca 2010” grant to BC (Grant N° RBF106NTE) and University of Rome SAPIENZA “AWARDS 2013” project.

References

1. Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, Patil AP, Temperley WH, Gething PW, Kabaria CW, Okara RM, Van Boeckel T, Godfray HC, Harbach RE, Hay SI: **The dominant *Anopheles* vectors of human malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and bionomic precis.** *Parasit Vectors* 2010, **3**:117.
2. 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.
3. Coetzee M, Hunt RH, Wilkerson R, Della Torre A, Coulibaly MB, Besansky NJ: ***Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex.** *Zootaxa* 2013, **3619**:246–274.
4. 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.
5. Oliveira E, Salgueiro P, Palsson K, Vicente JL, Arez AP, Jaenson TG, Caccone A, Pinto J: **High levels of hybridization between molecular forms of *Anopheles gambiae* from Guinea Bissau.** *J Med Entomol* 2008, **45**:1057–1063.
6. Caputo B, Nwakanma D, Jawara M, Adiamoh M, Dia I, Konate L, Petrarca V, Conway DJ, Della Torre A: ***Anopheles gambiae* complex along The Gambia river, with particular reference to the molecular forms of *An. gambiae* s.s.** *Malar J* 2008, **7**:182.
7. 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.
8. 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.
9. Lee Y, Marsden CD, Norris LC, Collier TC, Main BJ, Fofana 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 U S A* 2013, **110**:19854–19859.
10. Lehmann T, Diabate A: **The molecular forms of *Anopheles gambiae*: A phenotypic perspective.** *Infect Genet Evol* 2008, **8**:737–746.
11. Diabate A, Dabire RK, Heidenberger K, Crawford J, Lamp WO, Culler LE, Lehmann T: **Evidence for divergent selection between the molecular forms of *Anopheles gambiae*: role of predation.** *BMC Evol Biol* 2008, **8**:5.

12. 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.
13. Gimonneau G, Pombi M, Choisy M, Morand S, Dabire RK, Simard F: **Larval habitat segregation between the molecular forms of the mosquito *Anopheles gambiae* in a rice field area of Burkina Faso, West Africa.** *Med Vet Entomol* 2012, **26**:9–17.
14. Marsden CD, Lee Y, Nieman CC, Sanford MR, Dinis J, Martins C, Rodrigues A, Cornel AJ, Lanzaro GC: **Asymmetric introgression between the M and S forms of the malaria vector, *Anopheles gambiae*, maintains divergence despite extensive hybridization.** *Mol Ecol* 2011, **20**:4983–4994.
15. 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.
16. Petrarca V, Carrara GC, Di Deco MA, Petrangeli G: **The *Anopheles gambiae* complex in Guinea Bissau.** *Parassitologia* 1983, **25**:29–39.
17. Jaenson TGT, Gomes MJ, Dossantos RCB, Petrarca V, Fortini D, Evora J, Crato J: **Control of endophagic *Anopheles* mosquitoes and human malaria in Guinea-Bissau, West-Africa by permethrin-treated bed nets.** *Trans R Soc Trop Med Hyg* 1994, **88**:620–624.
18. Fonseca LF, di Deco MA, Carrara GC, Dabo I, Do Rosario V, Petrarca V: ***Anopheles gambiae* complex (Diptera: Culicidae) near Bissau City, Guinea Bissau, West Africa.** *J Med Entomol* 1996, **33**:939–945.
19. Dabire KR, Diabate A, Agostinho F, Alves F, Manga L, Faye O, Baldet T: **[Distribution of the members of *Anopheles gambiae* and pyrethroid knock-down resistance gene (*kdr*) in Guinea-Bissau, West Africa](in French).** *Bull Soc Pathol Exot* 2008, **101**:119–123.
20. Riehle MM, Guelbeogo WM, Gneme A, Eiglmeier K, Holm I, Bischoff E, Garnier T, Snyder GM, Li X, Markianos K, Sagnon N, Vernick KD: **A Cryptic Subgroup of *Anopheles gambiae* Is Highly Susceptible to Human Malaria Parasites.** *Science* 2011, **331**:596–598.
21. Palsson K, Pinto J, Do Rosario VE, Jaenson TGT: **The palpal ratio method compared with PCR to distinguish between *Anopheles gambiae* s.s. and *A. melas* from Guinea Bissau, West Africa.** *Acta Trop* 1998, **70**:101–107.
22. Palsson K, Jaenson TGT, Dias F, Laugen AT, Bjorkman A: **Endophilic *Anopheles* mosquitoes in Guinea Bissau, West Africa, in relation to human housing conditions.** *J Med Entomol* 2004, **41**:746–752.
23. WHO: *World Malaria Report 2013*. World Health Organization: Geneva; 2013.
24. Ursing J, Rombo L, Rodrigues A, Aaby P, Kofoed P-E: **Malaria transmission in Bissau, Guinea-Bissau between 1995 and 2012: malaria resurgence did not negatively affect mortality.** *PLoS One* 2014, **9**:e101167.

25. Sudia WD, Chamberlain RW: **Battery-operated light trap, an improved model.** *J Am Mosq Control Assoc* 1988, **4**:536–538.
26. Hervy J-P, Le Goff G, Geoffroy B, Hervé J-P LM, Brunhes J: *Logiciel d'identification et d'enseignement: les anophèles de la région afro-tropicale.* Paris: ORSTOM; 1998. CD-ROM.
27. Arez AP, Pinto J, Palsson K, Snounou G, Jaenson TGT, Do Rosario VE: **Transmission of mixed *Plasmodium* species and *Plasmodium falciparum* genotypes.** *Am J Trop Med Hyg* 2003, **68**:161–168.
28. Collins FH, Mehaffey PC, Rasmussen MO, Brandlingbennett AD, Odera JS, Finnerty V: **Comparison of DNA-probe and isoenzyme methods for differentiating *Anopheles gambiae* and *Anopheles arabiensis* (Diptera, Culicidae).** *J Med Entomol* 1988, **25**:116–120.
29. 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.
30. Santolamazza F, Mancini E, Simard F, Qi Y, Tu Z, della Torre A: **Insertion polymorphisms of SINE200 retrotransposons within speciation islands of *Anopheles gambiae* molecular forms.** *Malar J* 2008, **7**:163.
31. Santolamazza F, Caputo B, Calzetta M, Vicente JL, Mancini E, Petrarca V, Pinto J, Pinto J: **Comparative analyses reveal discrepancies among results of commonly used methods for *Anopheles gambiae* molecular form identification.** *Malar J* 2011, **10**:215.
32. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R: **DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates.** *Mol Mar Biol Biotechnol* 1994, **3**:294–299.
33. Hebert PDN, Cywinska A, Ball SL, DeWaard JR: **Biological identifications through DNA barcodes.** *Proc Biol Sci* 2003, **270**:313–321.
34. Hall T: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucleic Acids Symp Ser (Oxf)* 1999, **41**:95–98.
35. Ratnasingham S, Hebert PDN: **BOLD: The Barcode of Life Data System (www.barcodinglife.org).** *Mol Ecol Notes* 2007, **7**:355–364.
36. Zheng LB, Benedict MO, Cornel AJ, Collins FH, Kafatos FC: **An integrated genetic map of the African human malaria vector mosquito, *Anopheles gambiae*.** *Genetics* 1996, **143**:941–952.
37. Wang R, Kafatos F, Zheng L: **Microsatellite markers and genotyping procedures for *Anopheles gambiae*.** *Parasitol Today* 1999, **15**:33–37.
38. Lanzaro GC, Toure YT, Carnahan J, Zheng LB, Dolo G, Traore S, Petrarca V, Vernick KD, Taylor CE: **Complexities in the genetic structure of *Anopheles gambiae* populations**

in west Africa as revealed by microsatellite DNA analysis. *Proc Natl Acad Sci U S A* 1998, **95**:14260–14265.

39. Coluzzi M, Sabatini A, Della Torre A, Di Deco MA, Petrarca V: **A polytene chromosome analysis of the *Anopheles gambiae* species complex.** *Science* 2002, **298**:1415–1418.

40. Donnelly MJ, Cuamba N, Charlwood JD, Collins FH, Townson H: **Population structure in the malaria vector, *Anopheles arabiensis* Patton, in East Africa.** *Heredity* 1999, **83**:408–417.

41. Newcombe RG: **Two-sided confidence intervals for the single proportion: Comparison of seven methods.** *Stat Med* 1998, **17**:857–872.

42. **VassarStats: Website for Statistical Computation.** [<http://www.vassarstats.net/>].

43. ElMousadik A, Petit RJ: **High level of genetic differentiation for allelic richness among populations of the argan tree *Argania spinosa* (L) Skeels endemic to Morocco.** *Theor Appl Genet* 1996, **92**:832–839.

44. Nei M: *Molecular Evolutionary Genetics*. New York: Columbia University Press; 1987.

45. Weir BS, Cockerham CC: **Estimating F-statistics for the analysis of population-structure.** *Evolution* 1984, **38**:1358–1370.

46. Goudet J: **FSTAT (Version 1.2): A computer program to calculate F-statistics.** *J Hered* 1995, **86**:485–486.

47. Wessa P: **Bootstrap Plot for Central Tendency (v. 1.0.13) in Free Statistics Software (v. 1.1.23-r7).** [http://www.wessa.net/rwasp_bootstrapplot1.wasp/].

48. Raymond M, Rousset F: **GENEPOP (version 1.2) - population-genetics software for exact tests and ecumenicism.** *J Hered* 1995, **86**:248–249.

49. Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P: **MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data.** *Mol Ecol Notes* 2004, **4**:535–538.

50. Hutter CM, Schug MD, Aquadro CF: **Microsatellite variation in *Drosophila melanogaster* and *Drosophila simulans*: A reciprocal test of the ascertainment bias hypothesis.** *Mol Biol Evol* 1998, **15**:1620–1636.

51. Vowles EJ, Amos W: **Quantifying ascertainment bias and species-specific length differences in human and chimpanzee microsatellites using genome sequences.** *Mol Biol Evol* 2006, **23**:598–607.

52. Released IC: *IBM SPSS Statistics for Windows. 20.0 edition*. Armonk, NY: IBM Corp.; 2013.

53. Queller DC, Goodnight KF: **Estimating relatedness using genetic markers.** *Evolution* 1989, **43**:258–275.
54. Lynch M, Ritland K: **Estimation of pairwise relatedness with molecular markers.** *Genetics* 1999, **152**:1753–1766.
55. Peakall R, Smouse PE: **GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update.** *Bioinformatics* 2012, **28**:2537–2539.
56. Kalinowski ST, Wagner AP, Taper ML: **ML-RELATE: a computer program for maximum likelihood estimation of relatedness and relationship.** *Mol Ecol Notes* 2006, **6**:576–579.
57. Kalinowski ST, Taper ML: **Maximum likelihood estimation of the frequency of null alleles at microsatellite loci.** *Conserv Genet* 2006, **7**:991–995.
58. Waples RS, Do C: **LDNE: a program for estimating effective population size from data on linkage disequilibrium.** *Mol Ecol Resour* 2008, **8**:753–756.
59. Do C, Waples RS, Peel D, Macbeth GM, Tillett BJ, Ovenden JR: **NEESTIMATOR v2: re-implementation of software for the estimation of contemporary effective population size (N-e) from genetic data.** *Mol Ecol Resour* 2014, **14**:209–214.
60. Cornuet JM, Luikart G: **Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data.** *Genetics* 1996, **144**:2001–2014.
61. Holm S: **A simple sequentially rejective multiple test procedure.** *Scand Stat Theory Appl* 1979, **6**:65–70.
62. Chen H, Fillinger U, Yan G: **Oviposition behavior of female *Anopheles gambiae* in western Kenya inferred from microsatellite markers.** *Am J Trop Med Hyg* 2006, **75**:246–250.
63. Lehmann T, Light M, Gimnig JE, Hightower A, Vulule JM, Hawley WA: **Spatial and temporal variation in kinship among *Anopheles gambiae* (Diptera : Culicidae) mosquitoes.** *J Med Entomol* 2003, **40**:421–429.
64. Blouin MS: **DNA-based methods for pedigree reconstruction and kinship analysis in natural populations.** *Trends Ecol Evol* 2003, **18**:503–511.
65. Fornadel CM, Norris LC, Franco V, Norris DE: **Unexpected anthropophily in the potential secondary malaria vectors *Anopheles coustani* s.l. and *Anopheles squamosus* in Macha, Zambia.** *Vector Borne Zoonotic Dis* 2011, **11**:1173–1179.
66. Mwangangi JM, Muturi EJ, Muriu SM, Nzovu J, Midega JT, Mbogo C: **The role of *Anopheles arabiensis* and *Anopheles coustani* in indoor and outdoor malaria transmission in Taveta District Kenya.** *Parasit Vectors* 2013, **6**:114.

67. Molineaux L, Shidrawi GR, Clarke JL, Boulzaguet JR, Ashkar TS: **Assessment of insecticidal impact on the malaria mosquito's vectorial capacity, from data on the man-biting rate and age-composition.** *Bull World Health Organ* 1979, **57**:265–274.

68. Molineaux L, Gramiccia G: *The Garki Project. Research on the Epidemiology and Control of Malaria in the Sudan Savanna of West Africa.* Geneva: World Health Organization; 1980.

Additional files

Additional_file_1 as DOCX

Additional file 1 Estimates of genetic diversity per microsatellite locus.

Additional_file_2 as DOCX

Additional file 2 Species identification by DNA barcoding using BOLD database.

Additional_file_3 as DOCX

Additional file 3 Number and proportion (percentages in parenthesis) of *Anopheles coluzzii*, *Anopheles gambiae* and admixed individuals in larval and adult samples.

Additional_file_4 as DOCX

Additional file 4 Correlation between allele richness and length of the original microsatellite clone isolated from *Anopheles gambiae*.

Additional_file_5 as DOCX

Additional file 5 Proportion (in percentage) of individuals assigned to each pedigree class by ML-RELATE [56].

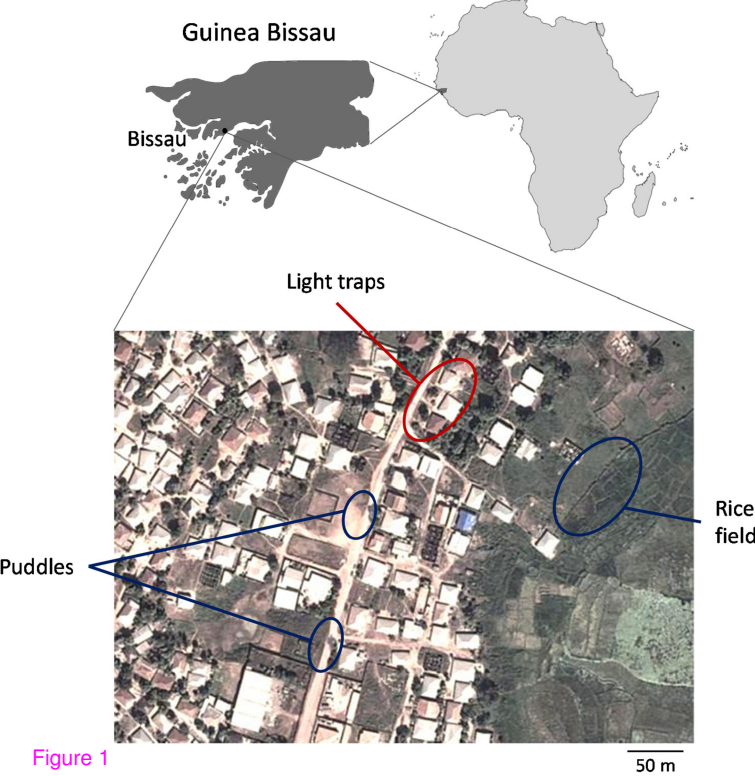
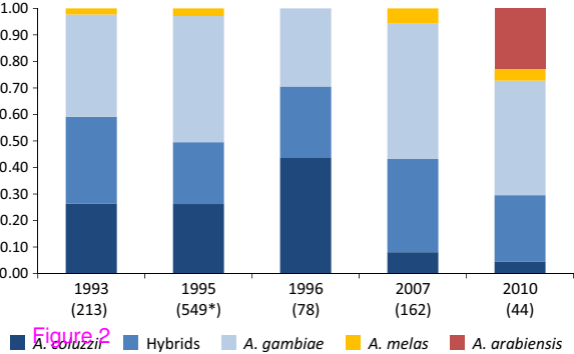


Figure 1



Additional files provided with this submission:

Additional file 1: 2647811691390333_add1.docx, 49K

<http://www.malariajournal.com/imedia/2195869861486396/supp1.docx>

Additional file 2: 2647811691390333_add2.docx, 26K

<http://www.malariajournal.com/imedia/1311370778148639/supp2.docx>

Additional file 3: 2647811691390333_add3.docx, 22K

<http://www.malariajournal.com/imedia/2093231242148639/supp3.docx>

Additional file 4: 2647811691390333_add4.docx, 19K

<http://www.malariajournal.com/imedia/9609653101486396/supp4.docx>

Additional file 5: 2647811691390333_add5.docx, 20K

<http://www.malariajournal.com/imedia/9845098251486396/supp5.docx>